DNA barcoding, an approach for molecular identification of Huyen-sam (*Scrophularia* L.) samples collected in Northern Vietnam

Manh Minh Bui[1], Anh Tuan Vu[2], Phuong Nhung Vu[1], Quang Cu Pham[2], Dang Ton Nguyen[1,3], Thi Thu Hue Huynh[1,3*]

[1]*Institute of Genome Research, Vietnam Academy of Science and Technology*
[2]*General Department of Logistics - Techniques*
[3]*Graduate University of Science and Technology, Vietnam Academy of Science and Technology*

Received 4 December 2017; accepted 26 March 2018

Abstract:

Huyen-sam (Vietnamese name) which belongs to *Scrophularia* L. genus is a valuable herb. This medicinal plant is classified as *Scrophularia ningpoensis* Hemsl. Huyen-sam roots, which contain a large amount of bioactive compounds, have a similar morphology to its relatives. DNA barcodes promise to be a precise and reliable tool for distinguishing the processed Huyen-sam materials from their counterfeits. However, studies about using DNA barcodes for classification of *Scrophularia* L. in Vietnam are not available. Here, we conducted a taxonomic analysis of eight *Scrophularia* L. samples collected from the mountain areas of Northern Vietnam. Based on the combined sequence data of ribosomal nuclear *ITS*, a part of chloroplast *rbcL* gene and *trnL-trnF* intergenic spacer, phylograms of *Scrophularia* L. were generated by both Bayesian inference and maximum likelihood bootstrap method. The phylogenetic analysis showed that the tested samples have a sister relationship to *S. ningpoensis*. Hopefully, the analysis strategy that we used would contribute to further phylogenetic analyses of medicinal plants of Vietnam in the future.

Keywords: ***Bayesian inference, DNA barcodes, ITS, maximum likelihood, medicinal plants, phylogenetic, rbcL, Scrophularia L., trnL-trnF.***

Classification numbers: ***3.3, 3.5***

Introduction

Scrophularia L., which is commonly called "figwort" is a plant genus belonging to family *Scrophulariaceae*. The genus comprises about 200-300 species distributed in Central Asia, Europe (Mediterranean), North America and China [1-3]. Huyen-sam (Vietnamese name) or *Scrophularia ningpoensis* Hemsl., whose root is a valuable natural herb, is usually used for the treatment of inflammation, constipation and fever [4-6]. The main bioactive compounds present in *S. ningpoensis*'s root are harpagoside, angroside C, acteoside and cinnamic acid, which have anti-inflammatory, antimicrobial and antioxidant effects [3, 6, 7]. Sourced from Southeast China, this herb is also domestically grown in some northern districts of Vietnam, such as Lao Cai, Ha Giang and Cao Bang [6]. Due to the similarity in the morphology, *S. ningpoensis*'s root can be mistaken for its close relatives, such as *S. buergeriana* Miq. or *S. kakudensis* Franch. Consequently, the demand for new molecular markers that support the identification of processed *S. ningpoensis*'s samples has become increasingly necessary. However, phylogenetic studies based on molecular markers of *Scrophularia* L. are very limited in Vietnam. These discoveries would play an important role in assuring the quality of processed herb in Vietnam market.

DNA barcoding is a conventional method for the identification of unknown living organism specimens. This approach can be applied to a wide range of species from microbes to higher animals. By analysing the evolutionary rate of small genome fragments as substitutes for morphology aspects, the method provides a quick and cost-effective species identification, especially for higher plant taxons [8, 9]. The searching for universal DNA

*Corresponding author: Email: hthue@igr.ac.vn

barcodes for plants is still ongoing; however, there is a common agreement that more than one region is needed for performing the taxonomy consensus analysis [10, 11].

The selected plant DNA barcodes are usually the genome regions which have a suitable evolutionary rate for generating enough changes in various nucleotide sites during generations. The majority of plant DNA barcoding studies have utilized the DNA regions located on the plastid genome, e.g. *ribulose 1,5-bisphosphate carboxylase large subunit (rbcL)*, *maturase K (matK)* and multiple intergenic regions such as tRNA Leucine - tRNA Phenylanaline (*trnL-trnF*), tRNA Histidine - photosystem binding protein A (*trnH-psbA*), tRNA Glutamine - ribosomal protein S16 (*trnQ-rps16*), tRNA Cysteine - tRNA Asparagine (*trnC-trnD*), and tRNA Alanine - tRNA Histidine (*trnA-trnH*) [11-15]. In addition, the nuclear internal transcribed spacer (ITS) and 18s RNA can also perform as useful barcodes for classifying flower plants [11, 16-18].

During the last decade, taxonomy studies of *Scrophularia* L. based on the DNA barcodes have been gradually conducted. The phylogenetic relationships among *Scrophularia* L. taxa collected from different parts of American continent were analysed from the combined data from the sequence of *ITS*, the chloroplast *trnQ-rps16* and *psbA-trnH* intergenic spacers [2]. Another study on the evolutionary relationships of *Scrophularia* L. species in Western Mediterranean and Macaronesia was completed by the data of ITS and *trnQ-rps16* by Bayesian binary MCMC (BBM) analysis [17].

In this paper, we perform a phylogenetic analysis of *Scrophularia* L. samples in Northern Vietnam. More particularly, the analysis is based on the sequence data from nuclear *ITS*, chloroplast *rbcL,* and *trnL-trnF*. The combined data set from these DNA regions are objects for generating the phylogenetic tree by Bayesian inference (BI) and maximum likelihood (ML) analysis. The main aim of this project is to contribute to the molecular classification study of genus *Scrophularia* L. in Vietnam.

Materials and methods

Plant materials

Eight leaf samples of *Scrophularia* L. were collected from the cultivated gardens located in different mountain districts in Northern Vietnam namely HSa-1A, HSa-1B, HSa-2A, HSa-3A, HSa-3B, HSa-4B, HSa-5A, HSa-8A. All the samples were preserved in silica gel for a completed desiccation.

DNA extraction, amplification and sequencing

Total DNA was extracted from about 100 mg of the dried leaf following the CTAB extraction method [19]. The extracted DNA was resuspended in 50 µl miliQ water, and standard 50 ng of the DNA was used for amplification. The primers for amplification of target regions were designed based on the reference sequence on Genbank (Table 1).

Table 1. List of primers used in the study.

Primer	DNA regions	Primer sequences (5'→3')	Amplicon size (bp)
ITS-AB-101	*ITS*	ACGAATTCATGGTCCGGTGAAGTGTTCG	800
ITS-AB-102		TAGAATTCCCCGGTTCGCTCGCCGTTAC	
rbcL-F	*rbcL*	ATTTGAACTGGTGACACGAG	600
rbcL-R		CGAAATCGGTAGACGCTACG	
TrnL-PF	*TrnL-trnF*	AGTGTTGGATTCAAGCTGGTG	1100
TrnL-PR		TGGTTGTGAGTTCACGTTCT	

The condition of amplification was optimized for 20 µl of PCR, including 50 ng extracted DNA, 2.5 µM of each primer, 0.75 unit of Phusion polymerase (Thermo Scientific), 1 mM of each dNTP and Phusion PCR buffer. The amplification thermocycles were performed as follows: 1 cycle of denaturing at 94°C for 4 minutes, 35 cycles of amplification including 94°C/30s followed by annealing at 52°C/30s (*trnL-trnF* and *rbcL*) or 54°C/30s (*ITS)* and extension 72°C/1 min 30s; ending with a final extension step of 72°C/7 mins. The PCR products were checked by electrophoresis on 0.8% agarose gel. Successful PCR products were purified by Thermofisher Scientific DNA

purification kit (K0512). Sequencing was carried out using the BigDyeTM terminator v3.1 cycle sequencing kit (Applied Biosystems) in a final volume of 20 µl. Sequence runs were performed on an ABI 3500 genetic analyser following Sanger's principle.

Alignment and phylogenetic construction

All the DNA sequences generated from this study were assembled, edited and aligned manually using Bioedit v7.0.5.9 which embedded the ClustalW v1.8 [20]. To access the closeness of the relationships between tested plant samples and the species of *Scrophularia* L., the DNA sequences of 3 examined regions namely *ITS, rbcL* and trnL-*trnF* of species involving in genus *Scrophularia* L. and some other genera of *Lamiales* as outgroup were downloaded from Genbank (www.ncbi.nlm.nih.gov) and aligned (Table 2).

Table 2. Taxons included in this study, with Genbank accession numbers.

Species	GenBank accession numbers		
	ITS	***rbcL***	***trnL-trnF***
Scrophularia ningpoensis	FJ609731.1	GQ436721.1	AY695886.1
Scrophularia buergeriana	JQ065663.1	NC031437.1	KP718626.1
Scrophularia takesimensis	JQ065681.1	KP718628.1	AY695886.1
Scrophularia kakudensis	JQ065674.1	-	KM979600.1
Scrophularia zvartiana	KY067618.1	-	-
Scrophularia arguta	-	-	AJ430936.1
Scrophularia californica	-	-	HQ412946.1
Orobanche gracilis	JX193303.1	AY582198.1	-
Orobanche californica	KC480368.1	AY582178.1	-
Phelipanche ramosa	AY209315.1	AY582252.1	-
Conopholis americana	AY209289.1	-	-
Epifagus virginiana	AY209290.1	-	-
Lathraea squamaria	KC480353.1	-	-
Plantago maritima	AY101879.1	KR297244.1	AY101934.1
Plantago media	AJ548964.1	KF602241.1	AY101920.1
Coffea canephora	MF417755.1	NC030053.1	AF102405.2
Coffea arabica	MF417758.1	EF044213.1	DQ153829.1
Jasminum nudiflorum	AF534817.1	DQ673255.1	EU281146.1
Olea europaea	KF805102.1	DQ673304.1	AF231866.1
Olea woodiana	JX862658.1	NC015608.1	LN515476.1
Utricularia macrorhiza	-	NC025653.1	AF482657.1
Utricularia gibba	-	NC021449.1	AF482657.1
Ajura reptans	EF508061.1	Z37385.1	GU381470.1
Salvia miltiorrhiza	DQ132863.1	KC473307.1	DQ667523.1
Pogostemon stellatus	KP718621.1	NC031434.1	NC031434.1
Phyllostegia velutina	KF529547.1	NC029820.1	KU724134.1
Stachys sylvatica	JN680361.1	AF502022.1	NC029824.1

-: the sequence is unavailable.

The phylogenetic taxonomy analysis was conducted with a BI and an ML approach from 3 datasets (ITS, chloroplast and combined data). The BI was calculated by MrBayes 3.2.6 using a Metropolis-coupled Markov Chain Monte Carlo (MCMCMC) algorithm [21, 22]. The certainty of the node generated by BI were supported by the posterior probability (PP) value, which ranged from 0 to 1. The Combined chloroplast partition evolution was assumed to follow the general time reversible model with a proportion of the sites invariable and the rate for the remaining sites drawn from a gamma distribution (GRT+Γ+I model), while ITS region follows a SYM model (SYM+Γ model) [23]. The MrBayes was executed with 2 runs and four chains (3 hot - 1 cold) with the default temperature of hot chain t = 0.2 for 10 million generations, sampling every 2000th generation to generate 10,000 trees. A burn-in ratio of 10% of sampled trees was discarded and the BI consensus tree were generated from 80% of the remainders. Besides, the alignment of the combined dataset (*ITS, rbcL* and *trnL-trnF*) and ML was performed by the software MEGA 6.06 with the bootstrap method of 1,000 replications [24]. The consensus tree was drawn using Figtree v.1.2.3.

For controlling the incongruence between phylogenetic trees generated by BI and ML bootstrap method, the phylograms received from the chloroplast and nuclear datasets were analysed separately before combining. The incongruence taxons and nodes with a high level of BS (70%) or surpassing the Bayesian support of 85% were discarded [25, 26].

Results

DNA extraction and amplification

The extracted DNA from examined samples were used as templates for amplification of *ITS, rbcL* and *trnL-trnF* regions using designed primers. The size of PCR products was checked by electrophoresis on Agarose gel 0.8%. The correct PCR products were purified and sequenced for generating the DNA sequences afterward (Fig. 1).

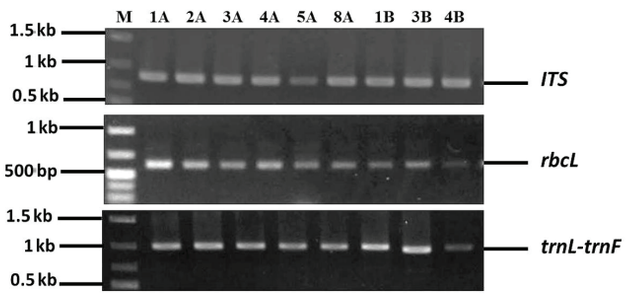

Fig. 1. Gel electrophoresis image of PCR products from *Scrophularia* samples on 0.8% agarose gel. 1A, 2A, 3A, 5A, 8A, 1B, 3B, and 4B are correspondent to HSa-1A, HSa-2A, HSa-3A, HSa-5A, HSa-8A, HSa-1B, HSa-3B and HSa-4B, respectively. The sample *ITS, rbcL* and *trnL-trnF* fragments have the size 800, 600 and 1,100 bp, respectively. The PCR products clearly showed the DNA bands with the correct size of desired fragments.

DNA sequence alignment

In this study, a total of 24 sequences were generated from tested *Scrophularia* samples including 8 each for the nuclear *ITS* regions, chloroplast *rbcL* genes and *trnL-trnF* intergenic spaces. The sequences obtained from examined samples were aligned with correspondent references for creating 3 separated alignment datasets namely ITS, Chloroplast combination (*rbcL+trnL-trnF*) and Mixed combination (*ITS+rbcL+trnL-trnF*). The Chloroplast combination is a merger of 2 plastid sequence, while the Mixed combination was generated by adding the ITS sequence to the Chloroplast combination. The mixed combined data matrix contained 2,065 aligned characters with the average sequence length of 1,826.1 bp. The average sequence length was 544.9 bp for ITS alignment data, 551.7 bp for *rbcL* and 729.8 bp for *trnL-trnF.* We also estimated the mean of evolution distance between the taxons included in each dataset using the Maximum composite likelihood model with Gamma distribution and assuming rate variation and pattern heterogeneity among sites. The *trnL-trnF* showed the highest overall mean of evolution distance (0.414) followed by the ITS regions (0.374), which indicates that these regions have relatively high evolution rates on *Scrophularia* L. genus. The detailed statistic information about the aligned dataset, including mean G+C content, number of conserved nucleotides and parsimony-informative sites, were provided in Table 3.

Table 3. Alignment characteristics and statistics for ITS, *rbcL* region, *trnL-trnF* intergenic space, combined chloroplast, and combined dataset.

	ITS	rbcL	trnL-trnF	Comb. Chloroplast	Combined
Number of taxa	32	31	33	26	23
Average sequence length (bp)	544.9	551.7	729.8	1277.6	1826.1
Aligned sequence length	608	580	1019	1457	2065
Conserved characters	228	376	223	788	1109
Parsimony-informative characters	298	112	538	383	601
Overall mean evolution distance	0.347	0.08	0.414	0.111	0.09
%G+C content	61.5	45.2	35.2	38.7	45.6

The table contains the number of conserved characters, parsimony - informative characters and mean % GC content of aligned sequences. Overall mean evolution distance was estimated through maximum composite likelihood model with Gamma distribution and assuming rate variation and pattern heterogeneity among sites.

Phylogenetic tree construction

In this study, we generated phylogenetic trees for three separately aligned datasets. For all the BI consensus trees, the average standard deviations of split frequencies when 2-run coverage at stationary distribution remained at lower than 0.002. The low split frequency indicated an increasing similarity of runs, and the results were adequate for the next analysis. The phylogenetic trees were obtained from 9,002 sampled trees after the Bayesian runs. The BI and ML analyses of each dataset showed high congruence on topologies, especially on the *Scrophulariaceae* family. Therefore, we plotted the BS value of ML analysis onto the respective BI consensus tree.

Individual phylogenetic trees generated from the nuclear barcode *ITS* and Chloroplast combination dataset (*rbcL+trnL-trnF*) were shown in Figs. 2 and 3, respectively. The clades of the examined sample in both consensus trees are highly similar in branch length with each other and grouped with *Scrophularia ningpoensis,* suggesting a close relationship. However, the node between *S. ningpoensis* and HSa-8A was not well-supported by the BI analysis (PP: 0.53, BS: 97). The BI and ML analyses also did not support the node of *S. takesimensis* and *S. zvartiana* (PP: 0.68, BS: 43) and the node of outgroup and tested samples (PP: 0.5, BS: 7), indicating the uncertainty of the ITS trees (Fig. 2). On the outgroup clades, the *Lathraea squamaria* was incorrectly ordered into the clade of *Orobanchaceae* instead

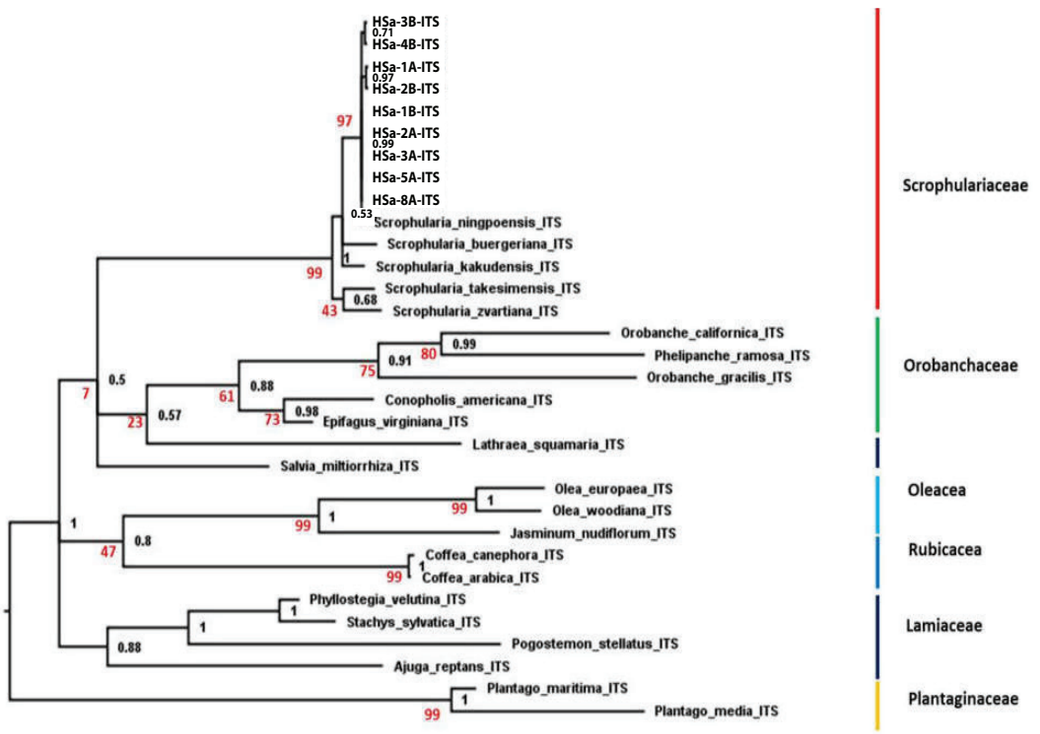

Fig. 2. Bayesian consensus tree of *Scrophularia* L., generated from the ITS dataset and reference sequences from 5 other genera of *Lamiales*. PP values are given in black number next to each node, while the corresponding BS values are given in red numbers. PP values are obtained from 9,002 trees. The scale bar indicates the average expected changes per site of sequences in the study.

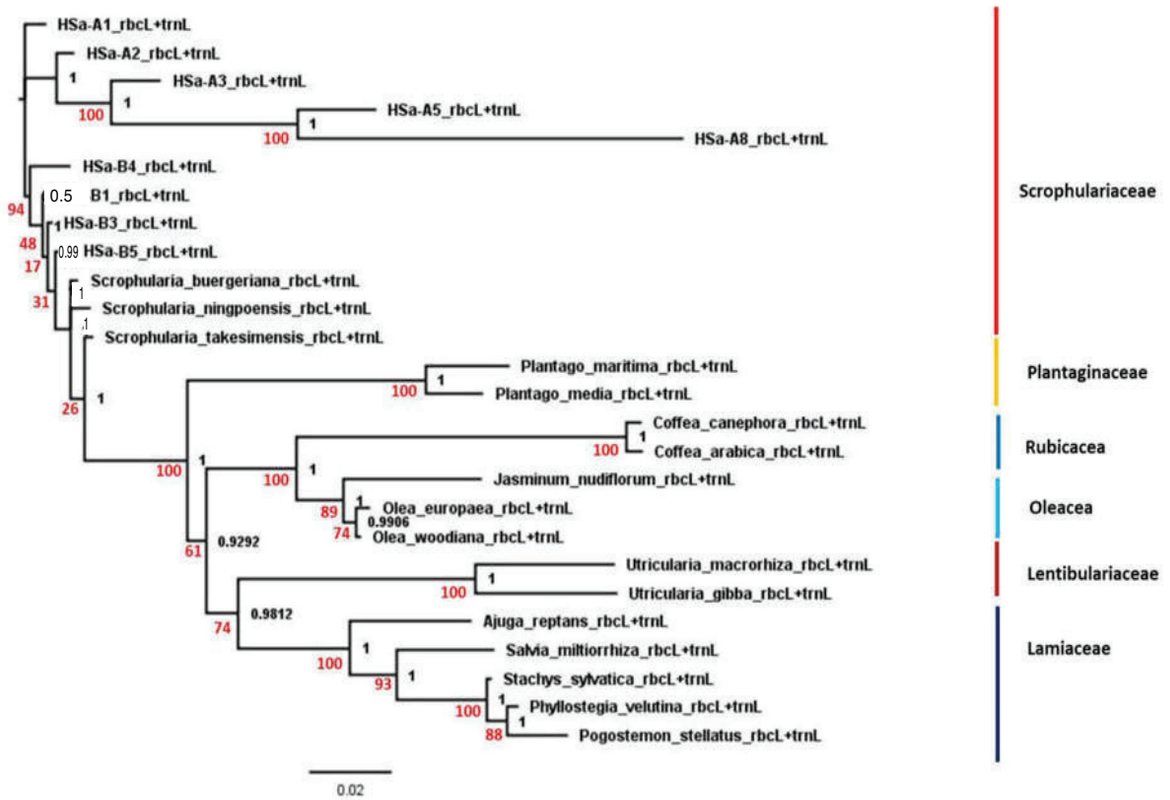

Fig. 3. Bayesian consensus tree *Scrophularia* L., generated from the Chloroplast combination dataset (*rbcL* and *trnL-trnF* intergenic spacers) of tested samples and reference sequences from 5 other genera of *Lamiales*. PP are given in black number next to each node, while the corresponding BS values are given in red numbers. PP values are obtained from 9,002 sampled trees. The scale bar indicates the average expected changes per site of sequences in the study.

of Lamiaceae; however, the PP and BS values for this classification were quite low (PP: 0.57, BS: 23), indicating uncertainty.

Turning to Combined chloroplast consensus tree (Fig. 3), all the nodes were well supported by the BI analysis (i.e. PP values are larger than 0.9) and the clade of outgroups had high values of both BS and PP. The branch length of taxons HSa-A1, HSa-AB1, HSa-A2, HSa-B3, HSa-B5 were similar and grouped together on a clade of the consensus tree. This clade also had a close relationship with *S. buergeriana* and *S. ningpoensis*. Nevertheless, the node between the tested group and *S. buergeriana* received a weak BS value from ML bootstrap analysis (BS: 31). A similar circumstance occurred with the node between the outgroups and tested samples (BS: 26).

The topologies involving the phylogram generated from the mixed combination dataset were relatively congruent with the Chloroplast combination tree (Fig. 4). All the nodes received a good support from BI analysis. The clades including examined samples (HSa-3A, HSa-1B, HSa-1A, HSa-4B, HSa-2A) were highly supported to be the sister of *S.ningpoensis* (PP: 1, BS: 91), while it is quite weak on ITS tree (PP: 0.53, BS: 97). The HSa-5A and HSa-8A were classified into a separated clade with high PP and BS value (PP: 1, BS: 100), suggesting that they belong to a completely different group involved in the tested samples. In addition, all the outgroup clades showed a consistency with the APGII classification (Angiosperm Phylogeny Group, 2003) with a high support from BI and ML bootstrap analysis. This result indicated the high efficiency of using Combined dataset for building phylogenetic trees not only for *Scrophulariaceae,* but also other families of *Lamiales*.

Discussion and conclusions

Scrophularia L. is one small genus which belongs to *Scrophulariaceae* family. The genus includes a relatively small number of species (about 200-300) in comparison

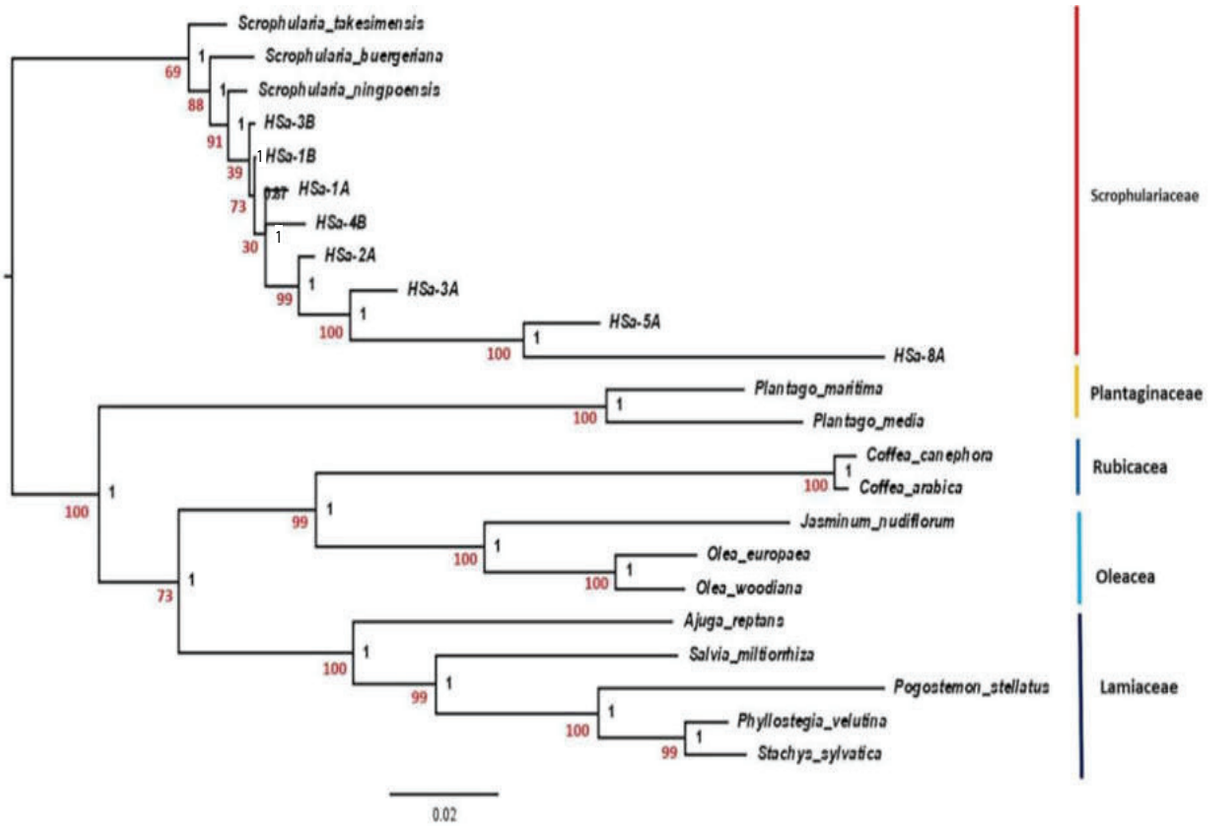

Fig. 4. Bayesian consensus tree *Scrophularia* L., generated from a mixed combination dataset (*ITS, rbcL* and *trnL-trnF* intergenic spacers) of tested samples and reference sequences from 4 other genera of *Lamiales*. PP are given in black number next to each node, while the corresponding BS values are given in red numbers. PP values are obtained from 9,002 sampled trees. The scale bar indicates the average expected changes per site of sequences in the study.

with the total number species of *Scrophulariaceae* family; however, a lot of them are utilized as traditional medicine. In Vietnam, the first study on *Scrophularia* L. was instituted by Do Tat Loi in 1962 under a catalogue of medical plants in Vietnam. The study classified Huyen-sam as *S. buergeriana* Miq. or *S. oldhami* with numerous benefits for human health, such as pulse-quickening, anti-inflammation, and antibiotic [27]. In the Vietnam Plant data Centre, the classification of *Scrophulariaceae* is largely based on the morphology of reproductive trait; e.g. stamen exsertion, corolla shape or leaf organization structure (www.botanyvn.com). However, the morphology knowledge for distinguishing *Scrophularia* L's processed roots which are valuable herbs is unavailable. In this study, we aim to develop a strategy using DNA barcoding, a molecular approach, to support the classification of *Scrophularia* L. samples in Vietnam.

DNA barcode is versatile and cost-effective for the plant taxonomist. The most remarkable advantage of using DNA barcoding is the wide range of applicable plant samples. The method could be applied for DNA samples obtained from different parts of the plant (e.g. leaf, root, flower...) in various kinds of preservation conditions (fresh, dry…). Upon DNA barcode analysis, the taxon identification can be processed without a detailed description of morphology [28]. In addition, DNA barcode information could support the finding of new species from a collection or confirmation of preserved plant materials [29].

Here, we followed a general pipeline of taxonomic analysis of *Scrophularia* L. genus on three controversial barcode regions *(ITS, rbcL* and *trnL-trnF*). For more details, *ITS* is well-known as one of the most importance barcodes for plant classification. This region includes 2 more variable partitions namely ITS1, ITS2 and a conserved 5.8S sequence. Due to the convenience of amplification, the ITS regions are widely used for performing taxonomy analysis of the fungi, monocot and dicot [10, 11]. However, the

ITS has a quite complex evolution pattern that correlates with nuclear genome and causes difficulties for analysis [8]. In this study, we applied a SYM+Γ substitution model which was suggested by Scheunert and colleagues [23] for generating the phylogram from the ITS data. The utilization of ITS dataset was also integrated into a large number of Scrophulariaceae family classification studies [2, 17, 30].

We also did the taxonomy analysis with the two-locus barcode located on chloroplast (*rbcL* and *trnL-trnF*) which are also widely used as plant DNA barcodes for *Scrophulariaceae* [28, 31, 32]. By merging the Chloroplast dataset with the ITS data, we have improved the reliability level of the analysis with a higher value of BI and ML bootstrap analysis. The combination of multi-loci for generating phylogenetic trees is a controversial method for reducing the inconsistency from different single locus analyses and creating a 'total evidence' approach [25]. In addition, the utilization of ITS locus combined with two plastid loci is proposed as the silver standard method for land plant classification [10]. The adding of a locus which has lower rates of evolution in plant plastids such as spacer regions (*trnH-psbA, trnL-trnF*…) and *rbcL* has shown more effective and precise results in separating closely related plants [33].

In conclusion, we have identified a close relationship between the Huyen-sam samples HSa-3A, HSa-1B, HSa-1A, HSa-4B, HSa-2A with the *S. ningpoensis* using combined DNA barcodes generated from 3 loci namely *ITS, rbcL* and *trnL-trnF.* The sequence data generated from this project could enhance the further studies about the diversity of medicinal plant in Vietnam or confirmation of plant material in Vietnam herb market. The phylogenetic analysis followed a general pipeline with BI and ML bootstrap analysis. This strategy could be promisingly applied for not only *Scrophularia* L., but also other valuable herbs of the other genera in Vietnam.

ACKNOWLEDGEMENTS

This study was supported by the Preservation of Vietnamese Herbarium genetic resources for medicine development project. All the experiments and analysis in the research were performed at Institute of Genome Research, Vietnam Academy of Science and Technology, Hanoi, Vietnam.

REFERENCES

[1] E. Fischer (2004), *Scrophulariaceae (in The families and genera of vascular plants)*, Springer, **7**, pp.333-432.

[2] A. Scheunert, G. Heubl (2011), "Phylogenetic relationships among new world *Scrophularia* L. (Scrophulariaceae): New insights inferred from DNA sequence data", *Plant Systematics and Evolution,* **291(1-2)**, pp.69-89.

[3] J. Tian, X. Ye, Y. Shang, Y. Deng, K. He, X. Li (2012), "Preparative isolation and purification of harpagoside and angroside C from the root of *Scrophularia ningpoensis* Hemsley by high-speed counter-current chromatography", *Journal Separation Science*, **35(19)**, pp.2659-2664.

[4] T. Kajimoto, M. Hidaka, K. Shoyama, T. Nohara (1989), "Iridoids from *Scrophularia ningpoensis*", *Phytochemistry,* **28(10)**, pp.2701-2704.

[5] A.P. Sagare, C.L. Kuo, F.S. Chueh, H.S. Tsay (2001), "De novo regeneration of scrophularia yoshimurae Yamazaki (*Scrophulariaceae*) and quantitative analysis of harpagoside, an iridoid glucoside, formed in aerial and underground parts of in vitro propagated and wild plants by HPLC", *Biological and Pharmaceutical Bulletin*, **24(11)**, pp.1311-1315.

[6] Nguyen Tien Ban (2005), *Scrophulariaceae (in Plant category of Vietnam)*, Agricultural Publisher, **3**, p.233.

[7] S. Yang, J. Li, Y. Zhao, B. Chen, C. Fu (2011), "Harpagoside variation is positively correlated with temperature in *Scrophularia ningpoensis* Hemsl", *J. Agricultural Food Chemistry,* **59(5)**, pp.1612-1621.

[8] O. Seberg, G. Petersen (2009), "How many loci does it take to DNA barcode a crocus?", *PLOS ONE,* **4(2)**, p.e4598, doi: 10.1371/journal.pone.0004598.

[9] L.T.T. Hien, H.D. Boer, N.V. Hai, L.T. Huong, N.M. Huong, L. Bjork (2012), "DNA molecule bar code and bar code data system", *Journal of Biotechnology (VN),* **10**, pp.393-405.

[10] M.W. Chase, et al. (2005), "Land plants and DNA barcodes: short-term and long-term goals", *Philosophical Transactions of the Royal Society, Biological Sciences,* **360(1462)**, pp.1889-1895.

[11] W.J. Kress, et al. (2005), "Use of DNA barcodes to identify flowering plants", in *Proceedings of the National Academy of Sciences of the United States of America,* **102**, pp.8369-8374.

[12] D.Y. Wang, et al. (2017), "Evaluation of DNA barcodes in Codonopsis (Campanulaceae) and in some large angiosperm plant genera", *PLOS ONE,* **12(2)**, p.e0170286, doi.org/10.1371/journal.pone.0170286.

[13] J.M. Saarela, et al. (2013), "DNA barcoding the Canadian Arctic flora: core plastid barcodes (*rbcL* + *matK*) for 490 vascular plant species", *PLOS ONE,* **8(10)**, p.e77982, doi.org/10.1371/journal.pone.0077982.

[14] X. Guo, X. Wang, W. Su, G. Zhang, R. Zhou (2011), "DNA barcodes for discriminating the medicinal plant Scutellaria baicalensis (Lamiaceae) and its adulterants", *Biological and Pharmaceutical Bulletin,* **34(8)**, pp.1198-1203.

[15] C. Garcia Robledo, et al. (2013), "Tropical plant-herbivore networks: reconstructing species interactions using DNA barcodes", *PLOS ONE,* **8(1)**, p.e52967, doi.org/10.1371/journal.pone.0052967.

[16] L.T. Hương, N.N. Linh, B.M. Minh, H.H. Hạnh, H.T.T. Hue, N.V. Hai, H.V. Huan, L.T.T. Hien (2017), "Applied DNA barcode assists

in the identification of some ginseng species in the genus ginseng (*Panax L.*)", *Journal of Biotechnology (VN)* **15**, pp.63-72.

[17] A. Scheunert, G. Heubl (2014), "Diversification of Scrophularia (Scrophulariaceae) in the western mediterranean and macaronesia-phylogenetic relationships, reticulate evolution and biogeographic patterns", *Molecular Phylogenetics and Evolution,* **70**, pp.296-313.

[18] G.N. Feliner, J.A. Rosselló (2007), "Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants", *Molecular Phylogenetics and Evolution,* **44(2)**, pp.911-919.

[19] J.D. Clarke (2009), "Cetyltrimethyl ammonium bromide (CTAB) DNA miniprep for plant DNA isolation", *Cold Spring Harbor Protocols*, **2009(3)**, doi: 10.1101/pdb.prot5177.

[20] T.A. Hall (1999), "BioEdit (version 7.0.5.1): a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT", *Nucl. Acids Symp. Ser.,* **41**, pp.95-98.

[21] F. Ronquist, J.P. Huelsenbeck (2003), "MrBayes 3: Bayesian phylogenetic inference under mixed models", *Bioinformatics*, **19(12)**, pp.1572-1574.

[22] F. Ronquist, et al. (2012), "MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space", *Systematic Biology*, **61(3)**, pp.539-542.

[23] A. Scheunert, et al. (2012), "Phylogeny of tribe Rhinantheae (Orobanchaceae) with a focus on biogeography, cytology and re-examination of generic concepts", *Taxon*, **61(6)**, pp.1269-1285.

[24] K. Tamura, et al. (2013), "MEGA6: molecular evolutionary genetics analysis version 6.0", *Molecular Biology and Evolution,* **30(12)**, pp.2725-2729.

[25] J.P. Huelsenbeck, J.J. Bull, C.W. Cunningham (1996), "Combining data in phylogenetic analysis", *Trends in Ecology & Evolution*, **11(4)**, pp.152-158.

[26] M.A. Perez Gutierrez, et al. (2012), "Phylogeny of the tribe Fumarieae (Papaveraceae s.l.) based on chloroplast and nuclear DNA sequences: evolutionary and biogeographic implications", *American Journal of Botany*, **99(3)**, pp.517-528.

[27] D.T. Loi (2004), *Vietnamese medicinal plants and herbs*, Medical Publishing House.

[28] D. Do, L.Z. Drábková (2017), "Herbarium tale: The utility of dry specimens for DNA barcoding Juncaceae", *Plant Systematics and Evolution*, **304(2)**, pp.1-14.

[29] D.E. Schindel, S.E. Miller (2005), "DNA barcoding a useful tool for taxonomists", *Nature*, **435**, p.17, doi:10.1038/435017b.

[30] A. Scheunert, G. Heubl (2017), "Against all odds: Reconstructing the evolutionary history of Scrophularia (Scrophulariaceae) despite high levels of incongruence and reticulate evolution", *Organisms Diversity & Evolution*, **17(2)**, pp.323-349.

[31] K.L. Bell, V.M. Loeffler, B.J. Brosi (2017), "An *rbcL* reference library to aid in the identification of plant species mixtures by DNA metabarcoding", *Applications in Plant Sciences*, **5(3)**, doi: 10.3732/apps.1600110.

[32] J.H. Chau, N. O'Leary, W.B. Sun, R.G. Olmstead (2017), "Phylogenetic relationships in tribe Buddlejeae (Scrophulariaceae) based on multiple nuclear and plastid markers", *Botanical Journal of the Linnean Society*, **184(2)**, pp.137-166.

[33] E. Pennisi (2007), "Wanted: a barcode for plants", *Science*, **318(5848)**, pp.190-191, doi: 10.1126/science.318.5848.190.