Identification of sanchi samples based on DNA barcodes

Dong T. Pham, & Phong V. Nguyen*
Faculty of Biological Sciences, Nong Lam University, Ho Chi Minh City, Vietnam

ARTICLE INFO

Research Paper

Received: April 10, 2023
Revised: August 21, 2023
Accepted: September 20, 2023

Keywords

DNA barcodes
ITS1
MatK
Phylogenetic tree
RbcL
TrnH-psbA

***Corresponding author**

Nguyen Vu Phong
Email:
nvphong@hcmuaf.edu.vn

ABSTRACT

For centuries, sanchi has been used in traditional medicine in East Asian countries to promote health and fitness. To identify and distinguish sanchi from adulterants, this study was conducted to determine the sequence of four DNA barcodes. The PCR products of the four barcodes were 330 bp (*rbcL*), 822 bp (*matK*), 484 bp (*trnH-psbA*), and 438 bp (ITS1). Genetic relationship analysis showed that the four DNA barcode regions had high similarity with other *Panax* species. Three barcodes, *matK*, *trnH-psbA*, and ITS1, confirmed that the studied samples belonged to *Panax notoginseng*, distinguished from *Panax pseudoginseng*, and can be used to identify *Panax notoginseng*.

Cited as: Pham, D. T., & Nguyen, P. V. (2024). Identification of sanchi samples based on DNA barcodes. *The Journal of Agriculture and Development* 23(3), 1-10.

1. Introduction

Sanchi (*Panax notoginseng* (Burkill) F. H. Chen) is a precious medicinal herb belonging to the Araliaceae family. It contains valuable compounds such as saponins, polyacetylenes, and essential amino acids that can inhibit and destroy tumor cells, exhibit antibacterial activity, and promote human health. *Panax notoginseng* is a well-known and widely used herbal medication in Asian nations. Compared with other species of the genus *Panax*, it has unique chemical components and medicinal value. Thanks to its potential therapeutic benefits not only for blood illnesses but also for other types of chronic human ailments, *Panax notoginseng* has attracted a lot of attention and interest (Xu et al., 2018). The morphological similarities between *Panax notoginseng* and *Panax pseudoginseng* have led to confusion in their nomenclature. More importantly, sanchi rhizomes can be confused with other medicinal rhizomes (Huynh et al.,

2021). Therefore, an effective authentication approach for medicinal plants, including sanchi, is essential for developing the herbal medicine industry and for protection from adulterants. Traditionally, authentication of herbs relies on morphological and histological inspections. In many cases, such as in the authentication of different *Panax* species, this approach is far from reliable (Ngan et al., 1999). Recently, molecular biology techniques have been applied to identify and analyze genetic diversity through molecular taxonomy. DNA barcoding, introduced in the mid-1990s, has been utilized to classify species within the Panax genus. Barcodes used located in the nuclear genome such as the internal trancribed spacer (ITS) region, 18S-rRNA; in mitochondria such as nad1 or chloroplast genome such as *matK*, *psbA-trnH*, *psbK-I*, *pspM-trnD*, *rps16*, *trnC-trnD*. The ITS and *psbA-trnH* regions showed more single-nucleotide polymorphisms and could be used for species identification and taxonomy of the *Panax* genus (Zuo et al., 2011). In 2017, Trang et al. used four chloroplast DNA regions, including *matK*, *rbcL*, and *rpoB*, and one nuclear DNA region ITS for authentic *Panax vietnamensis*. The results showed that *matK* and *rpoB* were suitable for identification at the species and subspecies levels. Although not all scientists have agreed upon universal plant DNA barcodes, most applications use the standard markers, *rbcL*, *matK*, *trnH-psbA*, and ITS (Yu et al., 2021). In this study, four DNA barcodes were used to identify and distinguish sanchi from potential substitutes or adulterants.

2. Materials and Methods

2.1. Sample collection

Sanchi plantlets were collected from Lam Dong province and identified based on the basis of their morphology.

2.2. DNA extraction

Total DNA from leaf samples was extracted according to the protocol described by Aboul-Maaty & Oraby (2019). 100 mg of chopped leaf samples were pureed with buffer (3% CTAB (w/v), 1.4 M NaCl, 0.8 M Tris-HCl pH 8.0, 0.5 M EDTA pH 8.0; 0.3% 2-β-mercaptoethanol). The mixture was incubated for 60 min at 65°C and denatured with chloroform: isoamyl alcohol (C: I, 24: 1). Centrifuge 13,000 rpm for 15 min at room temperature, collect the supernatant and add 0.5 V 6M NaCl, 0.1 V potassium acetate. DNA was precipitated with isopropanol for 60 min at -4°C. The precipitate was centrifuged, and the DNA was washed with cooled 70% ethanol, collect and store the DNA at -20°C. After extraction, the DNA quality was examined by electrophoresis on 1% agarose gel, and a spectrometer (Nanodrop).

2.3. PCR and DNA sequencing

The composition of PCR to amplify four DNA barcodes (rbcL, matK, trnH-psbA, and ITS1) is as follows: 8.6 L Mastermix (Thermo Scientific), 50 ng DNA, 0.2 M of each primer (rbcL-F: GAC AAC TGT GTG GAC CGA TG, rbcL-R: CCA CCG CGA AGA CAT TCA TA) (Kress & Erickson, 2007); (1R_KIM-f: ACC CAG TCC ATC TGG AAA TCT TGG TTC, 3F_KIM-r: CGT ACA GTA CTT TTG TGT TTA CGA G) (Kim, unpublished); (trnHf: CGC GCA TGG TGG ATT CAC AAT CC, psbA3'f: GTT ATG CAT GAA CGT AAT GCT C) (Kress et al., 2005); (ITS-p5: CCT TAT CAY TTA GAG GAA GGAG, ITS-p4: CCG CTT AKT GAT ATG CTT AAA) (Cheng et al., 2016), and PCR water for a final volume of 20 L. PCR reactions were carried out under the following conditions: initial denaturation at 94°C for 5 min; 35 cycles of 30 s at 94°C, 30 s at Ta°C, 60 s at 72°C, and finally 7 min at 72°C to complete the reaction.

The PCR products were visualized by 1% gel electrophoresis using a 1 kb ladder (Bioline, UK) to estimate the amplified length. The correct PCR products were purified and bidirectionally sequenced by Sanger methods at the 1st BASE (Malaysia).

2.4. Phylogenetic analysis

The nucleotide sequences of both DNA strands were verified and edited to obtain a consensus sequence using the BioEdit Sequence Alignment Editor and Chromas 2.6.6. The assembled sequences were aligned to the available sequences in the NCBI GenBank database to obtain accession numbers and used for subsequent analysis. The phylogenetic tree was constructed using Molecular Evolution Genetic Analysis (MEGA X) with the maximum likelihood algorithm from DNA sequences with a bootstrap value set at 1,000 replicates (Kumar et al., 2018).

3. Results and Discussion

3.1. Morphology

The rhizome is radish-like in shape, with fibrous strands around it. The stem was upright, unbranched, and smooth with a spongy core. Compound leaves consisted of verticillate leaves arranged around the top of the stem. The leaflets were serrated; the leaf blade was elliptical, thin, biconvex, and green; the tip had a pointed tail; and the base was wedge-shaped, narrow, and long. Scattered trichomes were observed in the veins of both leaf surfaces (Figure 1).

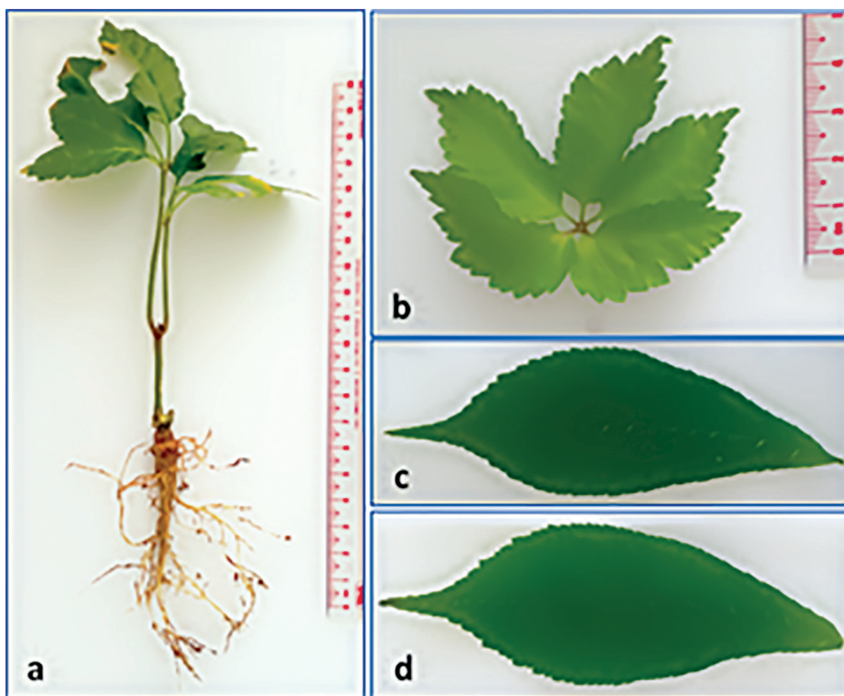

Figure 1. Sanchi sample. *a) Sanchi six months old; b) composite leaf; c-d) adaxial and abaxial sides.*

3.2. Amplify four regions of the DNA barcodes

The total DNA of the analyzed samples was extracted, quantified, and examined for purity. The results showed that the samples were purified with an OD 260/280 ratio of 1.89 - 1.97 and a concentration in the range of 48.8 ng/L - 380.1 ng/L. The finished product was clear and intact, and confirming its suitability for use in the PCR reaction (electrophoresis results are not shown). The results of the amplification of four DNA barcodes showed that the approximate product size was 400 bp for rbcL, 900 bp for matK, 500 bp for trnH - psbA, and in ITS1, which was close to the expected sequence. The PCR products were obtained, purified, and sequenced in two directions.

3.3. Nucleotide sequence analysis

Sequencing results were successful for all gene regions. The sizes of the products were 330 bp (*rbcL*), 822 bp (*matK*), 484 bp (*trnH-psbA*), and 438 bp (ITS1). BLAST results from NCBI showed 100% coverage and 99.7% to 100% similarity between the studied sequences and the corresponding *rbcL* gene sequences of *Panax* species; in the *matK* gene region were 99% and 93.79 - 93.30%, respectively; in the *trnH-psbA* gene region, 99 - 100% and 95.93 - 99.59%, respectively; and in the ITS1 gene region were 100% and 96.81% - 99.55%, respectively.

The studied *rbcL* region indicated a high similarity, reaching 100% similarity to sequences of all *Panax* species. The *rbcL* region is considered a universal highly conserved exhibiting and slow rate of evolution; therefore, it would only be used for classification at the genus level. The *matK* and *trnH-psbA* regions were highly similar to those of *P. notoginseng* (93.79% and 99.59%, respectively). The alignment results showed that the ITS1 sequence had the highest similarity (99.55%) to *P. notoginseng*, whereas *P. zingiberensis* showed the greatest difference (96.81%).

According to the rectification results, the research sample's *trnH - psbA* gene region contained a high similarity nucleotide to *P. notoginseng* (KP036468.1, South Korea), 3 different nucleotide positions and 5 gaps compared to *Panax* sp. 'sinensis', *P. quinquefolius*, *P. ginseng*, 7 different nucleotide positions from *P. wangianus*, *P. major*, *P. pseudoginseng* var. *elegantior*, *P. japonicus* var. *bipinnatifidus* and 9 - 10 nucleotide positions in *P. zingiberensis*, *P. vietnamensis* var. *fuscidiscus*, *P. vietnamensis*, *P. vietnamensis* var. *langbianesis* (Table 1).

In the ITS1 region, the research sample had a high nucleotide similarity to *P. notoginseng* (KP036468.1, South Korea) in the *trnH - psbA* region. There were nine nucleotide positions different from *P. wangianus*, *P. quinquefolius*, *P. japonicus* var. *japonicus*, *P.* sp 'sinensis, *P. ginseng*, *P. pseudoginseng* var. *bipinnatifidus*, *P. vietnamensis* and 10 - 12 positions different from *P. zingiberensis*, *P. japonicus*, *P. japonicus* var. *bipinnatifidus*, *P. variabilis*, *P. major*, respectively (Table 2). The diversity of nucleotides in the group of sanchi studied in Vietnam with other *Panax* species can be observed.

Table 1. Variable sites in *trnH - psbA* sequences from different *Panax* species

Species, Accession number / Nucleotide positions	140	147	188	232	235	250	271	272	276	280	281	289	389	405
Research sample	A	C	C	T	A	T	T	T	T	T	T	A	C	C
P. notoginseng KP036468.1														
P. wangianus MK408934.1		T		C		G	A	A				G	A	
P. major MW654097.1		T					A	A	A	A	A	G		
P. pseudoginseng var. *elegantior* NC062082.1		T					A	A	A	A	A	G		
P. japonicus var. *bipinnatifidus* OL543605.1		T					A	A	A	A	A	G		
P. zingiberensis MK408969.1		T		C		G	A	A	A	A	A	G		
P. vietnamensis var. *fuscidiscus* MT798587.1		T		C		G	A	A	A	A	A	G		
P. vietnamensis MF377623.1		T		C		G	A	A	A	A	A	G		T
P. vietnamensis var. *langbianensis* MT798584.1		T		C		G	A	A	A	A	A	G		T
Panax sp. 'sinensis' MK408967.1		T	A									G		
P. quinquefolius MK408953.1	T				G							G		
P. ginseng OL543607.1	T				G							G		

Table 2. Variable sites in internal transcribed spacer 1 sequences from different *Panax* species

Species, Accession number / Nucleotide positions	48	69	76	79	91	102	104	130	136	138	150	151	201	209	210	214	349
Studied sample	C	A	A	C	T	T	G	C		G	C	C	T	C	A	G	T
P. notoginseng MK408810.1	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-
P. wangianus MK408801.1	T	C	C		C	C	A			A			C				C
P. quinquefolius MK408799.1	T	C			C	C	A			A		T	C				C
P. japonicus var. *japonicus* MH345184.1	T	C			C	C	A			A		T	C				C
P. sp. '*sinensis*' MK408796.1	T	C	C		C	C	A			A		T	C				C
P. ginseng KM207668.1	T	C			C	C	A			A	A	T	C				C
P. pseudoginseng var. *bipinnatifidus* AY271912.1	T	T		T	C	C	A					T	C			A	C
P. zingiberensis MH345201.1	T	C	C		C	C	A	Y	Y	R		T	C				C
P. vietnamensis KT380922.1	T	C			C	C	A					T	C		G		C
P. japonicus KJ740652.1		C	C		C	C	A			A		T	C		G		C
P. japonicus var. *bipinnatifidus* MZ149947.1	T	C	C		C	C	A			A		T	C				C
P. variabilis AY233330.1	T	C	C		C	C	A			A			C	T			C
P. major MH345179.1	T	C	C		C	C	A			A		T	C				C
P. japonicus var. *angustifolius* MH345153.1	T	C	C	T	C	C	A			A			C				C

3.4. Phylogenetic relationship

Phylogenetic trees of the studied samples and available sequences of *Panax* in GenBank were constructed using the MEGA X software with the Maximum Likelihood algorithm, and the bootstrap coefficient was set to 1.000. Results presented in Figures 2 - 5 showed that all species belonged to the same genus, *Panax*, and formed a distinct evolutionary clade. Additionally, they were found to be closely related to each other, with varying levels of support from bootstrap analysis. The clustering scores ranged from 64% for the *rbcL* gene region (Figure 2), 64% to 87% for the *matK* gene region (Figure 3), 52% to 96% for *trnH-psbA* (Figure 4), and 29% to 99% for ITS1 (Figure 5). Particularly for the *matK* gene region, there was clear branching of the research sequence group with 14 sequences of *Panax* species in GenBank (Figure 3). The phylogenetic tree also showed that the studied sequence group formed a separate clade with *P. notoginseng* compared to the *Panax* species in two gene regions: *trnH-psbA* and ITS1. These results show that the *trnH-psbA* and ITS1 DNA barcodes can be used to identify *Panax notoginseng*.

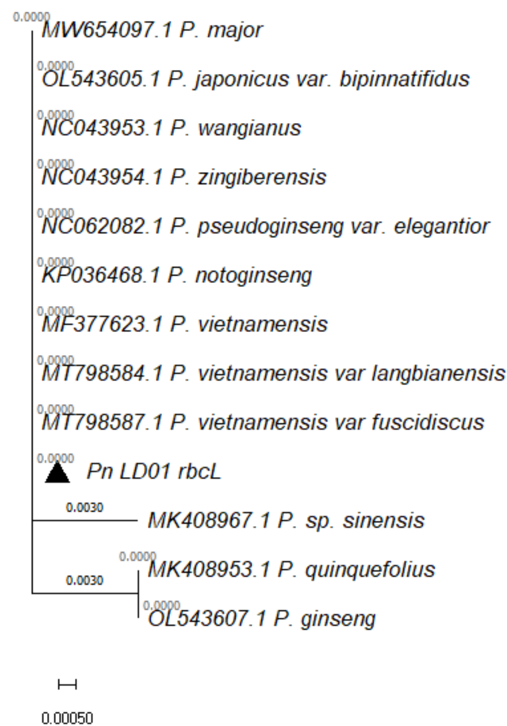

Figure 2. Phylogenetic relationship based on the *rbcL* sequence of the studied sample (Pn-LD01 *rbcL*) and 12 reference sequences with the Maximum Likelihood algorithm.

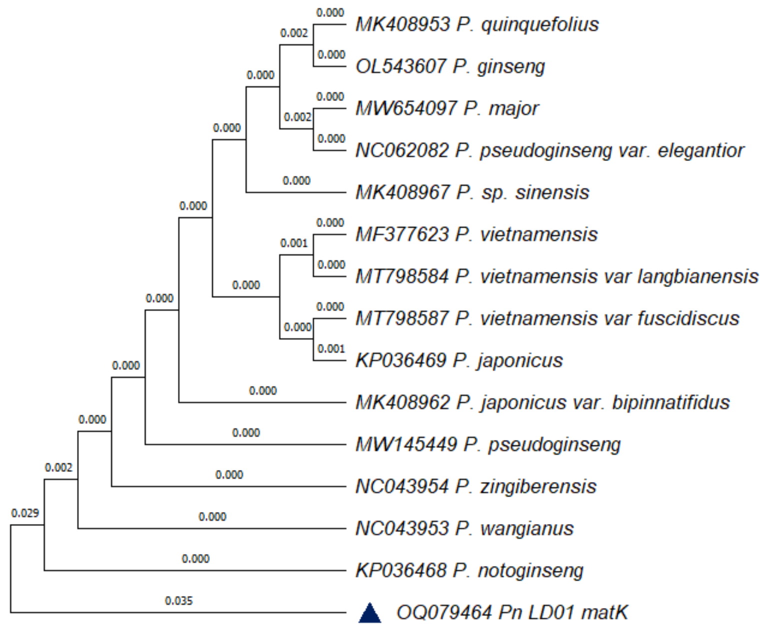

Figure 3. Phylogenetic relationship based on the *matK* sequence of the studied sample (Pn-LD01 *matK*) and 14 reference sequences with the Maximum Likelihood algorithm.

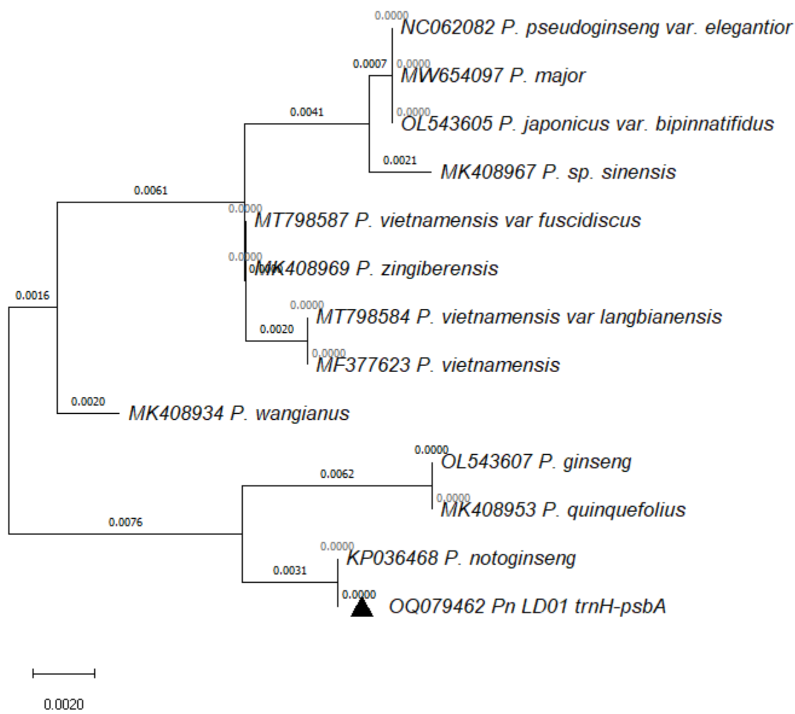

Figure 4. Phylogenetic relationship based on the *trnH-psbA* sequence of the studied sample (Pn-LD01 *trnH-psbA*) and 12 reference sequences with the Maximum Likelihood algorithm.

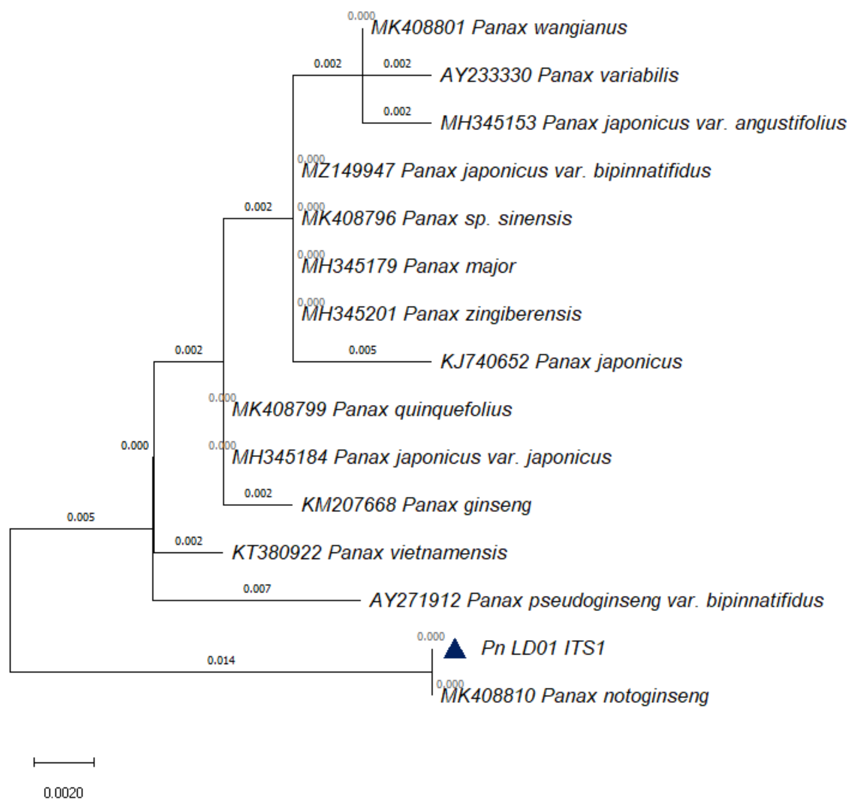

Figure 5. Phylogenetic relationship based on the internal transcribed spacer (ITS) 1 sequence of the studied sample (Pn-LD01 ITS1) and 14 reference sequences with the maximum likelihood algorithm.

The results proved the effectiveness of DNA barcodes in assessing the genetic relationships and species identification of *Panax*. The *trnH-psbA* and ITS1 sequences have almost identical nucleotide sequences in terms of the number of distinct nucleotides. However, compared with the *trnH-psbA* gene sequence, ITS1 displayed a larger differential nucleotide diversity. The *matK* region showed the greatest nucleotide difference. The four DNA barcodes displayed varying levels of differentiation, ranked from high to low as follows: *matK* > ITS1 > *trnH-psbA* > *rbcL*. This finding is consistent with the CBOL report on DNA barcoding in plants in 2009. These results provide a valuable database for further studies on this species.

4. Conclusions

The study successfully established a PCR procedure and sequenced four DNA barcode regions of the sanchi samples. The *matK*, *trnH-psbA*, and ITS1 sequences confirmed that the sanchi sample under study belonged to *Panax notoginseng*, distinguishing it from *Panax pseudoginseng* and other *Panax* species. These findings could be applied in the identification and discrimination of sanchi from adulterants.

Conflict of interest

The authors declare that they have no competing interests.

Acknowledgements

The authors gratefully acknowledge the support and encouragement of the Management Board of Nong Lam University (NLU). This research was funded by NLU under Grant No. CS-CB21-CNSH-01.

References

Aboul-Maaty, N. A. F., & Oraby, H. A. S. (2019). Extraction of high-quality genomic DNA from different plant orders applying a modified CTAB-based method. *Bulletin of the National Research Centre* 43(1), 1-10. https://doi.org/10.1186/s42269-019-0066-1.

Cheng, T., Xu, C., Lei, L., Li, C., Zhang, Y., & Zhou, S. (2016). Barcoding the kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and specificity. *Molecular Ecology Resources* 16(1), 138-149. https://doi.org/10.1111/1755-0998.12438.

Huynh T. N., N. P. M., & Ho T. V. (2021). Design PCR primer for distinguishing between *Panax stipuleanatus* and *Panax notoginseng*. *Journal of Science, Technology, and Food* 21(2), 73-79.

Kress, W. J., & Erickson, D. L. (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. https://doi.org/10.1371/journal.pone.0000508.

Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A., & Janzen, D. H. (2005). Use of DNA barcodes to identify flowering plants. *Proceedings of The National Academy of Sciences* 102(23), 8369-8374. https://doi.org/10.1073/pnas.0503123102.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35(6), 1547-1549. https://doi.org/10.1093/molbev/msy096.

Ngan, F., Shaw, P., But, P., & Wang, J. (1999). Molecular authentication of *Panax* species. *Phytochemistry* 50(5), 787-791. https://doi.org/10.1016/S0031-9422(98)00606-2.

Trang, N. T., Mai, N. T., Zhuravlev, Y. N., & Sciences. (2017). Application of DNA barcoding to authentic *Panax vietnamensis*. *American Scientific Research Journal for Engineering, Technology, and Science* 29(1), 60-67.

Xu, Y., Tan, H. Y., Li, S., Wang, N., & Feng, Y. (2018). *Panax notoginseng* for inflammation-related chronic diseases: a review on the modulations of multiple pathways. *The American Journal of Chinese Medicine* 46(05), 971-996. https://doi.org/10.1142/S0192415X18500519.

Yu, J., Wu, X., Liu, C., Newmaster, S., Ragupathy, S., & Kress, W. J. (2021). Progress in the use of DNA barcodes in the identification and classification of medicinal plants. *Ecotoxicology and Environmental Safety* 208, 111691. https://doi.org/10.1016/j.ecoenv.2020.111691.

Zuo, Y., Chen, Z., Kondo, K., Funamoto, T., Wen, J., & Zhou, S. (2011). DNA barcoding of *Panax* species. *Planta Medica* 77(2), 182-187. https://doi.org/10.1055/s-0030-1250166.