A DISSERTATION FOR

THE DEGREE OF DOCTOR OF PHILOSOPHY

DNA Super-barcoding and Authentication of Species in Araliaceae Family

By

VO NGOC LINH GIANG

FEBRUARY, 2022

MAJOR IN CROP SCIENCE AND BIOTECHNOLOGY

DEPARTMENT OF PLANT SCIENCE

COLLEGE OF AGRICULTURE AND LIFE SCIENCES

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

DNA Super-barcoding and Authentication of Species in Araliaceae Family

UNDER THE DIRECTION OF DR. TAE-JIN YANG

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

BY

VO NGOC LINH GIANG

MAJOR IN CROP SCIENCE AND BIOTECHNOLOGY

DEPARTMENT OF PLANT SCIENCE

FEBRUARY, 2022

APPROVED AS QUALIFIED DISSERTATION OF VO NGOC LINH GIANG

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

BY THE COMMITTEE MEMBERS

CHAIRMAN	Nam-Chon Paek, Ph.D.
VICE-CHAIRMAN	Tae-Jin Yang, Ph.D.
MEMBER	Do-Soon Kim, Ph.D.
MEMBER	Jeong Hill Park, Ph.D.
MEMBER	Van Binh Nguyen, Ph.D.

DNA Super-barcoding and Authentication of Species in Araliaceae Family

VO NGOC LINH GIANG

Department of Plant Science

The Graduate School of Seoul National University

GENERAL ABSTRACT

The Araliaceae family includes 50 genera and around 1500 species worldwide. Overharvesting and adulteration of high-value medicinal plants from this family, particularly *Panax* species, have caused trouble in supply systems. The whole plastid genome (plastome) and 45S nuclear ribosomal DNA (nrDNA) sequences, referred to as super-barcodes, have emerged as a promising approach for correctly identifying plant species, a vital first step in the authentication system for combating economically motivated adulteration (EMA). This dissertation was conducted to assess the efficiency of these super-barcodes in classifying Araliaceae species and identifying specific barcoding loci for *Panax* species authentication.

Eleven newly characterized complete plastomes and 64 previously reported complete plastomes were included, representing 56 species from 26 genera in the Araliaceae family and *Centella asiatica* (Apiaceae). The maximum likelihood (ML) tree was constructed from 75 plastome sequences and showed the monophyly of 25 genera with 100% bootstrap support. The genus Aralia was polyphyletic. The relaxed molecular clock estimation based

on the 78 plastid genes indicated that subsequent speciation occurred within the Araliaceae genera from 8.04 to 0.67 million years ago (MYA) during the Middle Miocene and Pliocene. The collection of 74 complete plastomes revealed 43,458 variants. The 45S nrDNA sequences of 22 Araliaceae species showed 97.5 to 100% identities with 383 variants. Critical examination of plastome and 45S nrDNA super-barcoding data would provide new insights to improve knowledge of the evolutionary history of Araliaceae species. Furthermore, based on the phylogenetic tree, a total of fifteen SNPs in the plastid coding gene sequences were selected and successfully developed Kompetitive allele-specific PCR (KASP) markers, enabling clear identification among three genera: *Panax, Aralia,* and *Eleutherococcus*.

Panax species native to Vietnam are designated as endangered species due to illegal harvesting and EMA from related species. Conservation of *Panax* species in Vietnam is hindered by a lack of genetic resources and molecular techniques. Seven *P. vietnamensis* individuals, representing three varieties: var *vietnamensis,* var. *fuscidiscus*, and var. *langbianesis,* were collected from Vietnamese habitats to generate NGS data for assemblies of whole plastome and 45S nrDNA sequences. The ML phylogenetic trees of the complete plastomes and 45S nrDNA indicated nine major *Panax* clades. The plastomes of *P. vietnamensis* var. *vietnamensis* and var. *langbianesis* were almost identical and clustered within the *P. vietnamensis* clade. However, *P. vietnamensis* var. *fuscidiscus* formed a distinct clade with *P. zingiberensis*, which was sister to the clade of *P. vietnamensis*. Significant intraspecific variation and phylogenetic incongruence within the *P. japonicus* clade and within the *P. vietnamensis* var. *fuscidiscus* and *P. zingiberensis* clade suggested an ambiguous species delimitation. Sixteen *Panax*-specific SNPs were developed as KASP markers, a high throughput SNP genotyping platform, which were highly accurate in the authentication of 114 *Panax*

individuals collected from Vietnam.

In conclusion, the two investigations demonstrated the effectiveness of the DNA super-barcoding approach through the use of whole plastome and 45S nrDNA sequences. The results of this study widened the field of phylogenetic analysis and species identification in the Araliaceae family. Furthermore, the markers developed through the analysis of *Panax* species entire plastome and 45S nrDNA sequences will aid in the detection of counterfeit ginseng in Vietnam's ginseng industry.

Keywords: Araliaceae, *Panax* species, Next-generation sequencing (NGS), Plastomes, 45S nrDNA.

Student Number: 2017-35949

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

InDel	Insertion or Deletion
CDS	Coding sequence
dCAPS	Derived cleaved amplified polymorphic sequences
dnaLCW	*de novo* assembly using low coverage of WGS
EMA	Economically motivated adulteration
nDNA	Nuclear DNA
45S nrDNA	45S nuclear ribosomal DNA
NOR	Nucleolus organizer regions
NGS	Next-generation sequencing
PCR	Polymerase chain reaction
rDNA	Ribosomal DNA
RE	Restriction Enzyme
SNP	Single nucleotide polymorphism
WGS	Whole-genome sequence

GENERAL INTRODUCTION

The Araliaceae family (also known as the ginseng family) belongs to the order Apiales and is the sister family of the Apiaceae family (Chase et al., 2016; Plunkett et al., 2004). The family comprises approximately 1500 species, which are divided into 41~55 genera. Araliaceae species are found across the tropics and subtropics of Asia, South and Central America, Oceania, and sub-Saharan Africa-Madagascar (Liu et al., 2012; Mitchell et al., 2012; Plunkett et al., 2004; Valcárcel et al., 2014; Valcárcel & Wen, 2019). The Araliaceae family has a wide range of adaptogenic medicinal plants, such as *Panax*, *Aralia*, and *Eleutherococcus* (Clement & Clement, 2014; Davydov & Krikorian, 2000; Irfan et al., 2020; Panossian et al., 2021). In recent years, due to increased globalization of trade, accompanied with financial gain, medicinal plants in the Araliaceae family have been frequently adulterated, especially *Panax* or ginseng. Price differentials between the major species of *Panax* are the primary motivator for intentional adulteration (Ichim & de Boer, 2020). Economically motivated adulteration (EMA) through the use of low-cost varieties, alternative species, or nothing at all may pose health concerns to consumers. It is necessary to establish an effective and practical authentication system that utilizes molecular markers to identify the principal *Panax* species in effort to detect potential adulteration.

The three plant genomes, nuclear, mitochondrial, and chloroplast (plastids), reside in distinct cellular compartments and exhibit distinct evolutionary patterns. Plastid genomes (plastomes) and nuclear DNA regions encoding ribosomal RNA components (45S nrDNA) are highly conserved structures with a large copy number, and they are particularly important for better understanding the evolution and genetic diversity in plants (Bobik & Burch-Smith, 2015; Daniell et al., 2016; Gitzendanner et al., 2018; Soltis et

al., 1999). The plastome of photosynthetically active land plants is mostly maternally inherited, and it has a conserved pattern of approximately 120 genes that are densely packed onto a circular DNA molecule of 120–220 kb in size. The mapping plastome is typically depicted as a quadripartite structure, consisting of a ring of four major segments: large and small single-copy (LSC and SSC) regions connected by a pair of inverted repeats (IRs) (Bock, 2007; Daniell et al., 2016; Greiner et al., 2015; Palmer, 1985, 1991; Wakasugi et al., 2001). The nuclear genomes of plants contain hundred to thousand copies of a tandem array of the 45S nrDNAs. Tandem repeats of 45S nrDNA present in the nucleolus organizer regions (NOR) of plant chromosomes and separated by an intergenic spacer (IGS) (De Winter & Moss, 1986; Hamby, 1990; Labhart & Reeder, 1986; Sáez-Vásquez & Delseny, 2019). Each unit comprises three cistrons that codes for 26S (or 28S), 18S, and 5.8S rRNA genes. With the presence of two internal transcribed spacers (ITS1 and ITS2), five segments finally come together to create a 5′-18S-ITS1-5.8S-ITS2-26S-3′ unit, which is frequently denoted by 45S nrDNA. The ITS regions are the most divergent regions in the 45S nrDNA and have been widely used in phylogenetic studies (Hamby, 1990).

In recent years, advances in DNA sequencing technology, specifically the next-generation sequencing (NGS) platform, have enabled the rapid and cost-effective sequencing of whole genomes at high throughput. Massive amounts of NGS data are created, requiring data mining techniques to extract valuable information for downstream studies. The method developed in our laboratory for *de novo* assembly of low coverage WGS (dnaLCW) (Kim et al., 2015b) has been extensively used in worldwide researches to analyze raw next-generation sequencing data in order to retrieve plastome and 45S nrDNA sequences. These sequences are regarded as super-barcodes since they

comprise plenty of the genomic information to distinguish between the species of interest (Dong et al., 2021; Li et al., 2015; Ren et al., 2021; C. Wu et al., 2021; L. Wu et al., 2021). Additionally, recent researches have demonstrated that these super-barcodes also have a high discriminatory capacity in Araliaceae species (Kim et al., 2015a; Kim et al., 2017; Kim et al., 2018; Nguyen et al., 2018; Valcárcel et al., 2014; Valcárcel & Wen, 2019; Wen & Zimmer, 1996). There was a vast amount of Araliaceae species genetic data accessible, but the data came from various sources and narrowly focused on several genera. Thus, the phylogenetic relationships of the Araliaceae are poorly understood.

By adding more whole plastome and 45S nrDNA sequences from Araliaceae species, this research focused on the concept of "super-barcoding", which opened up new platforms for molecular plant identification. In this study, the initial objective was to conduct a comprehensive analysis of phylogenetic relationships and genetic diversity among Araliaceae species by analyzing the DNA super-barcoding data. The second purpose was to investigate and to assess the Single Nucleotide Polymorphisms (SNPs) associated with plastome genes that could be switched to a high-throughput automated assay, such as the Kompetitive allele-specific polymerase chain reaction (PCR) (KASP) technology(LGC, Queens Road, Teddington, Middlesex, UK). A number of KASP assays were designed with the aim of diagnosing Araliaceae-specific genotypes efficiently. These findings will provide the groundwork for future research on genetic variation in Araliaceae species. More precisely, these KASP assays can differentiate between *Panax* species in Vietnam, which is necessary for properly managing and protecting these endangered species.

REFERENCES

Bobik, K., & Burch-Smith, T. M. (2015). Chloroplast signaling within, between and beyond cells [Review]. *Frontiers in Plant Science*, *6*, 781. https://doi.org/10.3389/fpls.2015.00781

Bock, R. (2007). Structure, function, and inheritance of plastid genomes. In *Cell and molecular biology of plastids* (pp. 29-63). Springer.

Chase, M. W., Christenhusz, M., Fay, M., Byng, J., Judd, W. S., Soltis, D., Mabberley, D., Sennikov, A., Soltis, P. S., & Stevens, P. F. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, *181*(1), 1-20.

Clement, J., & Clement, E. (2014). The medicinal chemistry of genus *Aralia*. *Current topics in medicinal chemistry*, *14*(24), 2783-2801.

Daniell, H., Lin, C.-S., Yu, M., & Chang, W.-J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biology*, *17*(1), 134. https://doi.org/10.1186/s13059-016-1004-2

Davydov, M., & Krikorian, A. D. (2000). *Eleutherococcus senticosus* (Rupr. & Maxim.) Maxim.(Araliaceae) as an adaptogen: a closer look. *Journal of ethnopharmacology*, *72*(3), 345-393.

De Winter, R. F., & Moss, T. (1986). The ribosomal spacer in *Xenopus laevis* is transcribed as part of the primary ribosomal RNA. *Nucleic acids research*, *14*(15), 6041-6052.

Dong, W., Sun, J., Liu, Y., Xu, C., Wang, Y., Suo, Z., Zhou, S., Zhang, Z., & Wen, J. (2021). Phylogenomic relationships and species identification of the olive genus *Olea* (Oleaceae). *Journal of Systematics and Evolution*.

Gitzendanner, M. A., Soltis, P. S., Wong, G. K. S., Ruhfel, B. R., & Soltis, D.

E. (2018). Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. *American Journal of Botany*, *105*(3), 291-301.

Greiner, S., Sobanski, J., & Bock, R. (2015). Why are most organelle genomes transmitted maternally? *Bioessays*, *37*(1), 80-94.

Hamby, R. K. (1990). *Ribosomal RNA and the early evolution of flowering plants* Louisiana State University and Agricultural & Mechanical College].

Ichim, M. C., & de Boer, H. J. (2020). A review of authenticity and authentication of commercial ginseng herbal medicines and food supplements. *Frontiers in Pharmacology*, *11*, 2185.

Irfan, M., Kwak, Y.-S., Han, C.-K., Hyun, S. H., & Rhee, M. H. (2020). Adaptogenic effects of *Panax ginseng* on modulation of cardiovascular functions. *Journal of Ginseng Research*, *44*(4), 538-543.

Kim, K., Lee, S.-C., Lee, J., Lee, H. O., Joh, H. J., Kim, N.-H., Park, H.-S., & Yang, T.-J. (2015a). Comprehensive survey of genetic diversity in chloroplast genomes and 45S nrDNAs within *Panax ginseng* species. *PloS one*, *10*(6), e0117159. https://doi.org/https://doi.org/10.1371/journal.pone.0117159

Kim, K., Lee, S.-C., Lee, J., Yu, Y., Yang, K., Choi, B.-S., Koh, H.-J., Waminal, N. E., Choi, H.-I., & Kim, N.-H. (2015b). Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of *Oryza* AA genome species. *Scientific Reports*, *5*, 15655.

Kim, K., Nguyen, V. B., Dong, J., Wang, Y., Park, J. Y., Lee, S.-C., & Yang, T.-J. (2017). Evolution of the Araliaceae family inferred from complete chloroplast genomes and 45S nrDNAs of 10 *Panax*-related species. *Scientific Reports*, *7*(1), 4917. https://doi.org/10.1038/s41598-017-05218-y

Kim, N.-H., Jayakodi, M., Lee, S.-C., Choi, B.-S., Jang, W., Lee, J., Kim, H. H., Waminal, N. E., Lakshmanan, M., van Nguyen, B., Lee, Y. S., Park, H.-S., Koo, H. J., Park, J. Y., Perumal, S., Joh, H. J., Lee, H., Kim, J., Kim, I. S., . . . Yang, T.-J. (2018). Genome and evolution of the shade-requiring medicinal herb *Panax ginseng*. *Plant Biotechnology Journal*, *16*(11), 1904-1917. https://doi.org/10.1111/pbi.12926

Labhart, P., & Reeder, R. H. (1986). Characterization of three sites of RNA 3′ end formation in the Xenopus ribosomal gene spacer. *Cell*, *45*(3), 431-443.

Li, X., Yang, Y., Henry, R. J., Rossetto, M., Wang, Y., & Chen, S. (2015). Plant DNA barcoding: from gene to genome. *Biological Reviews*, *90*(1), 157-166.

Liu, Z., Zeng, X., Yang, D., Chu, G., Yuan, Z., & Chen, S. (2012). Applying DNA barcodes for identification of plant species in the family Araliaceae. *Gene*, *499*(1), 76-80.

Mitchell, A., Li, R., Brown, J. W., Schönberger, I., & Wen, J. (2012). Ancient divergence and biogeography of *Raukaua* (Araliaceae) and close relatives in the southern hemisphere. *Australian Systematic Botany*, *25*(6), 432-446.

Nguyen, V. B., Giang, V. N. L., Waminal, N. E., Park, H.-S., Kim, N.-H., Jang, W., Lee, J., & Yang, T.-J. (2018). Comprehensive comparative analysis of chloroplast genomes from seven *Panax* species and development of an authentication system based on species-unique single nucleotide polymorphism markers. *Journal of Ginseng Research*. https://doi.org/https://doi.org/10.1016/j.jgr.2018.06.003

Palmer, J. D. (1985). Comparative organization of chloroplast genomes. *Annual review of genetics*, *19*(1), 325-354.

Palmer, J. D. (1991). Plastid chromosomes: structure and evolution. *The*

molecular biology of plastids, *7*, 5-53.

Panossian, A. G., Efferth, T., Shikov, A. N., Pozharitskaya, O. N., Kuchta, K., Mukherjee, P. K., Banerjee, S., Heinrich, M., Wu, W., & Guo, D. a. (2021). Evolution of the adaptogenic concept from traditional use to medical systems: Pharmacology of stress-and aging-related diseases. *Medicinal Research Reviews*, *41*(1), 630-703.

Plunkett, G. M., Chandler, G. T., Lowry II, P., Pinney, S., Sprenkle, T., Van Wyk, B.-E., & Tilney, P. (2004). Recent advances in understanding Apiales and a revised classification. *South African Journal of Botany*, *70*(3), 371-381.

Ren, F., Wang, L., Li, Y., Zhuo, W., Xu, Z., Guo, H., Liu, Y., Gao, R., & Song, J. (2021). Highly variable chloroplast genome from two endangered Papaveraceae lithophytes *Corydalis tomentella* and *Corydalis saxicola*. *Ecology and Evolution*, *11*(9), 4158-4171.

Sáez-Vásquez, J., & Delseny, M. (2019). Ribosome biogenesis in plants: from functional 45S ribosomal DNA organization to ribosome assembly factors. *The Plant Cell*, *31*(9), 1945-1967.

Soltis, P. S., Soltis, D. E., & Chase, M. W. (1999). Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature*, *402*(6760), 402-404.

Valcárcel, V., Fiz-Palacios, O., & Wen, J. (2014). The origin of the early differentiation of Ivies (*Hedera* L.) and the radiation of the Asian Palmate group (Araliaceae). *Molecular phylogenetics and evolution*, *70*, 492-503.

Valcárcel, V., & Wen, J. (2019). Chloroplast phylogenomic data support Eocene amphi-Pacific early radiation for the Asian Palmate core Araliaceae. *Journal of Systematics and Evolution*, *57*(6), 547-560.

Wakasugi, T., Tsudzuki, T., & Sugiura, M. (2001). The genomics of land plant

chloroplasts: gene content and alteration of genomic information by RNA editing. *Photosynthesis Research*, *70*(1), 107-118.

Wen, J., & Zimmer, E. A. (1996). Phylogeny and biogeography of *Panax* L.(the ginseng genus, Araliaceae): inferences from ITS sequences of nuclear ribosomal DNA. *Molecular phylogenetics and evolution*, *6*(2), 167-177.

Wu, C., Sudianto, E., Chiu, H., Chao, C., & Chaw, S. (2021). Reassessing Banana Phylogeny and Organelle Inheritance Modes Using Genome Skimming Data. *Frontiers in Plant Science*, 1661.

Wu, L., Wu, M., Cui, N., Xiang, L., Li, Y., Li, X., & Chen, S. (2021). Plant super-barcode: a case study on genome-based identification for closely related species of *Fritillaria*. *Chinese medicine*, *16*(1), 1-11.

CHAPTER I

Exploration of DNA super-barcoding for the Araliaceae family

ABSTRACT

Araliaceae species are notable for their highly diverse structures of flowers and issues on their classification. The ginseng family (Araliaceae) would be an ideal model group to reflect the impact of molecular techniques in resolving the issues of authentication and classification. This study successfully assembled the new complete plastid genomes (plastomes) and 45S nuclear ribosomal DNA (45S nrDNA) regions of 11 species from nine Araliaceae genera using WGS data provided by the Illumina MiSeq platform. Total 76 plastomes from 26 genera and 22 45S nrDNA sequences of 11 genera were employed to evaluate the effectiveness of these super-barcode sequences in classifying genera in the Araliaceae. The Araliaceae plastomes were determined from 152,880 bp to 156,939 bp in length and presented the same quadripartite architecture with common 80 protein-coding, four rRNA, and 30 tRNA genes. The variant calling of plastomes showed 1,521 InDels, 15,819 SNPs, and 13 variant hotspots. The synonymous substitution rates among the plastomes are 0.00 to 0.26. The maximum likelihood (ML) tree of the Araliaceae plastomes indicated that most genera included in this study were monophyletic, whereas Aralia was not monophyletic. Based on 78 plastid coding genes, the divergence times estimation indicated that most genera in Araliaceae had been diversifying into many evolutionary lineages between 0.34 and 8.29 million years ago (MYA). In the Araliaceae, the complete 45S nrDNA coding unit ranged from 5,849 to 5,892 bp in size and had 97.5 to 100 % of identities. Phylogenetic comparison utilizing plastomes and 45S nrDNA sequences revealed that *Panax stipuleanatus* and *Oplopanax elatus* have a conflicting phylogenetic position. The three genera, *Panax*, *Aralia*, and *Eleutherococcus*, in the Araliaceae, were excellently distinguished using 15 newly developed KASP markers in this study.

Keywords: Araliaceae species, plastomes, SNPs Molecular markers

INTRODUCTION

The Araliaceae is a flowering family rich in medicinal species (also known as the ginseng family), closely related to the Apiaceae family, under the order Apiales (Brussell, 2004; Chase et al., 2016; Plunkett et al., 2004b). The family consists of about 1500 species, classified into 41~55 genera (Liu et al., 2012; Mitchell et al., 2012; Plunkett et al., 2004b; Valcárcel et al., 2014; Valcárcel & Wen, 2019). For centuries, researchers have relied on floral morphological characters for diagnosing and classifying Araliaceae species (Chang-Jiang & Gin, 1982; Fiaschi et al., 2020; Plunkett et al., 2004b). However, the morphology of Araliaceae species is extremely diverse, and several characters were overlapped with Apiaceae in morphology (Nuraliev et al., 2010). The Araliaceae was traditionally viewed as a subfamily ("Aralineae") under Apiaceae (Calestani, 1905; Judd et al., 1994; Thorne, 1973). Recent classifications eventually overturned this classification. In particular, phylogenetic studies using molecular data revealed that the Araliaceae was a totally different and independent group from the other subfamilies of the Apiaceae (Plunkett et al., 2004a).

In the field of biodiversity classification and identification, the introduction of DNA barcoding has had a significant impact (Hebert et al., 2003). DNA barcoding with short DNA fragments has been widely used for rapid identification of plants (Group et al., 2009; Kress et al., 2005). DNA barcoding can be classified as single-locus barcodes, multi-locus DNA barcoded, super-barcodes, and specific barcodes (Li et al., 2015). Numerous DNA barcoding markers developed from nuclear and plastid genomes were used to classify the Araliaceae family. Three major groups, the *Aralia* group, the Asian Palmate group, and the *Polyscias-Pseudopanax* group within this family were able to be distinguished using nrITS region (ITS1-5.8S-ITS2) and

plastid *trn*L-*trn*F data; however, phylogenetic relationships of several lineages remained unclear (Plunkett et al., 2004a; Wen et al., 2001). Employing PCR and Sanger sequencing, plastid IGS regions (*trn*L–*trn*F, *atp*B–*rbc*L, and *psb*A–*trn*H) and plastid genes (*rps*16, *rpl*16, and *ndh*F) were utilized to reconstruct Araliaceae phylogenies but also failed to provide a fully resolved and well-supported phylogeny (Valcárcel & Wen, 2019).

Next-generation sequencing (NGS) technologies have been rapidly advanced offer an incredibly huge amount of accurate genomic sequence data. With the affordability of NGS approaches, nowadays, researchers are able to assemble the genomes in plants such as nuclear, mitochondrial, and chloroplast (plastid). The number of Araliaceae complete plastome sequences has rapidly been increased up to 142 (accessed on August 27, 2021), representing 25 genera and four major groups: (1) *Panax–Aralia*, (2) Asian Palmate, (3) *Polyscias–Pseudopanax*, and (4) Greater *Raukaua* (**Table S1-1**). Complete plastome sequences have been used as the super-barcoding, overcoming limitations of single- or multi-locus barcoding approaches and providing a high resolution for Araliaceae species identification (Kim et al., 2016a; Kim et al., 2015b; Kim et al., 2016b; Kim et al., 2017; Kim et al., 2018; Nguyen et al., 2018; Valcárcel & Wen, 2019). Several different research teams investigated the evolution of the Araliaceae family; however, they only focused on a few genera at a time. Therefore, it is necessary that a study should be conducted to describe the whole Araliaceae family.

It is possible to identify species by using a specific barcode that contains a region of DNA sequence with a high mutation rate. It could be based on molecular markers developed specifically for this purpose (Li et al., 2015). The most frequent type of polymorphism is single nucleotide polymorphism (SNP), which results from variation in a single base pair (Brookes, 1999). Through NGS technologies, the availability of large sets of plastome

sequences now allows SNPs to be explored for analyses, and marker-based SNPs can be designed. Among that, Kompetitive allele-specific PCR (KASP) technique, the third-generation markers, is referred to as the high throughput assays for population genotyping.

This research aimed to conduct a comprehensive survey of super-barcoding of 26 genera in the Araliaceae to evaluate the phylogenetic relationship and genetic diversity among Araliaceae species by including more plastome and 45S nrDNA sequences. Specifically, the efficacy of the KASP assay, which was a suitable marker for evaluating the genetic diversity of three economically important genera in the Araliaceae: *Panax*, *Aralia*, and *Eleutherococcus*, was emphasized.

MATERIAL AND METHODS

Taxon sampling and extraction of genomic DNA

Fresh leaves of 11 species: *Aralia elata*, *Aralia cordata*, *Dendropanax morbiferus*, *Hedera rhombea*, *Hedera helix*, *Fatsia japonica*, *Schefflera arboricola*, *Tetrapanax papyriferus*, *Oplopanax elatus*, *Polyscias fruticosa*, and *Centella asiatica* were mainly maintained and collected from Medicinal Plant Garden, College of Pharmacy, Seoul National University, Goyang, Korea. Following collection, the leaves were stored in freezers set to a freezing temperature (-70°C) until used. A modified cetyltrimethylammonium bromide (CTAB) method (Allen et al., 2006) was used to isolate whole genomic DNA. The isolated genomic DNA was analyzed for quality and quantity using agarose gel electrophoresis and a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA).

Illumina sequencing, assembly, and annotation of 45S nrDNA and plastomes

Library construction and sequencing were performed by PHYZEN (Seongnam, South Korea). An Illumina MiSeq instrument coping with 2 x 300 bp reads was used to generate whole-genome NGS data. Following the dnaLCW protocol (Kim et al., 2015b), low-quality read trimming and assembly were carried out by the CLC Genomic Workbench software, ver. 4.21 (https://digitalinsights.qiagen.com). *Panax ginseng* cultivar Chunpoong (GenBank: KM088019.1 (plastome) and KM036295.1 (45S nrDNA)) set as reference genomes. Plastid Genome Annotator (PGA) package (Qu et al., 2019) combining GeSeq (https://chlorobox.mpimp-golm.mpg.de/geseq.html) and manual corrections were processed to annotate the genes and other features along the genomes. The graphical representation of the resulting Araliaceae plastomes “ring” was established by OGDRAW

(https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) and Circos (http://circos.ca/).

Phylogenetic relationships and estimation of divergence times

Centella asiatica, belonging to the Mackinlayaceae, a subfamily of Apiaceae (Plunkett et al., 1996), was included as an outgroup for phylogenetic analyses. MAFFT(Katoh et al., 2019) performed the sequence alignment, and MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/) performed the alignment of 78 plastid protein-coding sequences (CDS). Maximum likelihood (ML) analyses were conducted using the GTR+CAT model in RAxML NG v0.9.0 (Kozlov et al., 2019), node support was assessed with 1000 bootstrap replicates. Two different data matrices were generated. First, a total of 75 plastomes, including 11 from this study, 12 from our previously reported data, and 52 from other previous studies, were used for a plastome-based phylogeny. Second, plastomes and 45S nrDNA sequences of 22 Araliaceae species and *C. asiatica* were used for phylogenetic comparison.

Under a lognormal relaxed clock model and the Yule speciation model, divergence-time estimation was inferred with concatenated data of 78 CDSs from 54 plastomes, using BEAST 2 (Bouckaert et al., 2019) for 10^8 generations sampling every 1000 generations. Two calibration points were applied: (1) the second calibration point was set for the root of the tree, at 51 MYA based on the divergence time estimation of *Panax* and carrot (Apiaceae); and (2) the divergence between *Panax* and *Aralia* lineages, at 7.50 MYA (Kim et al., 2018). Parameter log files were analyzed using Tracer v1.7.2 (Rambaut et al., 2018). The Tree files were summarized using TreeAnnotator v2.6.6 and visualized by FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Comparative analyses

The assembled Araliaceae plastomes and 45S nrDNA were compared with

the available genomes in GenBank for investigating alterations of the genomic sequences in several features: sequence size, genetic variations with SNPs and indels, nucleotide substitution rates and patterns. The alignments were input to Excel for variant calling and classifying by the COUNTIFS function. The variants that appeared every 1000 bp (1 kb) were measured, and their distribution was graphically represented. Total 78 protein-coding genes were employed to estimate nonsynonymous (Ka) and synonymous (Ks) substitution rates using the seqinr package in R (Charif & Lobry, 2007). The means of pairwise substitution rate values were estimated and plotted as a bar graph. Graphics were generated using GraphPad Prism software (https://www.graphpad.com/scientific-software/prism/).

Development and validation of KASP markers

The discovered SNPs were used extensively for developing KASP markers for the three genera: *Panax*, *Aralia*, and *Eleutherococcus*. 15 KASP assays were produced by LGC Biosearch Technologies (Teddington, United Kingdom) and experiments were done following their standard protocol (https://biosearch-cdn.azureedge.net/assetsv6/running-KASP-on-Roche-LC480.pdf). Developed assays were examined by using the gDNA isolated from 18 species, including six *Panax* (*P. ginseng*, *P. quinquefolius*, *P. japonicus*, *P. vietnamensis*, *P. notoginseng*, *P. stipuleanatus*), two *Aralia* (*A. elata*, *A. cordata*), *Dendropanax morbiferus*, *Eleutherococcus sessiliflorus*, 2 *Hedera* (*Hedera rhombea*, *Hedera helix*), *Fatsia japonica*, *Schefflera arboricola*, *Tetrapanax papyriferus*, *Oplopanax elatus*, *Kalopanax septemlobus*, and *Centella asiatica*.

RESULTS

The size and structure of Araliaceae plastome and 45S nrDNA

The summary statistics for NGS sequencing of ten Araliaceae and *Centella asiatica* were shown in Table 1-1. Including newly assembled plastomes, 74 plastomes that could represent 26 Araliaceae genera: 1_*Panax* (17 plastomes), 2_*Aralia* (8), 3_*Sciadodendron* (1), 4_*Osmoxylon* (1), 5_*Dendorpanax* (5), 6_ *Gamblea* (1), 7_*Chengiopanax* (1), 8_*Eleutherococcus* (7), 9_*Brassaiopsis* (1), 10_*Trevesia* (1), 11_*Metapanax* (1), 12_*Macropanax* (1), 13_*Kalopanax* (1), 14_*Hedera* (4), 15_*Merrilliopanax* (2), 16_*Fatsia* (3), 17_*Oreopanax* (1), 18_*Schefflera* (4), 19_*Heteropanax* (1), 20_*Tetrapanax* (2), 21_*Oplopanax* (2), 22_*Polyscias* (1), 23_ *Cheirodendron* (1), 24_*Raukaua* (3), 25_*Harmsiopanax* (1), 26_*Hydrocotyle* (4) were studied (Table S1-1). All Araliaceae plastomes included in this study were the same quadripartite structure and consistent in gene content and order. Total 115 genes were encoded by the plastomes, comprising four rRNAs, 31 tRNAs, and 80 proteins (Fig. 1-1A; Table S1-2). The sizes of 74 plastomes ranged from 152,880 bp (*Hydrocotyle sibthorpioides*) to 156,939 bp (*Harmsiopanax ingens*). The length of the large single copy region (LSC) varied between 84,067 bp (*Hydrocotyle nepalensis*) and 86,829 bp (*Eleutherococcus senticosus*). Each inverted repetition (IR) ranged in size from 25,060 bp (*Hydrocotyle verticillata*) to 26,134 bp (*Panax notoginseng*). The sizes of small single copy region (SSC) varied between 17,868 bp (*Fatsia japonica*) and 18,768 bp (*Hydrocotyle nepalensis*) (Table 1-2). Excluding the species of *Hydrocotyle* genus, the Araliaceae plastomes had a low degree of size variation, generally differed by less than 950 bp.

The structural organization of the 45S nrDNA in 23 Araliaceae species studied was almost identical. The alignment of the 45S nrDNA sequence was illustrated and color-coded for four regions: 18S, 5.8S, 26S, and ITS (Fig. 1-1 B).

These 45S nrDNA sequences were also highly conserved in length. 18S and 5.8S sequences were consistent in length, as 1,808 bp and 159 bp, respectively. The length of 26S sequences of the studied species ranged from 3,427 bp (*Dendropanax morbiferus*) to 3,457 bp (*Polyscias fruticosa*), indicating considerable size variation (Table 1-3).

Table 1-1. Summary of NGS data for plastomes and 45S nrDNAs of 11 species

No.	Species	NGS_ID	Total reads (Gb)	Average coverage (X)	
				Plastomes	45S
1	*Aralia elata*	IM180813_55	24.35	1,482	2,840
2	*Aralia cordata*	IM180813_32	3.75	500	483
3	*Dendropanax morbiferus*	IM180813_56	26.51	1,015	1,573
4	*H edera Rhombea*	IM180813_57	29.82	590	1,641
5	*Hedera helix*	IM180813_34	3.68	172	660
6	*Fatsia japonica*	IM180813_33	3.26	217	605
7	*Schefflera arboricola*	IM180813_58	20.42	373	2,612
8	*Tetrapanax papyriferus*	IM180813_35	2.82	759	761
9	*Oplopanax elatus*	IM180813_36	3.36	182	1,651
10	*Polyscias fruticosa*	IM1907233	4.49	378	421
11	*Centella asiatica*	IM180813_54	13.82	1,737	1,031

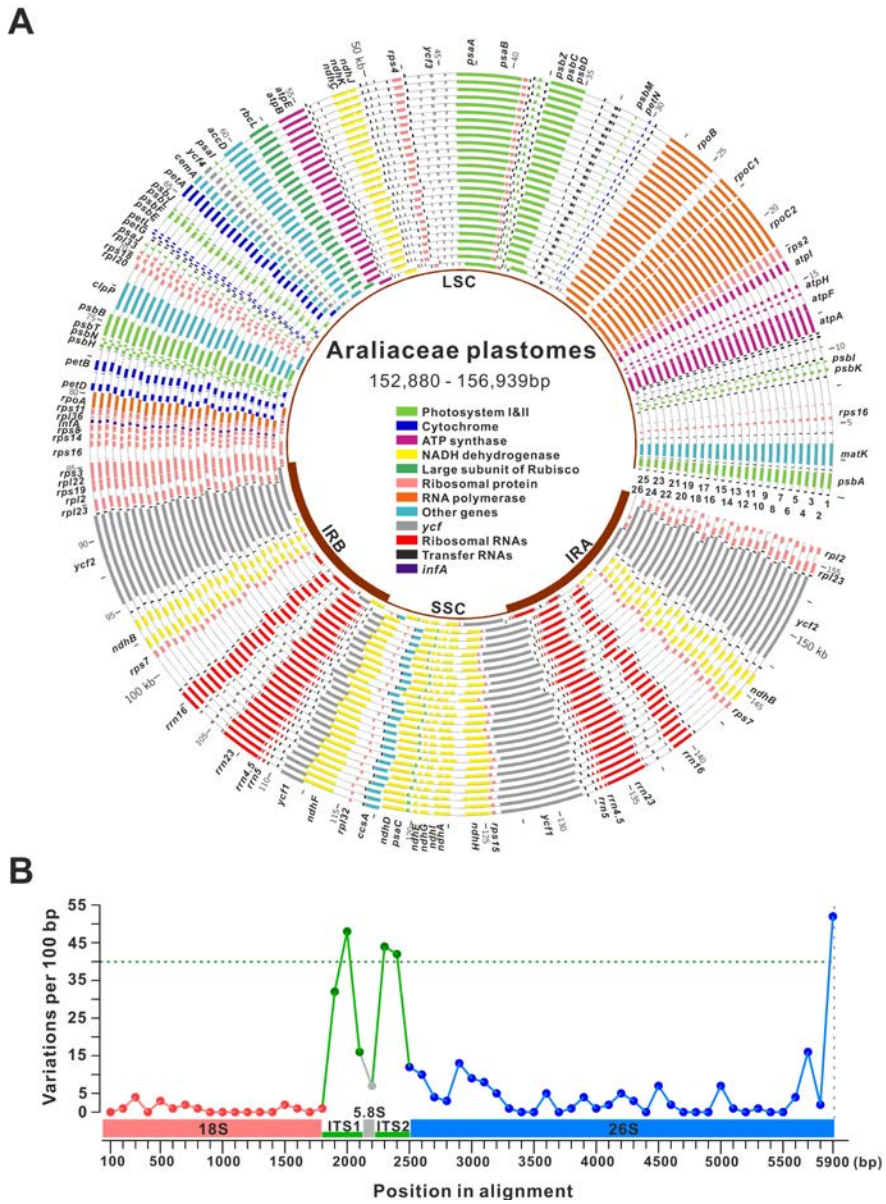

Figure 1-1. Araliaceae plastomes and 45S nrDNAs. (A) The plastomes map of representatives of 26 Araliaceae genera: 1_*Panax*, 2_*Aralia*, 3_*Sciadodendron*, 4_*Osmoxylon*, 5_*Dendorpanax*, 6_ *Gamblea*, 7_ *Chengiopanax*, 8_*Eleutherococcus*, 9_*Brassaiopsis*, 10_*Trevesia*, 11_*Metapanax*, 12_*Macropanax*, 13_*Kalopanax*, 14_*Hedera*, 15_*Merrilliopanax*, 16_*Fatsia*, 17_*Oreopanax*, 18_*Schefflera*, 19_*Heteropanax*, 20_*Tetrapanax*, 21_*Oplopanax*, 22_*Polyscias*, 23_ *Cheirodendron*, 24_*Raukaua*, 25_*Harmsiopanax*, 26_*Hydrocotyle.* Genes belonging to different groups are coded with different colors, transcribed clockwise genes were shown inside. (B) The graph shows the number of variants/100bp in the alignment of the 22 Araliaceae species nuclear 18S-ITS1-5.8S-ITS2-26S units.

Table 1-2. Summary of plastome sequences for 74 Araliaceae and *Centella asiatica*

No.	Species	Length (bp)				GeneBank /NGS_ID
		LSC	SSC	IR	Total	
1	*Panax ginseng* cv. Chunpoong	86,131	18,085	26,016	156,248	KM088019.1*
2	*Panax ginseng* cv. Gumpoong	86,132	18,078	26,073	156,356	KM067388.1*
3	*Panax quinquefolius*	86,127	18,080	26,076	156,359	KM088018.1*
4	*Panax quinquefolius*	86,080	17,994	25,998	156,070	MK408923.1
5	*Panax japonicus*	86,202	18,014	25,986	156,188	KP036469.1*
6	*Panax japonicus* var. *bipinnatifidus*	86,189	18,007	25,967	156,130	MK408940.1
7	*Panax wangianus*	86,179	17,967	26,015	156,176	MK408921.1
8	*Panax wangianus*	86,186	17,967	26,072	156,297	MK408930.1
9	*Panax vietnamensis*	86,181	17,930	25,941	155,993	KP036471.1*
10	*Panax vietnamensis*	86,181	17,930	25,941	155,993	KP036470.1*
11	*Panax zingiberensis*	86,118	17,967	26,053	156,191	MK408968.1
12	*Panax zingiberensis*	86,172	17,975	26,128	156,403	MK408966.1
13	*Panax notoginseng*	86,193	18,005	26,134	156,466	KP036468.1*
14	*Panax notoginseng*	86,160	18,005	26,077	156,319	MK408937.1
15	*Panax stipuleanatus*	86,119	18,175	25,885	156,064	KX247147.1*
16	*Panax stipuleanatus*	86,187	18,149	25,877	156,090	MF377622.1
17	*Panax trifolius*	86,325	18,048	25,892	156,157	MF100782.1
18	*Aralia elata*	86,266	18,112	25,921	156,220	KT153023.1*
19	***Aralia elata***	**86,262**	**18,113**	**25,921**	**156,217**	**IM180813_55***
20	*Aralia undulata*	86,032	18,091	26,105	156,333	KC456163.1
21	***Aralia cordata***	**86,099**	**18,053**	**25,965**	**156,082**	**IM180813_32***
22	*Aralia cordata*	86,104	18,053	25,965	156,087	MH778959.1
23	*Aralia continentalis*	86,052	18,053	25,947	155,999	MG914654.1
24	*Aralia atropurpurea*	86,269	18,081	25,961	156,272	MK809524.1
25	*Aralia atropurpurea*	86,234	18,070	25,961	156,226	MK778455.1
26	*Sciadodendron excelsum*	86,479	18,560	25,511	156,061	MK943809.1
27	*Osmoxylon novoguineense*	86,698	18,545	25,519	156,281	MK943807.1
28	*Dendropanax morbiferus*	86,478	18,126	25,881	156,366	KR136270.1*
29	***Dendropanax morbiferus***	**86,535**	**18,126**	**25,922**	**156,505**	**IM180813_56***
30	*Dendropanax dentiger*	86,683	18,150	25,927	156,687	KP271241.1
31	*Dendropanax oligodontus*	86,443	18,076	25,942	156,403	MT909827.1
32	*Dendropanax nutans*	86,506	18,146	25,939	156,530	MK943797.1
33	*Chengiopanax sciadophylloides*	86,498	18,037	25,931	156,397	MK930365.1
34	*Gamblea ciliata* var. *evodiifolia*	86,217	18,171	25,929	156,246	MK943799.1
35	*Eleutherococcus brachypus*	86,745	18,185	25,936	156,802	MN527993.1
36	*Eleutherococcus trifoliatus*	86,675	18,175	25,933	156,716	MN727298.1
37	*Eleutherococcus trifoliatus*	86,750	18,133	25,934	156,751	MT754220.1
38	*Eleutherococcus gracilistylus*	86,732	18,176	25,931	156,770	KT153020.1*
39	*Eleutherococcus senticosus*	86,829	18,200	25,917	156,863	KY085901.1
40	*Eleutherococcus senticosus*	86,758	18,154	25,928	156,768	JN637765.1
41	*Eleutherococcus sessiliflorus*	86,606	18,214	25,955	156,730	KT153019.1*

No.	Species	Length (bp)				GeneBank /NGS_ID
		LSC	SSC	IR	Total	
42	*Brassaiopsis hainla*	86,570	18,021	25,934	156,459	KC456164.1
43	*Trevesia sundaica*	86,366	18,113	25,999	156,477	MK943811.1
44	*Metapanax delavayi*	86,364	18131	25,924	156,343	KC456165.1
45	*Macropanax dispermus*	86,106	18,116	25,930	156,082	MK943802.1
46	*Kalopanax septemlobus*	86,470	18,119	25,912	156,413	KC456167.1
47	*Hedera nepalensis* var. *sinensis*	86,600	18,174	25,939	156,652	MK130890.1
48	***Hedera rhombea***	**86,628**	**18,174**	**25,939**	**156,680**	**IM180813_57***
49	*Hedera helix*	86,630	18,182	25,938	156,688	MK943800.1
50	***Hedera helix***	**86,630**	**18,182**	**25,938**	**156,688**	**IM180813_34***
51	*Merrilliopanax listeri*	86,389	18,127	25,939	156,394	MK943803.1
52	*Merrilliopanax listeri*	86,422	18,143	25,904	156,373	MK943804.1
53	*Fatsia polycarpa*	86,492	17,870	25,627	155,616	MK943798.1
54	*Fatsia japonica*	86,491	17,868	25,627	155,613	KR021045.1
55	***Fatsia japonica***	**86,487**	**17,869**	**25,626**	**155,608**	**IM180813_33***
56	*Oreopanax obtusifolius*	86,265	18,134	25,960	156,319	MK943806.1
57	*Schefflera heptaphylla*	86,613	18,148	25,962	156,685	KT748629.1
58	***Schefflera arboricola***	**86,618**	**18,148**	**25,969**	**156,704**	**IM180813_58***
59	*Schefflera actinophylla*	86,577	18,162	25,968	156,675	MT385083.1
60	*Schefflera delavayi*	86,126	18,063	26,076	156,341	KC456166.1
61	*Heteropanax fragrans*	86,600	18,145	25,920	156,585	MK943801.1
62	*Tetrapanax papyrifer*	86,137	18,015	25,969	156,090	MK943810.1
63	***Tetrapanax papyriferus***	**86,152**	**17,988**	**25,942**	**156,024**	**IM180813_35***
64	*Oplopanax horridus*	86,295	18,122	25,894	156,205	MK943805.1
65	***Oplopanax elatus***	**86,137**	**18,113**	**25,879**	**156,008**	**IM180813_36***
66	***Polyscias fruticosa***	**86,298**	**18,118**	**25,964**	**156,344**	**IM1907233***
67	*Raukaua anomalus*	86,603	18,171	25,945	156,664	MT385080.1
68	*Raukaua simplex*	86,599	18,222	25,946	156,713	MT385082.1
69	*Raukaua edgerleyi*	86,624	18,226	25,949	156,748	MT385081.1
70	*Cheirodendron bastardianum*	86,573	18,183	25,971	156,698	MT385071.1
71	*Harmsiopanax ingens*	86,808	18,175	25,978	156,939	MK922468.1
72	*Hydrocotyle nepalensis*	84,435	18,768	25,075	153,353	MT561038.1
73	*Hydrocotyle sibthorpioides*	84,067	18,691	25,061	152,880	KT589392.1
74	*Hydrocotyle verticillata*	84,355	18,732	25,060	153,207	HM596070.1
75	***Centella asiatica***	**86,169**	**17,971**	**25,310**	**154,760**	**IM180813_54***

*, works of Lab. of Functional Plants: 23 plastomes (Kim et al., 2016a; Kim et al., 2015a; Kim et al., 2016b; Kim et al., 2017; Kim et al., 2018; Nguyen et al., 2015; Nguyen et al., 2018). 11 new plastomes were bolded.

Table 1-3. Summary of 45S nrDNA sequences for 22 Araliaceae and *C. asiatica*

No.	Species	Length (bp)						GeneBank /NGS_ID
		18S	ITS1	5.8S	ITS2	26S	Total	
1	*P. ginseng*	1,808	224	159	233	3,453	5,877	KM036295.1[a]
2	*P. quinquefolius*	1,808	224	159	233	3,450	5,874	KM036297.1[a]
3	*P. japonicus*	1,808	224	159	233	3,450	5,874	KT380920.1[a]
4	*P. wangianus*	1,808	224	159	233	3,450	5,874	MK408764.1[b]
5	*P. vietnamensis*	1,808	224	159	233	3,450	5,874	KT380922.1[a]
6	*P. zingiberensis*	1,808	224	159	233	3,450	5,874	MK408787.1 [b]
7	*P. notoginseng*	1,808	224	159	233	3,450	5,874	KT380921.1[a]
8	*P. stipuleanatus*	1,808	223	159	232	3,451	5,873	MF091695.1[a]
9	*P. trifolius*	1,808	222	159	233	3,453	5,875	MF099781.1[a]
10	*A. elata*	1,808	222	159	233	3,451	5,873	KT380919.1[a]
11	*A. elata*	1,808	222	159	233	3,450	5,872	IM180813_55
12	*A. cordata*	1,808	223	159	233	3,451	5,874	IM180813_32
13	*D. morbiferus*	1,808	222	159	233	3,427	5,849	KT380923.1[a]
14	*D. morbiferus*	1,808	222	159	233	3,427	5,849	IM180813_56
15	*H. rhombea*	1,808	224	158	233	3,451	5,874	IM180813_57
16	*H. helix*	1,808	224	159	233	3,451	5,875	IM180813_34
17	*F. japonica*	1,808	223	158	233	3,447	5,869	IM180813_33
18	*E. sessiliflorus*	1,808	223	159	233	3,451	5,874	KT380924.1[a]
19	*O. elatus*	1,808	223	159	232	3,451	5,873	IM180813_36
20	*S. arboricola*	1,808	224	159	233	3,448	5,872	IM180813_58
21	*T. papyriferus*	1,808	222	159	234	3,440	5,863	IM180813_35
22	*P.fruticosa*	1,808	227	159	229	3,457	5,880	IM1907233
23	*C. asiatica*	1,808	229	159	237	3,459	5,892	IM180813_54

[a], obtained from our previous study. [b], obtained from Ji et al. 2019.

Variations of plastomes and 45S nrDNA sequences in Araliaceae

While the Araliaceae plastomes were highly conserved in terms of gene content and arrangement, the collection of 74 complete plastomes revealed 43,458 variable sites and 13 variant hotspots (Fig. 1-2). In total, the dataset encompassed 15,819 SNPs and 1,521 Indels. In IR regions, the number of variable sites was smaller than in other SC regions. The variation data was characterized at two levels of differentiation: genus level and species level. On the other hand, intra- and interspecific genetic variability were grouped as random variants. As a consequence, the variable sites were classified into three categories: (1) genus-specific variants (18.9%, 242 Indels and 3,286 SNPs); (2) random variants (30%, 404 Indels and 4,033 SNPs); and (3) species-specific variants (51.1%, 875 Indels and 8,500 SNPs). Using sliding windows of 1kb, ten hotspots had over 600 variable sites, and 3 hotspots of genus-specific variants were identified (Fig. 1-2). The species-specific variant was the most prevalent variant type with over 50% of variable sites in seven hotspots (7 intergenic spacers): *trn*K–*rps*16, *rps*16–*trn*Q, *trn*E–*trn*T, *trn*T–*psb*D, *trn*T–*trn*L, *ndh*C–*trn*L, and *ndh*F–*rpl*32. Random variants were abundant in 3 regions: the *trn*C–*pet*N intergenic spacer, *psa*I–*ycf*4, and *ycf*1 gene. The three regions: the *trn*S–*trn*G intergenic spacer, *accD* gene, and the *trn*R–*ycf*1 intergenic spacer, had the most genus-specific variants.

The 45S nrDNA pairwise sequence comparison assessed 97.5-100% identity and 98-100% similarity (Table 1-4). The overall 45S nrDNA variation sites were 383 (16 in 18S; 92 in ITS1; 10 in 5.8S; 99 in ITS2 and 169 in 26S) (Fig. 1-4 B); ITS1 and ITS2 were extremely variable (Fig. 1-1 B).

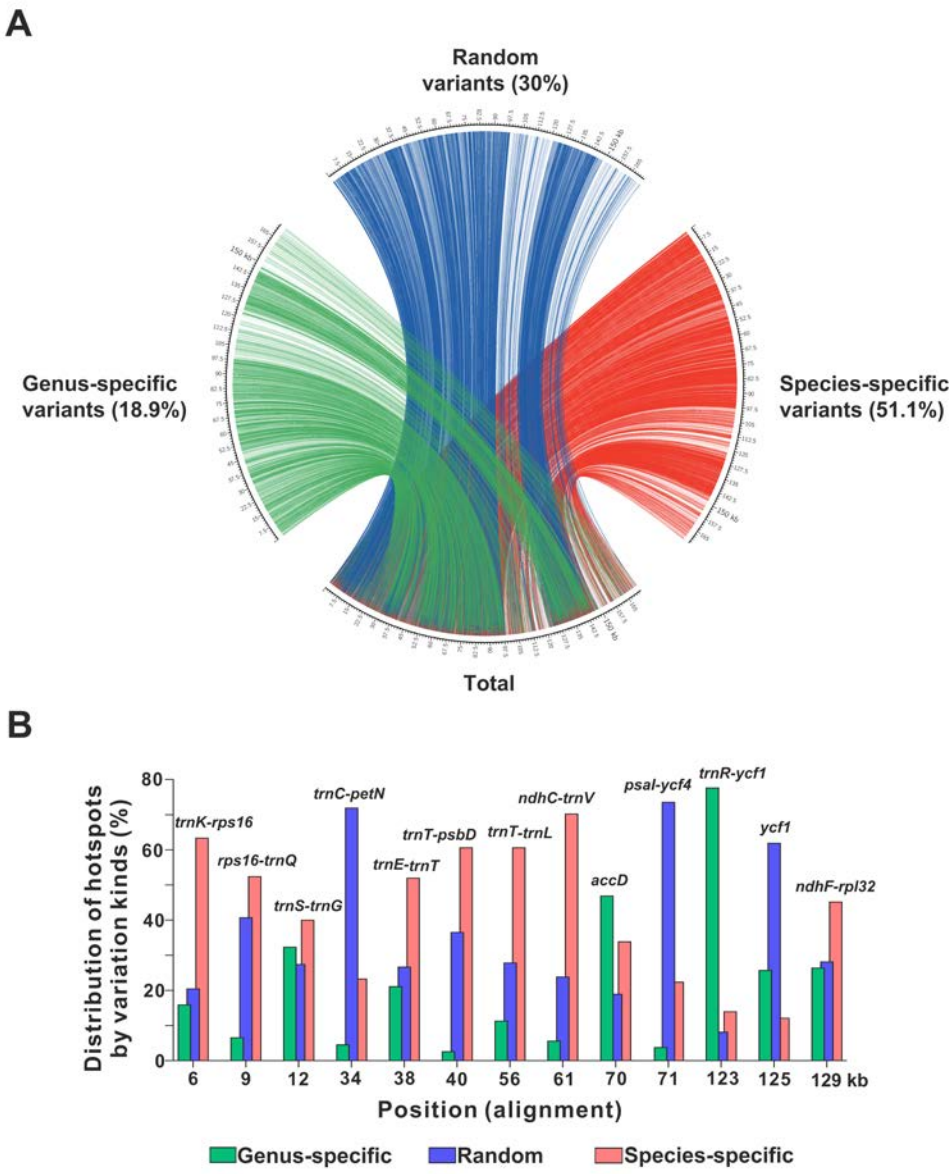

Figure 1-2. Araliaceae plastome-wide variation distribution. (A) Circos plot of plastome variations, which were separated into genus-specific variants, random variants, species-specific variants. The outermost black lines denote the position of variants. Green lines connect genus-specific variants, blue lines connect random variants, and red lines connect species-specific variants. (B) Percentile distributions of 3 kinds of genetic variants in 13 variant hotspots.

Table 1-4. Sequence identity and similarity matrix of 45S nrDNA

No.	Species	Percent identity (%)																				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	*P. ginseng*		99.8	99.5	99.6	99.6	99.6	99.3	99.1	98.8	99.1	99.1	99.0	98.1	98.1	97.9	98.0	98.5	98.5	98.7	98.3	98.0	98.2
2	*P. quinquefolius*	99.9		99.6	99.7	99.7	99.7	99.4	99.2	98.8	99.2	99.2	99.1	98.2	98.2	98.0	98.0	98.5	98.6	98.8	98.4	98.1	98.2
3	*P. japonicus*	99.6	99.7		99.7	99.6	99.6	99.2	99.0	98.7	99.1	99.0	98.9	98.1	98.1	97.8	97.9	98.4	98.4	98.6	98.2	97.9	98.2
4	*P. wangianus*	99.7	99.8	99.6		99.6	99.6	99.3	99.1	98.7	99.1	99.1	99.0	98.1	98.1	97.9	97.9	98.4	98.5	98.7	98.3	98.0	98.2
5	*P. vietnamensis*	99.7	99.8	99.4	99.7		99.9	99.4	99.2	98.8	99.2	99.1	99.1	98.2	98.2	98.0	98.0	98.5	98.6	98.8	98.4	98.1	98.2
6	*P. zingiberensis*	99.7	99.8	99.4	99.7	100.0		99.4	99.2	98.8	99.2	99.1	99.1	98.2	98.2	98.0	98.0	98.5	98.6	98.8	98.4	98.1	98.2
7	*P. notoginseng*	99.4	99.5	99.4	99.4	99.6	99.6		99.0	98.8	99.2	99.1	99.0	98.1	98.1	97.9	97.9	98.4	98.5	98.6	98.3	98.0	98.2
8	***P. stipuleanatus***	99.3	99.4	99.2	99.3	99.4	99.4	99.3		98.9	**99.4**	**99.4**	**99.3**	98.3	98.3	98.1	98.1	98.5	98.5	98.8	98.5	98.1	98.3
9	*P. trifolius*	99.1	99.1	98.9	99.0	99.1	99.1	99.0	99.1		99.2	99.2	99.1	98.1	98.1	97.9	97.9	98.4	98.4	98.7	98.2	98.0	98.1
10	*A. elata*	99.3	99.4	99.3	99.3	99.4	99.4	99.4	**99.6**	99.0		99.9	99.5	98.4	98.4	98.3	98.3	98.8	98.9	99.0	98.7	98.4	98.6
11	*A. elata**	99.3	99.3	99.2	99.3	99.3	99.3	99.3	**99.6**	99.0	99.9		99.5	98.4	98.4	98.2	98.2	98.8	98.8	99.0	98.7	98.3	98.5
12	*A. cordata**	99.2	99.3	99.1	99.2	99.3	99.3	99.2	**99.5**	98.9	99.6	99.6		98.3	98.3	98.3	98.3	98.7	98.7	99.0	98.6	98.3	98.4
13	*D. morbifer*	98.4	98.5	98.4	98.4	98.5	98.5	98.5	98.6	98.4	98.7	98.7	98.6		100.0	98.3	98.2	98.6	98.5	98.8	98.4	98.1	97.8
14	*D. morbifer**	98.4	98.5	98.4	98.4	98.5	98.5	98.5	98.6	98.4	98.7	98.7	98.6	100.0		98.3	98.2	98.6	98.5	98.8	98.4	98.1	97.8
15	*H. rhombea**	98.4	98.4	98.3	98.3	98.4	98.4	98.4	98.6	98.4	98.7	98.7	98.7	98.6	98.6		99.8	98.6	98.8	98.8	98.6	97.9	97.5
16	*H. helix**	98.4	98.4	98.3	98.3	98.4	98.4	98.4	98.6	98.3	98.7	98.6	98.7	98.5	98.5	99.9		98.6	98.8	98.8	98.6	97.8	97.6
17	*F. japonica**	98.8	98.8	98.8	98.7	98.8	98.8	98.8	98.9	98.7	99.0	99.0	98.9	98.9	98.9	98.2	98.3		99.1	**99.3**	98.7	98.3	98.1
18	*E. sessiliflorus**	98.8	98.9	98.7	98.7	98.8	98.8	98.8	98.9	98.7	99.1	99.1	99.0	98.8	98.8	98.4	98.4	98.9		**99.3**	98.7	98.4	98.2
19	***O. elatus****	99.0	99.0	98.9	98.9	99.0	99.0	98.9	99.1	98.9	99.2	99.2	99.1	98.9	98.9	98.5	98.5	**99.1**	**99.1**		98.8	98.6	98.4
20	*S. arboricola**	98.6	98.7	98.5	98.6	98.7	98.7	98.6	98.9	98.6	99.0	99.0	98.9	98.6	98.6	98.2	98.2	99.0	98.9	99.0		98.4	97.9
21	*T. papyriferus**	98.3	98.4	98.2	98.3	98.3	98.4	98.3	98.4	98.3	98.6	98.6	98.5	98.3	98.3	98.2	98.2	98.5	98.6	98.7	98.6		98.1
22	*P. fruticosa**	98.6	98.6	98.5	98.6	98.6	98.6	98.6	98.7	98.5	98.8	98.8	98.6	98.2	98.2	98.0	98.0	98.4	98.5	98.6	98.3	97.8	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
		Percent similarity (%)																					

*, the newly obtained sequence

Phylogenetic study

The ML phylogenetic analysis of the plastomes revealed five supergroups with high branch supporting values: (A) *Panax, Aralia, Sciadodendron,* and *Osmoxylon*; (B) *Dendorpanax, Gamblea, Chengiopanax, Eleutherococcus, Brassaiopsis, Trevesia, Metapanax, Macropanax, Kalopanax, Hedera, Merrilliopanax, Fatsia, Oreopanax, Schefflera, Heteropanax, Tetrapanax,* and *Oplopanax*; (C) *Polyscias*; (D) *Cheirodendron* and *Raukaua*; and (E) *Harmsiopanax* and *Hydrocotyle* (Fig. 1-3). Several internal nodes within the supergroup (B), also known as the Asian Palmate group, showed low bootstrap supports (<70), but most of the investigated genera formed monophyletic lineages (except genus *Aralia*). The newly obtained 11 plastomes in this study belong to eight genera: *Aralia, Dendorpanax, Hedera, Fatsia, Schefflera, Tetrapanax, Oplopanax,* and *Polyscias*. There were no incorrect phylogenetic placements of the newly obtained plastomes (Fig. 1-3).

However, there were phylogenetic incongruences between plastome-based phylogeny and 45S nrDNA phylogeny of 22 Araliaceae species. The incongruences were the phylogenetic positions of *Panax stipuleanatus* and *Oplopanax elatus*. In the plastome-based phylogeny, *P. stipuleanatus* was sister to other *Panax* species within the *Panax* clade, and *O. elatus* was sister to the *Schefflera–Tetrapanax* clade. Inconsistently, in the 45S nrDNA tree, *P. stipuleanatus* was nested within the *Aralia* clade, and *O. elatus* was sister to *Fatsia japonica*, which was associated with the estimated identity and similarity of these species (Fig. 1-4 A, Table 1-4). In the enlarged genetic investigation of 45S nrDNA, *P. stipuleanatus* exhibited 24 SNPs diversity with the remaining *Panax* (Fig. 1-4 B), and *O. elatus* displayed 33 SNPs, distinguishing it from *Schefflera* and *Tetrapanax* (Fig. 1-4 C).

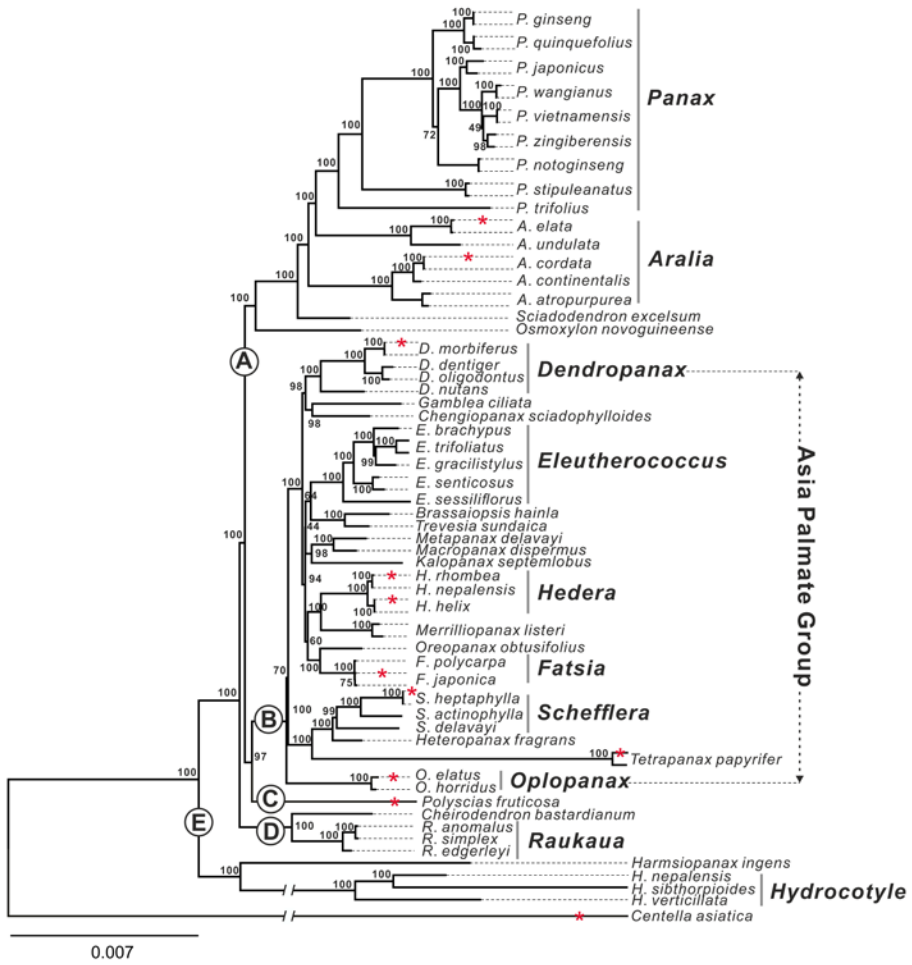

Figure 1-3. Maximum likelihood phylogeny of Araliaceae plastomes. 74 plastomes of 26 genera were employed. *C. asiatica* was designed as an outgroup. Red asterisks corresponded to the new plastomes. Bootstrap support (BS) was shown near the branches. The branches to *Hydrocotyle* and *C. asiatica* were truncated to allow a better display.

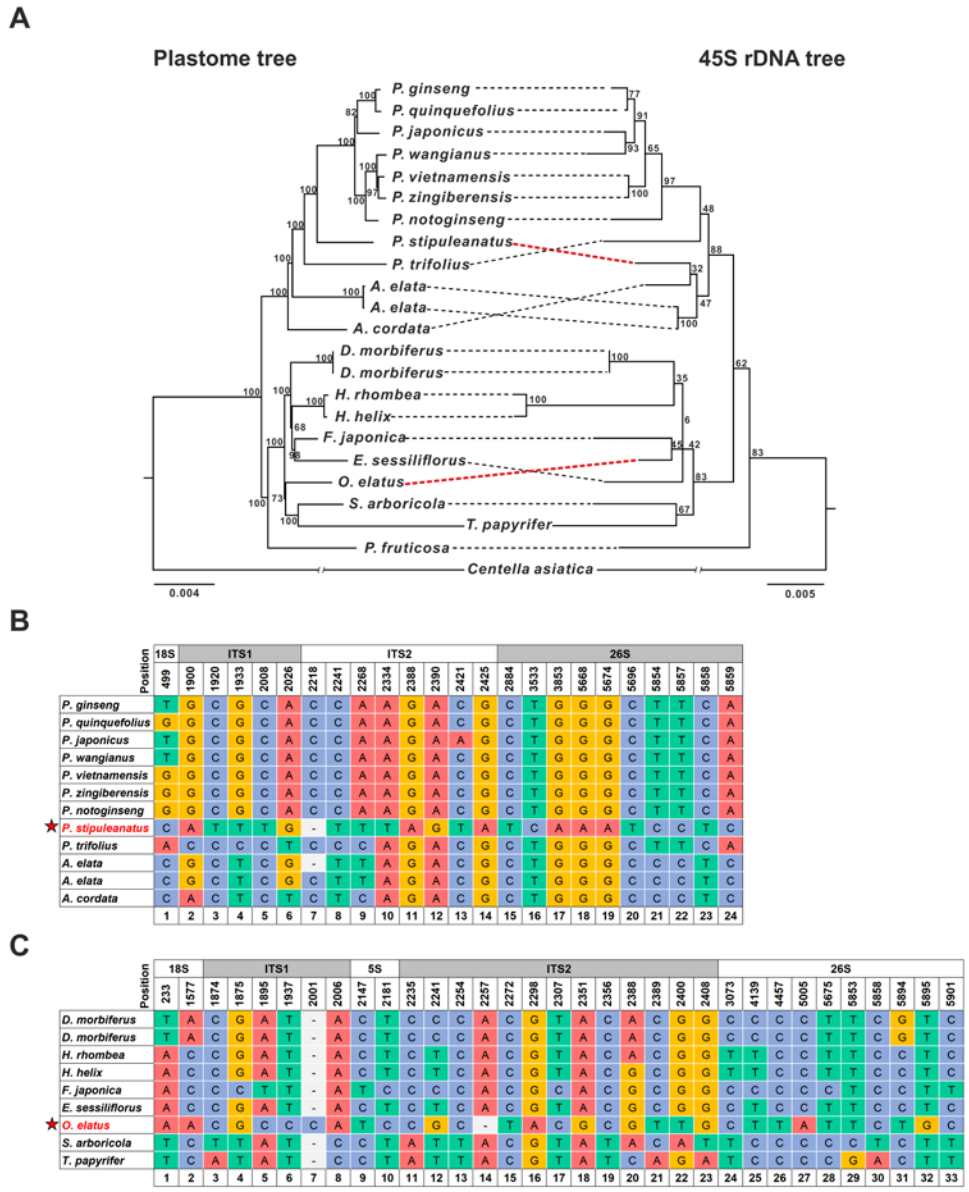

Figure 1-4. Maximum likelihood trees built with plastomes and 45S nrDNAs. (A) Phylogenetic analysis of 22 Araliaceae species was studied. Bootstrap values calculated for 1000 replicates were shown on the branches. Dashed lines connect the positions of each species in the two trees. (B) The 45S nrDNA of *P. stipuleanatus* showed 24 SNPs of genetic diversity with the remaining *Panax*. (C) The 45S nrDNA of *O. elatus* displayed 33 SNPs, distinguishing it from *Schefflera* and *Tetrapanax*

Substitution rates of plastomes and estimation of divergence time

The nucleotide substitutions were found in all 78 examined genes. The mean of Ks values were about 0.001 ~ 0.041, the mean of Ka values were about 0 ~ 0.017. The Ka of 5 genes: *psbI*, *psbF*, *petG*, *psaJ*, and *psbT* were 0.00 (Fig. 1-5). Ka was greater than Ks in 6 genes: *accD*, *clpP*, *psbH*, *rps8*, *ycf2*, and *ycf15* (Fig. 1-5).

The five Araliaceae supergroups were divided into 26 clades that correspond to 26 genera. There were two distinct subclades in the *Aralia* and *Schefflera* genera (Fig. 1-6, Fig. 1-7). A total of 1474 clade/subclade-specific nonsynonymous variants (ψ) were identified in 65 CDSs, highlighting the divergence of the 26 genera in Araliaceae (Fig. 1-6). The highest number of ψ were in 4 genes: *mat*K (127), *rpo*C2 (168), *ycf*2 (132), and *ndh*F (123). Furthermore, the number of ψ were also increased in *clp*P gene of *Panax* and *ndh*I gene of *Dendropanax* (Fig. 1-6).

The result of molecular dating showed that the first split within Araliaceae occurred during the Oligocene of the Paleogene Period (median = 26.8 MYA, 95% high posterior density [HPD] = 19.6–35.2 MYA) (Fig. 1-7). The major splits forming the four supergroups, A, B, C, and D, were estimated to have occurred between the Early and Middle of the Miocene in the Neogene (8.6–17 MYA). The divergence time between the *Panax-Aralia* group (A) and the Asian Palmate group (B) was estimated to be approximately 12.4 MYA (95% HPD = 9.8–15.9 MYA). Subsequently, most generic divergences in Araliaceae occurred in 2.01–7.39 MYA during the Late Miocene. Based on the synonymous substitutions (Ks) per year (λ) of 2×10^{-9} for plastome (Wolfe et al., 1987), the means of Ks were converted to divergence time (T) using the formula $T = Ks/2\lambda \times 10^{-6}$ MYA (Fig. S1-1). The result also indicated that Araliaceae species diverged 7.62 MYA.

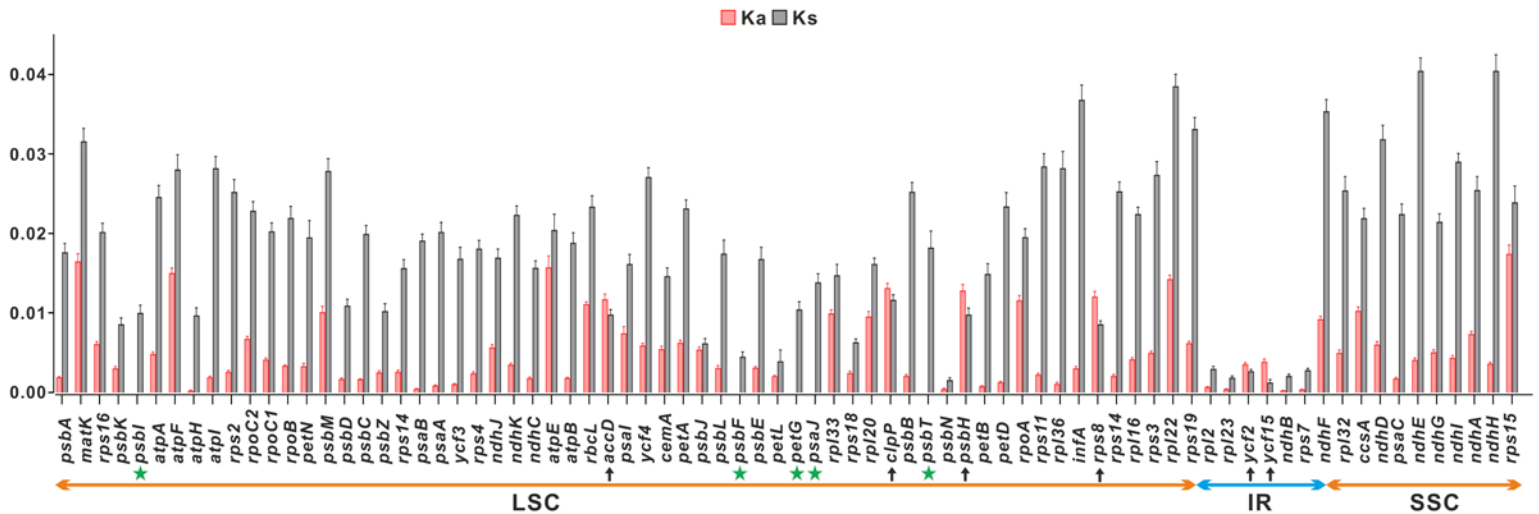

Figure 1-5. A summary of the non-synonymous (Ka) versus synonymous (Ks) in 78 CDSs of 54 Araliaceae plastomes. Bar graph illustrated the mean (bars) and 95% confidence interval (lines) of Ka and Ks. Green stars indicated genes with no value of Ka. The arrows indicated genes with Ka > Ks.

Figure 1-6. Heatmap showing number of clade/subclade-specific nonsynonymous variants (ψ).

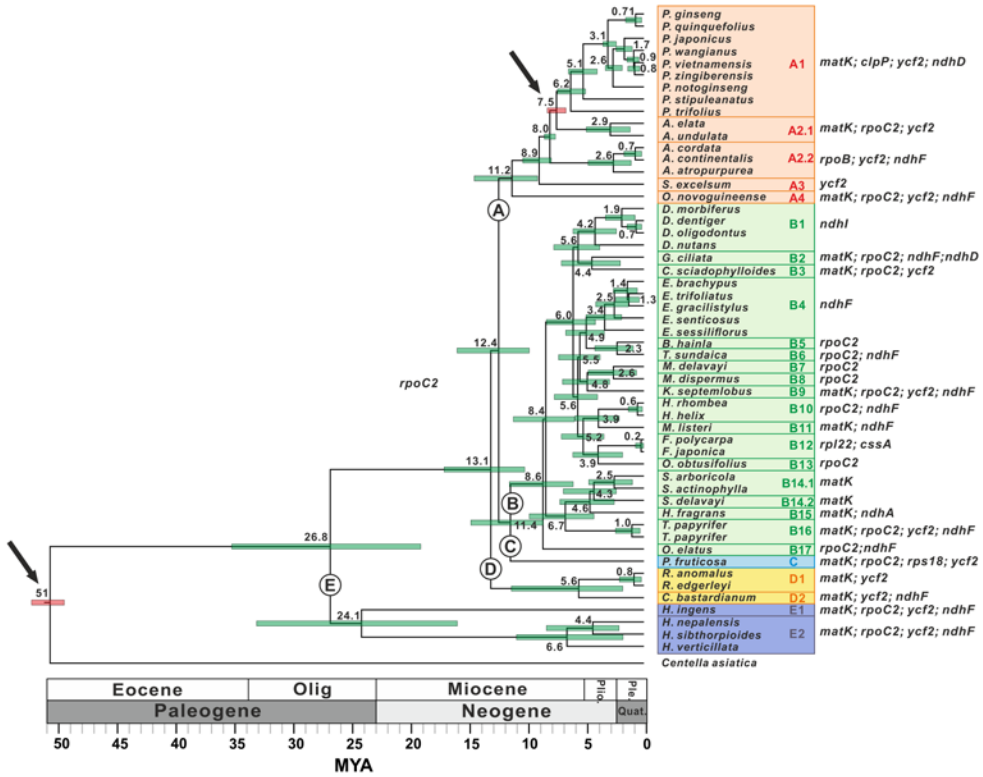

Figure 1-7. Molecular dating tree. The tree was obtained from the Bayesian analysis of 54 plastomes in BEAST. Values near nodes were the mean average ages of the nodes, bars representing the 95% HPD, and the arrows indicate two calibration points following the previous results. In the right part, genes with a large number of clade-specific nonsynonymous variants were indicated.

SNP markers associated with the important genera

The diversity of *Panax*, *Aralia*, and *Eleutherococcus* species was investigated utilizing SNPs in CDSs for genotyping: (1) *Panax–Aralia* group; (2) *Panax* clade; (3) *Panax* subclade A: *P. ginseng* and *P. quinquefolius*; (4) *Panax* subclade B: *P. japonicus*, *P. wangianus*, *P. vietnamensis*, *P. zingiberensis*, and *P. notoginseng*; (5) *Aralia* subclade A: *A. elata* and *A. undulata*; (6) *Aralia* subclade B: *A. cordata*, *A. continentalis*, and *A. atropurpurea*; (7) *Eleutherococcus* clade. Fifteen SNPs were identified and successfully developed to KASP markers as a genetic diagnosis (Table 1-5). The KASP markers were excellently distinguished the Araliaceae genotypes: PaAr.Km1, 2 & 3 clustered the accessions of *Panax* and *Aralia*; Pa.Km4, 5 & 6 differentiated *Panax* species with the other species; Pa.A.Km7 & 8 genotyped *P. ginseng* and *P. quinquefolius*; Pa.B.Km9 & 10 genotyped *P. japonicus*, *P. vietnamensis*, and *P. notoginseng*; Ar.A.Km11 & 12 genotyped *A. elata*; Ar.B.Km13 genotyped *A. cordata*; and El.Km14 & 15 genotyped *E. sessiliflorus* (Table 1-6; Fig. 1-8).

Table 1-5. KASP markers developed for genotyping *Panax*, *Aralia* and *Eleutherococcus* species

No	Gene	KASP ID	SNP	FAM primers	HEX primers	Common primers
1	*psaB*	**PaAr.Km1**	[T/G]	AAGAGTTTAACCATAAATAATCTCTTAACCAT	GAGTTTAACCATAAATAATCTCTTAACCAG	ATGGCAGGGTAACGTTTCRCAGTTTAA
2	*petA*	**PaAr.Km2**	[G/A]	GACGGGAGCAAGAGTAATAATACG	CGACGGGAGCAAGAGTAATAATACA	TACTATACCWGCTGCTGTAGCATTATAAA
3	*psbB*	**PaAr.Km3**	[T/C]	GAACAAAAGGATCAAAACCTTCCACA	GAACAAAAGGATCAAAACCTTCCACG	GGACTAACGGGAAAAGTACAAKCTGTA
4	*rpoC2*	**Pa.Km4**	[C/A]	CATTTGTCCTTGGGGATCCGC	AATCATTTGTCCTTGGGGATCCGA	GTACATCAATTAGTAGGTATGAGAGGAYTA
5	*rpoB*	**Pa.Km5**	[C/T]	CCCGAGTGAACATTCAAATATCTGC	AACCCGAGTGAACATTCAAATATCTGT	GTTGATATGGTCTTCAAYCCATTAGGAGT
6	*ndhK*	**Pa.Km_6**	[A/G]	TTATGGCCGCTTCTCTATGGTACT	ATGGCCGCTTCTCTATGGTACC	CCTAKTAGTGAAGCAAATTCAATGAAGCAA
7	*ndhC*	**Pa.A.Km7**	[C/T]	TTGCTAATCGGGGCTAAAACC	CCTTTGCTAATCGGGGCTAAAACT	TATCAAGTCTTATTCCTATTTTGGCATTTT
8	*ndhF*	**Pa.A.Km8**	[C/T]	AAAACCCTATTAATAGATACGAACACATC	CAAAACCCTATTAATAGATACGAACACATT	GTTACTAGTTCTAATTTGATMCAAATTTAT
9	*psbA*	**Pa.B.Km9**	[T/A]	CATATTCAGCTCCTGTTGCAGCA	CATATTCAGCTCCTGTTGCAGCT	CCTTGACCGATTGGGTAGATCAAGAA
10	*ndhE*	**Pa.B.Km10**	[C/T]	CGATAAATTGATGAAACAATAGCCAGC	ACGATAAATTGATGAAACAATAGCCAGT	KATTGCAGCCGCTGAAGCAGCTAT
11	*psbZ*	**Ar.A.Km11**	[A/G]	ACTTGACCAACCATCAGGAGAAGT	CTTGACCAACCATCAGGAGAAGC	ATCTTRTTAATTGGCGTACCCGTTGTATTT
12	*ndhK*	**Ar.A.Km12**	[G/T]	GATTCTTATAGTACTGTTCGGGGC	CTGATTCTTATAGTACTGTTCGGGGA	CCTGGCAAATAGACATCSACAGGAA
13	*atpA*	**Ar.B.Km13**	[A/G]	GCCAATTGATTCTGAGTAGCTTTATCA	CCAATTGATTCTGAGTAGCTTTATCG	GAAGCCTTTGCRCAATTTGCTTCTGAT
14	*psaB*	**El.Km14**	[A/G]	GGAATCCCCCGAGAAGGGTTAA	GAATCCCCCGAGAAGGGTTAG	CCCARGGAGCAGGAACTGCCAT
15	*psaA*	**El.Km15**	[T/G]	GCAACTTTTAATTTATTATGAGCCCAAACT	CAACTTTTAATTTATTATGAGCCCAAACG	GACGTGGTTATTGGCAAGAACTTATTGAA

Abbreviations: PaAr, *Panax–Aralia* group; Pa, *Panax* clade; Pa.A, *Panax* subclade (A); Pa.B, *Panax* subclade (B); Ar.A, *Aralia* subclade (A); Ar.B, *Aralia* subclade (B); El, *Eleutherococcus* clade.

Table 1-6. KASP genotyping results of tested samples

No.	Sample	KASP markers ID														
		Pa.Ar.Km1	PaAr.Km2	PaAr.Km3	Pa.Km4	Pa.Km5	Pa.Km_6	Pa.A.Km7	Pa.A.Km8	Pa.B.Km9	Pa.B.Km10	Ar.A.Km11	Ar.A.Km12	Ar.B.Km13	El.Km14	El.Km15
1	*P. ginseng*	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B
2	*P. quinquefolius*	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B
3	*P. japonicus*	A	A	A	A	A	A	B	B	A	A	B	B	B	B	B
4	*P. vietnamensis*	A	A	A	A	A	A	B	B	A	A	B	B	B	B	B
5	*P. notoginseng*	A	A	A	A	A	A	B	B	A	A	B	B	B	B	B
6	*P. stipuleanatus*	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B
7	*A. elata*	A	A	A	B	B	B	B	B	B	B	A	A	B	B	B
8	*A. cordata*	A	A	A	B	B	B	B	B	B	B	B	B	A	B	B
9	*E. sessiliflorus*	B	B	B	B	B	B	B	B	B	B	B	B	B	A	A
10	Other species	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B

Abbreviations: A, targeted genotype and B, non_targeted genotype

Figure 1-8. KASP genotyping of *Panax*, *Aralia* and *Eleutherococcus* species results. Screenshots of 15 KASP assay results show differentiation between target species (FAM-labeled, blue) with other species (HEX-labeled, green).

DISCUSSION

The potential utility of plastomes super-barcodes in phylogenetic studies of Araliaceae

The advanced high-throughput sequencing technologies have allowed the characterization of whole plastome sequences at a relatively low cost. Plastomes of the Araliaceae species have significantly increased in the public database over the last few years (roughly 2017–2021). The genetic information in the plastomes has also been extensively utilized in the molecular phylogeny of the Araliaceae. Our previous study assembled whole plastome sequences to perform a comparative analysis of ten Araliaceae species that were divided into two supergroups: the *Panax-Aralia* group and the *Eleutherococcus–Dendropanax* group (also known as the Asian Palmate group) (Kim et al., 2017). These two groups were also well-supported by a recent study using 29 Araliaceae (Valcárcel & Wen, 2019).

The current study enhanced the number of Araliaceae plastomes by including ten new plastomes from Araliaceae species and *C. asiatica* (Apiaceae). In the ML phylogenetic analysis, apart from the two well-supported groups (A and B in Fig. 1-3), three other groups such as *Polyscias* group, *Cheirodendron–Raukaua* group, and *Harmsiopanax–Hydrocotyle* group were determined (C, D, and E in Fig. 1-3), respectively. The phylogenetic analysis was also successful in accurately classifying nine newly assembled plastomes belonging to seven genera: *Aralia*, *Dendorpanax*, *Hedera*, *Fatsia*, *Schefflera*, *Tetrapanax*, and *Oplopanax* (Fig. 1-3). Besides, this study provided the first complete plastome sequence for *Polyscias fruticosa* (*Polyscias* genus, group C), which was the sister species of the group (B). Overall, the results demonstrated that complete plastomes as super-barcoding had a large number of informative sites for species identification (Fig. 1-2) and that they would be excellent tools for distinguishing closely

related Araliaceae species. Further analysis integrating high-throughput genome sequencing, morphological characteristics, and geographic distributions will achieve a better understanding of species delimitation and phylogenetic relationship in this problematic family.

Incongruence between plastome and 45S nrDNA phylogenies

For 22 Araliaceae species, phylogenies based on plastomes and the 45S nrDNA tandem repeat showed a similar result in the topology. The 45S nrDNA tree revealed poor branch supporting values for the Asian Palmate group, which had been observed in previous studies (Plunkett et al., 2004b; Plunkett et al., 1996; Valcárcel et al., 2014). Between phylogenies of plastomes and 45S nrDNA, phylogenetic incongruences in the positions of *P. stipuleanatus* and *O. elatus* were presented (Fig. 1-4A). Both species retained the common polymorphism observed in 45S nrDNA sequences (Fig. 1-4B), resulting in discordance of phylogenetic position with the plastome-based phylogeny. Unfortunately, this study was unable to trace the evolutionary fate of these species through time to provide direct evidence for explaining the observed phenomenon.

Interspecific or intergeneric hybridization (Fehrer et al., 2007; Rieseberg & Willis, 2007; Soltis & Soltis, 2009) and horizontal gene transfer (Bock, 2010; Gao et al., 2014) resulted in genomic alterations that may have contributed to the phylogenetic discordance. Furthermore, inconsistency between trees might arise not just due to hybridization/introgression but also as a result of incomplete lineage sorting (ILS). In ILS, ancestral polymorphisms persist during subsequent speciation, making the speciation and hybridization history more difficult to reveal (Charlesworth et al., 2005; Fehrer et al., 2007; Sousa & Hey, 2013). Therefore, research with extensive sampling and large-scale genomic data, such as complete nuclear and

mitochondrial genome sequences, would be recommended to improve the comprehension of the evolutionary processes of Araliaceae species.

Timing of diversification

In this study, the molecular clock estimation using 78 plastid CDSs demonstrated that most generic divergences of Araliaceae occurred between the Middle -Miocene to Pleistocene (Fig. 1-7). The Cenozoic Era spans the interval from 66 MYA to present. It is separated into two periods: the Paleogene (66–23 MYA) and the Neogene (23 MYA to present). Neogene epochs include the Miocene, Pliocene, Pleistocene, and Holocene (Fig. 1-7) (Selley et al., 2005). Miocene and Pliocene, 23.03 to 2.56 MYA, is a critical period in the evolution of the Earth's climate, i.e., from a greenhouse to an icehouse world (O'Brien et al., 2020).

The divergence time of *Panax_ Aralia* was consistent with the previous reports (Kim et al., 2017; Kim et al., 2018). Although the estimated times from another study (Valcárcel & Wen, 2019) were older than the estimated time in this study, the speciation times in Araliaceae were still estimated within the Miocene. However, around 50 genera of Araliaceae are found worldwide (Liu et al., 2012; Mitchell et al., 2012; Plunkett et al., 2004b; Valcárcel et al., 2014; Valcárcel & Wen, 2019), raising the migration argument for their origins. At the moment, our data do not permit us to conduct the hypotheses of intercontinental disjunctions in the distribution of the family. Phylogeographic analysis may be used to investigate the phylogenetic and geographic congruence of Araliaceae species. Additional phylogeographic research may shed light on the population genetic diversity across space and time.

Plastid super-barcodes

Araliaceae plastomes exhibited tremendous genus and species-specific

variants (Fig. 1-2). The nucleotide substitutions occurred in all 78 Araliaceae CDSs, and most plastid genes were under purifying selection; 73 genes had Ka < Ks (Fig. 1-5). Therefore, most plastid genes in the Araliaceae maintain high conservation. On the other hand, *acc*D, *clp*P, *rps*8, *ycf*2, and *ycf1*5 exhibited Ka> Ks, indicating that these genes could be undergoing adaptation to environmental conditions. Accelerated nucleotide substitution rates in acetyl-CoA carboxylase subunit β (*acc*D), ATP-dependent Clp protease proteolytic subunit (*clp*P), some subunits of ribosomal proteins (*rps*8), and the chloroplast factors ycf (*ycf*15 and *ycf*2) are frequently observed in plants (Guisinger et al., 2008; Jung et al., 2021; Li et al., 2018; Park et al., 2018; Park et al., 2017; Sloan et al., 2014). Numerous hypotheses have been advanced to account for this acceleration, including dysfunction of the DNA replication, repair, and recombination machinery, localized hypermutation, and pseudogenization (Guisinger et al., 2008; Magee et al., 2010; Weng et al., 2014). Without more data, it is impossible to determine whether there is a link between positively selected genes and morphology; however, these genes may serve as candidates for further comprehensive population genetics research to establish their functions.

The Araliaceae genetic-variant hotspots identified by assessing variation counts in 1,000 bp sliding windows over the plastomes may provide critical information for Sanger analysis of Araliaceae species. Intergenic areas were found to be more variable than coding regions, with 12 of 13 hotspots located in intergenic spacers: *trn*K–*rps*16, *rps*16–*trn*Q, *trn*S–*trn*G, *trn*E–*trn*T, *trn*C–*pet*N, *trn*T–*psb*D, *trn*T–*trn*L, *trn*L–*ndh*C, *psa*I gene–*ycf*4, *trn*R–*ycf*1 (Fig. 1-2). The genetic-variant hotspots found here would be a useful resource for future researchers and managers seeking precise information on the Araliaceae species diversity and ultimately contribute to creating effective conservation plans to protect endangered species in this family.

Four genes: *mat*K, *rpoc*2, *ycf*2, and *ndh*F, displayed a significant number of group-specific nonsynonymous variations (ψ) in over 23 genera (Fig. 1-6), suggesting that these genes could often be targets of natural selection during Araliaceae diversification. On the other side, the two genes *clpP* and *ndhI* could also serve as genomic markers for speciation between *Panax* and *Aralia*, and *Dendropanax* and *Chengiopanax*, respectively (Fig. 1-6).

Specific barcodes based on SNP sites

Panax, *Aralia*, and *Eleutherococcus* have adaptogenic characteristics, i.e., they enhance the capacity to overcome fatigue and stress, enhance memory and concentration, combat disease, and reduce the aging and degenerative processes (Brekhman & Dardymov, 1969; Navrátilová & Patočka, 2013). These genera species are of pharmaceutical and economic importance due to the large number of valuable therapeutic compounds they possess. This study attempted to develop a molecular taxonomic key for these Araliaceae genera using a simple and low-cost SNP marker technique. SNPs are the most frequent forms of genetic variation, appearing every 100–500 bp in the plastomes, and might be utilized as input data for identifying specific SNPs of Araliaceae (Table 1-7). The KASP approach to SNPs of plastome CDS enabled a graphical comparison of tested samples at a high resolution (Fig. 1-8). The KASP assays established in this research could be used to conduct initial observations on these significant taxa prior to developing species-specific identifications.

CONCLUSION

The presented work established a workflow for exploring plastome and 45S nrDNA super-barcoding in the Araliaceae family. A total of 74 accessions from 56 species representing 26 Araliaceae genera were investigated for phylogenetic analysis, genomic diversity, and species differentiation. The phylogenetic trees generated utilizing DNA super-barcoding would serve as a backbone for integrating molecular and morphological markers to characterize the diversity of Araliaceae species. Comparisons of sequence divergences across Araliaceae plastomes revealed 13 mutation hotspot regions that contributed to building a reference library to identify Araliaceae species. Regarding the amino acid mutations, the most Araliaceae group-specific nonsynonymous variations were in *mat*K, *rpoc*2, *ycf2*, and *ndh*F. The present study also demonstrated the power of phylogenetic approach with plastome sequences in conjunction with 45S nrDNA data to improve phylogenetic resolution. With the frequent reports of cytonuclear discordance and organelle genome capture in plants, the DNA barcoding community should progress toward including nuclear data into species discrimination. This work was based on an expanded sampling to explore the divergence patterns among the genera of Araliaceae and established diagnostic-genera SNPs. 15 KASP markers, diagnostic of three Araliaceae genera, were successfully developed with a very fine resolution. Imperatively, the robustness of these markets is a decisive advantage for identifying *Panax* species in markets. It will be highly beneficial in terms of protecting consumer health, providing adequate pre-selling warnings, managing ginseng cultivation properly, and protecting the species from overharvesting or habitat destruction.

REFERENCES

Allen, G., Flores-Vergara, M., Krasynanski, S., Kumar, S., & Thompson, W. (2006). A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nature protocols*, *1*(5), 2320-2325.

Bock, R. (2010). The give-and-take of DNA: horizontal gene transfer in plants. *Trends in plant science*, *15*(1), 11-22.

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., & De Maio, N. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS computational biology*, *15*(4), e1006650.

Brekhman, I., & Dardymov, I. (1969). New substances of plant origin which increase nonspecific resistance. *Annual review of pharmacology*, *9*(1), 419-430.

Brookes, A. J. (1999). The essence of SNPs. *Gene*, *234*(2), 177-186.

Brussell, D. E. (2004). Araliaceae species used for culinary and medicinal purposes in Niigata-ken, Japan. *Economic botany*, *58*(4), 736-739.

Calestani, V. (1905). Contributo alla sistematica: Ombrellifere D'Europa *Webbia*, *1*(1), 89-280.

Chang-Jiang, T., & Gin, H. (1982). A new classification scheme for the family Araliaceae. *Journal of Systematics and Evolution*, *20*(2), 125.

Charif, D., & Lobry, J. R. (2007). SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In *Structural approaches to sequence evolution* (pp. 207-232). Springer.

Charlesworth, B., Bartolomé, C., & NoëL, V. (2005). The detection of shared and ancestral polymorphisms. *Genetics Research*, *86*(2), 149-157.

Chase, M. W., Christenhusz, M., Fay, M., Byng, J., Judd, W. S., Soltis, D., Mabberley, D., Sennikov, A., Soltis, P. S., & Stevens, P. F. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, *181*(1), 1-20.

Fehrer, J., Gemeinholzer, B., Chrtek Jr, J., & Bräutigam, S. (2007). Incongruent plastid and nuclear DNA phylogenies reveal ancient intergeneric hybridization in *Pilosella hawkweeds* (Hieracium, Cichorieae, Asteraceae). *Molecular phylogenetics and evolution*, *42*(2), 347-361.

Fiaschi, P., Lowry, P. P., & Plunkett, G. M. (2020). Studies in Neotropical Araliaceae. III. Resurrection of the New World genus *Didymopanax* Decne. & Planch., previously included in *Schefflera* (Araliaceae). *Brittonia*, 1-7.

Gao, C., Ren, X., Mason, A. S., Liu, H., Xiao, M., Li, J., & Fu, D. (2014). Horizontal gene transfer in plants. *Functional & integrative genomics*, *14*(1), 23-29.

Group, C. P. W., Hollingsworth, P. M., Forrest, L. L., Spouge, J. L., Hajibabaei, M., Ratnasingham, S., van der Bank, M., Chase, M. W., Cowan, R. S., & Erickson, D. L. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences*, *106*(31), 12794-12797.

Guisinger, M. M., Kuehl, J. V., Boore, J. L., & Jansen, R. K. (2008). Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. *Proceedings of the National Academy of Sciences*, *105*(47), 18424-18429.

Hebert, P. D., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal*

Society of London. Series B: Biological Sciences, *270*(1512), 313-321.

Judd, W. S., Sanders, R. W., & Donoghue, M. J. (1994). Angiosperm family pairs: preliminary phylogenetic analyses. *Harvard papers in Botany*, 1-51.

Jung, J., Kim, C., & Kim, J.-H. (2021). Insights into phylogenetic relationships and genome evolution of subfamily Commelinoideae (Commelinaceae Mirb.) inferred from complete chloroplast genomes. *BMC genomics*, *22*(1), 1-12.

Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in bioinformatics*, *20*(4), 1160-1166.

Kim, K., Lee, J., Lee, S.-C., Kim, N.-H., Jang, W., Kim, S., Sung, S., Lee, J., & Yang, T.-J. (2016a). The complete chloroplast genome of *Eleutherococcus gracilistylus* (WW Sm.) SY Hu (Araliaceae). *Mitochondrial DNA Part A*, *27*(5), 3741-3742.

Kim, K., Lee, S.-C., Lee, J., Lee, H. O., Joh, H. J., Kim, N.-H., Park, H.-S., & Yang, T.-J. (2015a). Comprehensive survey of genetic diversity in chloroplast genomes and 45S nrDNAs within *Panax ginseng* species. *PloS one*, *10*(6), e0117159. https://doi.org/https://doi.org/10.1371/journal.pone.0117159

Kim, K., Lee, S.-C., Lee, J., Yu, Y., Yang, K., Choi, B.-S., Koh, H.-J., Waminal, N. E., Choi, H.-I., & Kim, N.-H. (2015b). Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of *Oryza* AA genome species. *Scientific Reports*, *5*, 15655.

Kim, K., Lee, S.-C., & Yang, T.-J. (2016b). The complete chloroplast genome sequence of *Dendropanax morbifera* (Leveille). *Mitochondrial DNA Part A*, *27*(4), 2923-2924.

Kim, K., Nguyen, V. B., Dong, J., Wang, Y., Park, J. Y., Lee, S.-C., & Yang,

T.-J. (2017). Evolution of the Araliaceae family inferred from complete chloroplast genomes and 45S nrDNAs of 10 *Panax*-related species. *Scientific Reports*, *7*(1), 4917. https://doi.org/10.1038/s41598-017-05218-y

Kim, N.-H., Jayakodi, M., Lee, S.-C., Choi, B.-S., Jang, W., Lee, J., Kim, H. H., Waminal, N. E., Lakshmanan, M., van Nguyen, B., Lee, Y. S., Park, H.-S., Koo, H. J., Park, J. Y., Perumal, S., Joh, H. J., Lee, H., Kim, J., Kim, I. S., . . . Yang, T.-J. (2018). Genome and evolution of the shade-requiring medicinal herb *Panax ginseng*. *Plant Biotechnology Journal*, *16*(11), 1904-1917. https://doi.org/10.1111/pbi.12926

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, *35*(21), 4453-4455.

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.

Li, J., Su, Y., & Wang, T. (2018). The repeat sequences and elevated substitution rates of the chloroplast accD gene in cupressophytes. *Frontiers in Plant Science*, *9*, 533.

Li, X., Yang, Y., Henry, R. J., Rossetto, M., Wang, Y., & Chen, S. (2015). Plant DNA barcoding: from gene to genome. *Biological Reviews*, *90*(1), 157-166.

Liu, Z., Zeng, X., Yang, D., Chu, G., Yuan, Z., & Chen, S. (2012). Applying DNA barcodes for identification of plant species in the family Araliaceae. *Gene*, *499*(1), 76-80.

Magee, A. M., Aspinall, S., Rice, D. W., Cusack, B. P., Sémon, M., Perry, A. S., Stefanović, S., Milbourne, D., Barth, S., & Palmer, J. D. (2010). Localized hypermutation and associated gene losses in legume

chloroplast genomes. *Genome research*, *20*(12), 1700-1710.

Mitchell, A., Li, R., Brown, J. W., Schönberger, I., & Wen, J. (2012). Ancient divergence and biogeography of *Raukaua* (Araliaceae) and close relatives in the southern hemisphere. *Australian Systematic Botany*, *25*(6), 432-446.

Navrátilová, Z., & Patočka, J. (2013). Arálie-rostlinné adaptogeny. *Kontakt*, *15*(1), 82-88.

Nguyen, B., Kim, K., Kim, Y.-C., Lee, S.-C., Shin, J. E., Lee, J., Kim, N.-H., Jang, W., Choi, H.-I., & Yang, T.-J. (2015). The complete chloroplast genome sequence of Panax vietnamensis Ha et Grushv (Araliaceae). *Mitochondrial DNA*, 1-2.

Nguyen, V. B., Giang, V. N. L., Waminal, N. E., Park, H.-S., Kim, N.-H., Jang, W., Lee, J., & Yang, T.-J. (2018). Comprehensive comparative analysis of chloroplast genomes from seven *Panax* species and development of an authentication system based on species-unique single nucleotide polymorphism markers. *Journal of Ginseng Research*. https://doi.org/https://doi.org/10.1016/j.jgr.2018.06.003

Nuraliev, M. S., Oskolski, A. A., Sokoloff, D. D., & Remizowa, M. V. (2010). Flowers of Araliaceae: structural diversity, developmental and evolutionary aspects. *Plant Diversity and Evolution*, *128*(1), 247.

O'Brien, C. L., Huber, M., Thomas, E., Pagani, M., Super, J. R., Elder, L. E., & Hull, P. M. (2020). The enigma of Oligocene climate and global surface temperature evolution. *Proceedings of the National Academy of Sciences*, *117*(41), 25302-25309.

Park, S., An, B., & Park, S. (2018). Reconfiguration of the plastid genome in *Lamprocapnos spectabilis*: IR boundary shifting, inversion, and intraspecific variation. *Scientific Reports*, *8*(1), 1-14.

Park, S., Ruhlman, T. A., Weng, M.-L., Hajrah, N. H., Sabir, J. S., & Jansen,

R. K. (2017). Contrasting patterns of nucleotide substitution rates provide insight into dynamic evolution of plastid and mitochondrial genomes of Geranium. *Genome biology and evolution*, *9*(6), 1766-1780.

Plunkett, G., Wen, J., & Lowry Ii, P. (2004a). Infrafamilial classifications and characters in Araliaceae: Insights from the phylogenetic analysis of nuclear (ITS) and plastid (*trnL-trnF*) sequence data. *Plant Systematics and Evolution*, *245*(1), 1-39.

Plunkett, G. M., Chandler, G. T., Lowry II, P., Pinney, S., Sprenkle, T., Van Wyk, B.-E., & Tilney, P. (2004b). Recent advances in understanding Apiales and a revised classification. *South African Journal of Botany*, *70*(3), 371-381.

Plunkett, G. M., Soltis, D. E., & Soltis, P. S. (1996). Higher level relationships of Apiales (Apiaceae and Araliaceae) based on phylogenetic analysis of *rbcL* sequences. *American Journal of Botany*, *83*(4), 499-515.

Qu, X.-J., Moore, M. J., Li, D.-Z., & Yi, T.-S. (2019). PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods*, *15*(1), 50.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic biology*, *67*(5), 901.

Rieseberg, L. H., & Willis, J. H. (2007). Plant speciation. *Science*, *317*(5840), 910-914.

Selley, R. C., Cocks, L. R. M., & Plimer, I. R. (2005). *Encyclopedia of geology*. Elsevier Academic.

Sloan, D. B., Triant, D. A., Forrester, N. J., Bergner, L. M., Wu, M., & Taylor, D. R. (2014). A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe Sileneae (Caryophyllaceae).

Molecular phylogenetics and evolution, 72, 82-89.

Soltis, P. S., & Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annual review of plant biology*, *60*, 561-588.

Sousa, V., & Hey, J. (2013). Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics*, *14*(6), 404-414.

Thorne, R. F. (1973). Inclusion of the Apiaceae (Unbelliferae) in the Araliaceae. *Edinb Roy Bot Gard Notes*.

Valcárcel, V., Fiz-Palacios, O., & Wen, J. (2014). The origin of the early differentiation of Ivies (*Hedera* L.) and the radiation of the Asian Palmate group (Araliaceae). *Molecular phylogenetics and evolution*, *70*, 492-503.

Valcárcel, V., & Wen, J. (2019). Chloroplast phylogenomic data support Eocene amphi-Pacific early radiation for the Asian Palmate core Araliaceae. *Journal of Systematics and Evolution*, *57*(6), 547-560.

Wen, J., Plunkett, G. M., Mitchell, A. D., & Wagstaff, S. J. (2001). The evolution of Araliaceae: a phylogenetic analysis based on ITS sequences of nuclear ribosomal DNA. *Systematic Botany*, 144-167.

Weng, M.-L., Blazier, J. C., Govindu, M., & Jansen, R. K. (2014). Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Molecular biology and evolution*, *31*(3), 645-659.

Wolfe, K. H., Li, W.-H., & Sharp, P. M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences*, *84*(24), 9054-9058.

CHAPTER II

Authentication of *Panax* collections

in Vietnam and Southern China

based on multi plastomes and 45S nrDNA

ABSTRACT

Due to illegal harvesting and economically motivated adulteration of *Panax vietnamensis* var. *vietnamensis*, *Panax* species native to Vietnam were designated as endangered, precious, and rare species. However, a lack of genetic resources and tools for molecular identification is one of the primary impediments to the conservation of *Panax* species in Vietnam. Seven *P. vietnamensis* individuals, representing three varieties: var *vietnamensis* (Pv_v), var. *fuscidiscus* (Pv_f), and var. *langbianesis* (Pv_l), were collected from Vietnamese habitats to generate NGS data. The phylogenetic trees of whole plastome and 45S nrDNA sequences from 34 *Panax* accessions supported nine major *Panax* clades. Pv_v and Pv_l exhibited almost identical plastome sequences and clustered within the *P. vietnamensis* clade. However, Pv_f was distinct from the *P. vietnamensis* clade and was classified as a member of the *P. zingiberensis* clade. 84 *Panax* plastomes with lengths ranging from 155,993 to 156,466 bp served as the ultimate source of variation, containing 3,007 SNPs and 491 InDels. 65 plastid genes shared among *Panax*, *Aralia*, and *Eleutherococcus* were under purifying selection. Additionally, 176 variation sites were identified in the 5,895 bp alignment of 76 45S nrDNA sequences. The significant intraspecific variation in two clades: *P. japonicus* clade; *P. vietnamensis* var. *fuscidiscus* and *P. zingiberensis* clade, implied ambiguous intra-clade speciation. Eight of nine *Panax* clades can be identified by 1,790 clade-specific SNPs. Sixteen *Panax* clades-specific SNPs were discovered and converted to KASP markers that genotyped 114 *Panax* specimens obtained in Vietnam with excellent precision. In practice, the marker system will aid in identifying counterfeit ginseng in the worldwide ginseng industry in general and in the Vietnamese ginseng market in particular.

Keywords: *Panax* species authentication, plastome, 45S nrDNA, KASP assay.

INTRODUCTION

Ginseng (*Panax*) is a shade-obligate perennial herb, classed under the Araliaceae family. A bicentric (disjunct) distribution pattern is observed in the genus, comprising about 22 taxon/taxa (Hou et al., 2021; Secretariat, 2019; Zhengyi, 1983; Zuo et al., 2017). Among these species, *P. ginseng*, *P. japonicus*, *P. wangianus*, *P. vietnamensis*, *P. zingiberensis*, *P. notoginseng*, and *P. stipuleanatus* are distributed in Eastern Asia, while *P. quinquefolius* and *P. trifolius* are distributed in North America. *Panax* species have been used in Asian traditional medicine for over a thousand years and are highly valued for their ginsenosides, a class of unique secondary metabolites with a broad range of pharmacological activity in preserving health and treating disease (Bai et al., 2018). While the amount of cultivated area for ginseng is increasing in several nations, including China, Korea, Japan, Vietnam, Russia, the United States, and Canada, it is still inadequate to reach world demands of ginseng. Consequently, illegal exploitation of the wild *Panax* population has increased dramatically. Thus, *Panax* species have been threatened to survive in the wild. CITES (https://cites.org/eng) designated *P. ginseng* (in 2000) and *P. quinquefolius* (in 1975) under Appendix II, in which listed species are at risk of extinction if the trade is not regulated.

In Vietnam, investigations on *Panax* species started in 1973 (Komatsu et al., 2001). Three major Asiatic ginseng are recorded: *P. vietnamensis* (three varieties: var. *vietnamensis* Ha et Grushv., var. *fuscidiscus* K. Komatsu, S. Zhu et S. Q. Cai, and var. *langbianesis* N.V. Duy. V.T. Tran & L.N. Trieu), *P. stipuleanatus*, and *P. bipinnatifidus* (Le et al., 2018; Long et al., 2020; Nong et al., 2016a). It is currently more expensive to get *P. vietnamensis* var. *vietnamensis* (Ngoc Linh ginseng) (price ranges from 1000 to 3000 USD/1kg) (Le et al., 2018). Due to pricing pressures on suppliers of Ngoc Linh ginseng,

other varieties of *P. vietnamensis* or other *Panax* species have often been claimed to be fraudulently labeled as *P. vietnamensis* var. *vietnamensis*. Consequently, illegal overharvesting and habitat damage have accelerated the decline of *Panax* species, which are recognized in the Vietnam Red Book (2007) and classified as endangered, precious, or rare species that require special protection (Decree 06/2019/ND-CP). In light of this, promoting research and technology transfering in the identification of *Panax* species is crucial to the conservation of *Panax* species in Vietnam (Ho & Pham, 2020). Research on *Panax* species in Vietnam has been conducted on morphological and biochemical levels (Bon et al., 2019; Duc et al., 1993; Dung & Grushvitski, 1985; Phan et al., 2013; Yamasaki, 2000), but only a few molecular studies have been established (Ngoc et al., 2020; Nong et al., 2016a; Nong et al., 2016b; Vu et al., 2020; Zhang et al., 2015).

The genetic-based management system offers critical insight and accuracy to conservation efforts (Coates et al., 2018). Recently, numerous molecular approaches have been applied to characterize genetic variation in *Panax* species (Goodwin & Proctor, 2019). Significantly, the evolution of Next-generation sequencing (NGS) technology provided high-throughput information of nuclear ribosomal DNA (nrDNA) and chloroplast (plastid) genome (plastome) sequences of *Panax* species (Ji et al., 2019; Kim et al., 2015a; Kim et al., 2017; Kim et al., 2018; Liu et al., 2018; Nguyen et al., 2018). The NGS strategy has the potential to inspire the development of numerous efficient molecular techniques for species discrimination, thereby enhancing our knowledge of *Panax* species diversity, phylogeny, and evolution. Chloroplast DNA (plastomes) is different from nuclear DNA in that it is maternally inherited and highly conservative (Birky, 1995). Previous studies reported *Panax* plastomes size between 155,993 bp to 156,466 bp (Ji et al., 2019; Nguyen et al., 2018). *Panax* plastomes are circular double-

stranded DNA molecules organized similarly to most plants. They contain two copies of inverted repeats (IRs) that divide the plastome into a large single copy (LSC) region and a small single copy (SSC) region. Twelve large mitochondrial plastid DNAs (MTPTs) ranging in size from 2,297 to 8,250 bp were discovered in the plastome of *P. ginseng* (Jang et al., 2020; Nguyen et al., 2018). These MTPTs could result in inaccurate inferences (Park et al., 2020) when evaluating genetic variation between *Panax* species using a plastome-specific polymorphism and should be minimized.

An earlier study successfully authenticated seven *Panax* species by utilizing 18 dCAPS markers unique to each *Panax* species (Nguyen et al., 2018). Unfortunately, the Pvdm13 and 14 markers were insufficient to differentiate *P. vietnamensis* varieties. Higher-resolution markers for *P. vietnamensis* varieties should be established. In this study, we obtained NGS data of seven *P. vietnamensis* varieties (2 var. *vietnamensis*, 2 var. *langbianesis*, and 3 var. *fuscidiscus*), assembled their complete plastomes and 45S nrDNA sequences, and performed a comparative analysis with other *Panax* plastomes and 45S nrDNA sequences derived from NCBI (National Center for Biotechnology Information) (https://www.ncbi.nlm.nih.gov/). The phylogenetic relationships among *Panax* species were investigated. KASP genotyping assays (http://www.lgcgenomics.com) were adopted for high-throughput analysis. 2 new dCAPS markers developed based on the extensive survey of complete *Panax* plastomes will be helpful to produce a systematic approach in the authenticity of *P. vietnamensis* varieties.

MATERIAL AND METHODS

Sample collection, DNA extraction, and sequencing

Between 2016 and 2020, leaf and root tissues of wild and cultivated *Panax* species were collected in Vietnam. 114 specimens from six provinces were primarily identified through leaf and root morphology or phytochemical investigation. In Figure S2-1, representative photos of several specimens are shown. Table S2-1 contains the accession numbers, geographic origin, and primary classification of each specimen.

Genomic DNA was isolated from leaves and roots using a modified cetyltrimethylammonium bromide (CTAB) technique (Allen et al., 2013). A NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA) and agarose gel electrophoresis were used to determine the quality and quantity of isolated genomic DNA. Additionally, genomic DNA obtained from previously published samples (Nguyen et al., 2018) was employed.

Total genomic DNA from two samples of *P. vietnamensis* var. *vietnamensis* (KT1 and DL7), two samples of *P. vietnamensis* var. *langbianesis* (DL34 and DL46), and three samples of *P. vietnamensis* var. *fuscidiscus* (LC1, SP9, and DL1) were sequenced on the Illumina MiSeq platform in paired-end mode 2 × 300 bp (overlapping reads) by PHYZEN (Seongnam, South Korea).

Illumina sequencing, assembly, and annotation of plastomes and 45S nrDNA

The raw Illumina paired-end sequence reads were first trimmed of low-quality reads and adaptor sequences to obtain high-quality reads. The CLC Genomics Workbench version 4.21 (https://digitalinsights.qiagen.com) and dnaLCW protocol (Kim et al., 2015b) were utilized to assemble plastome and 45S nrDNA sequences. Scaffolds corresponding to plastome and 45S nrDNA

sequences were separated using a BLASTZ tool (http://phyzen.iptime.org/tools/cv.php) with *P. vietnamensis* (GenBank: KP036470.1 and KT380922.1) set as reference genomes. The final plastome sequences were generated after orienting and sorting the representative scaffolds. Plastid Genome Annotator (PGA) package (Qu et al., 2019) was used to annotate the genes of the chloroplasts, ribosomal RNAs, and transfer RNAs. Following that, manual comparisons with the reference plastome were proceeded to correct protein-coding sequences (CDS) (star and stop codon) and exon/intron boundaries. Finally, circular plastome maps of *P. vietnamensis* varieties were drawn via the OGDRAW program (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html). In the case of 45S nrDNA, boundaries of 18S, Internal transcribe spacer 1 (ITS1), 5.8S, ITS2, and 26S regions were identified by manual comparisons with the reference genome (KT380922.1).

Phylogenetic analysis

Phylogenetic analysis was performed on plastomes and 45S nrDNA sequences of 33 *Panax* accessions (**Supplementary Table S2-2**). The alignments of plastomes and 45S nrDNA sequences were computed to SeaView(Gouy et al., 2010) that drives the PhyML v3.1 program (Guindon & Gascuel, 2003) to draw the maximum likelihood (ML) phylogenetic trees with 1,000 bootstrap replications.

Variations in plastome and 45S nrDNA sequences

We investigated the nucleotide diversity of *Panax* species using 84 *Panax* complete plastome sequences and 76 *Panax* 45S nrDNA sequences. First, the newly completed plastomes and 45S nrDNA sequences of *P. vietnamensis* varieties in this study were aligned with the plastomes and 45S nrDNA sequences of *Panax* species available in NCBI. The alignments were

performed by MAFFT v7 (Katoh et al., 2019) (https://mafft.cbrc.jp/alignment/server) and edited by BioEdit (Hall, 1999). The alignments were imported into Excel and analyzed using the COUNTIFS function for nucleotide variants. The histogram of SNPs and InDels distribution across 84 plastomes measured in windows of 10 bp were visualized in the Circos plot (Krzywinski et al., 2009). On the other hand, the 45S nrDNA nucleotide variants distribution in a sliding window of 100 bp along the sequences was summarized and graphically illustrated.

Analysis of single nucleotide polymorphisms (SNPs) across the *Panax* plastomes

The pairwise number of SNP differences (Δ) designed to describe the genetic distance among 84 *Panax* plastomes was a tool to capture the correlation between genetic distances and phylogenetic relationships. Δ is defined as the number of SNP differences between 2 plastomes. Two *Panax* plastome sequences with smaller values of Δ (Δ = 0 being the smallest) were regarded to be more closely related taxa. Snp-dists (version 0.6.3, https://github.com/tseemann/snp-dists) program was used to calculate Δ. The output was organized according to the topology of ML phylogenetic analyses of 84 *Panax* plastomes and was visualized as a heatmap created by the ComplexHeatmap package (https://jokergoo.github.io/ComplexHeatmap-reference/book). Δ_{intra} and Δ_{inter} are intra- and interspecific pairwise SNP differences values according to the level of taxonomic categories. Density plots of Δ_{intra} and Δ_{inter} for plastomes from same species and all 84 were produced using the ggridges package (https://cran.r-project.org/web/packages/ggridges/).

In Excel, the COUNTIF function was used for filtering *Panax* clade-specific SNPs. The distribution of *Panax* clade-specific SNPs along the

plastome was plotted using Circos (http://circos.ca/).

Rate of substitution

In the investigation of the patterns of nucleotide substitution rates, a total of 78 protein-coding genes from 84 complete *Panax* plastome sequences were collected and aligned using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo). Pairwise comparisons of nucleotide substitution rates (Ka: nonsynonymous substitution and Ks: synonymous substitution) were performed on each of the 78 aligned protein-coding genes using the R package seqinr (Charif & Lobry, 2007). Fivenum() function in R was used to return Tukey's five-number summary of Ka and Ks. A bar graph was used to display Ka and Ks median values (GraphPad Prism software, https://www.graphpad.com/scientific-software/prism/).

Development and validation of KASP genotyping assays

The *Panax* clade-specific SNPs in coding regions that do not overlap with MTPT or IR regions were potentially positioned to develop KASP markers. From the 18 dCAPS markers developed in the previous study (Nguyen et al., 2018), SNP loci that met the KASP assay criteria were collected to convert into KASP markers. Additional KASP markers were developed to ensure that each *Panax* clade possessed at least two KASP markers indicating species differentiation. A total of 14 KASP markers were produced (50-bp left and 50-bp right-flanking sequence of each *Panax* clade-specific SNP site were used to design two allele-specific forward primers and a common reverse primer). KASP primers were produced by LGC Biosearch Technologies (Teddington, United Kingdom), and experiments were done following their standard protocol (https://biosearch-cdn.azureedge.net/assetsv6/running-KASP-on-Roche-LC480.pdf).

Development and validation of dCAPS markers

To further assess the diversity of *P. vietnamensis* varieties, we utilized dCAPS Finder 2.0 (http://helix.wustl.edu/dcaps/dcaps.html) to develop 2 dCAPS markers that are species-specific for the group of *P. vietnamensis* var. *fuscidiscus* and *P. zingiberensis* (see Results). dCAPS marker genotyping was performed in a 25 µl reaction mixture: 2 µl 10 ng/µl template DNA, 1 µl 10 ng/ml of each forward and reverse primer, 2 µl 10 × PCR buffer with $MgCl_2$, 0.5 µl 10 mM of dNTP, and 0.5 U Taq DNA polymerase. PCR amplification conditions were 95°C for 5 min; [95°C for 30 s., 58–60°C (depending on primers) for 30 s. and 72°C for 30 s.] x 35 cycles and finished with a final elongation at 72°C for 7 min. The PCR products were digested with 0.2 Unit restriction enzyme (NEB) recognizing dCAPS sites overnight at 37°C. The digested products were recorded after electrophoresed on 3% agarose gel added Safety Gel stain in 0.5 × TBE buffer.

Authentication of *Panax* collections in Vietnam and Southern China

To demonstrate the efficacy of the established clade-specific SNP markers, 114 *Panax* specimens collected from Vietnam and Southern China were evaluated using the 14 KASP and two dCAPS markers. The results were displayed in the form of a bar chart, with specimens from each collection region categorized into the appropriate *Panax* clade.

RESULTS

Assembly and annotation of the plastome and 45S nrDNA sequences of seven *P. vietnamensis* varieties

This study successfully sequenced and assembled the complete plastomes and 45S nrDNA sequences of seven *P. vietnamensis* varieties. Each of the seven plastomes was assembled into a circular structure with a similar GC content (38%) and length (155,984–156,465 bp) (Table 2-1). In comparison to previously released *Panax* plastomes (Ji et al., 2019; Kim et al., 2015a; Kim et al., 2017; Kim et al., 2018; Liu et al., 2018; Nguyen et al., 2018), these plastomes exhibited a remarkable degree of conservation in terms of size, structure, gene content, and gene order. Their gene repertoires encompassed a set of 115 genes, including 80 protein-coding genes, 4 ribosomal RNA genes, and 31 tRNA genes (Fig. 2-1A; Table S1-2).

The 45S nrDNA sequences of seven *P. vietnamensis* varieties assembled into a unit length of 6,400 bp (Table 2-1), including the 5,874 bp coding region (18S-ITS1-5.8S-ITS2-26S) (Fig. 2-1B) and the 526 bp partial intergenic spacer region (IGS).

Table 2-1. Samples information and summary of plastomes and 45S nrDNAs characteristics for two *P. vietnamensis* var. *vietnamensis*, two *P. vietnamensis* var. *langbianensis*, and three *P. vietnamensis* var. *fuscidiscus*.

***P. vietnamensis* var.**		*vietnamensis*[a]	*vietnamensis*	*vietnamensis*	*langbianensis*	*langbianensis*	*fuscidiscus*	*fuscidiscus*	*fuscidiscus*
Accession IDs[b]			KT1	DL7	DL34	DL46	LC1	SP9	DL1
Filtered NGS Reads (Mbp)		4586	3180	1010	1890	1019	2350	2070	2990
CP Assembly	**GenBank**	KP036470.1	KP036471.1	KU059178.1	MT798583	MT798584	MT798585	MT798586	MT798587
	Coverage (X)	1005	259	150	41	18	370	141	255
	Total Length (bp)	155993	155992	155993	155984	155986	156284	156310	156465
	LSC (bp)	86180	86179	86180	86172	86172	86171	86198	86172
	IRA/B (bp)	25942	25942	25942	25941	25941	26071	26072	26063
	SSC (bp)	17929	17929	17929	17930	17930	17971	17968	17968
45S nrDNA Assembly	**GenBank**	KT380922.1	MW374467.1	MW374330.1	MW374329.1	MW374462.1	MW374464.1	MW374471.1	MW374463.1
	Coverage (X)	2267	542	265	558	3008	2810	2417	1046
	Total Length (bp)	7280	6400	6400	6400	6400	6400	6400	6400
	18S	1808	1808	1808	1808	1808	1808	1808	1808
	5.8S	160	160	160	160	160	160	160	160
	26S	3450	3450	3450	3450	3450	3450	3450	3450

Abbreviations: [a] Sequences obtained in our previous studies (Kim et al., 2017). [b] *Panax* accessions ID in Table S2-1, KT: Kon Tum province, LC: Lai Chau province, SP: Sa Pa district, and DL: Da Lat city.

Sequences variant distribution in *Panax* plastomes and 45S nrDNA

The alignment of 45S nrDNA coding units from 76 *Panax* accessions (7 assembled in this study and 69 from GenBank) showed 97% sequence similarity and a total of 179 nucleotide variant sites (3%) (10 in 18S; 37 in ITS1; 5 in 5.8S; 48 in ITS2 and 79 in 26S) were detected (Fig. S2-2). The ITS1 and ITS2 subregions had a high rate of variations per 100bp; thus, ITS regions could be used to characterize *Panax* species utilizing sequencing methods or DNA markers (Fig. 2-1). 179 variant sites were identified, 17 from individual accessions and 79 from parallel variants in different species. The remaining 80 variants sites were species-specific polymorphisms, including 3 for *P. ginseng*; 1 for *P. quinquefolius*; 1 for *P. wangianus*; 1 for *P. vietnamensis*; 2 for *P. zingiberensis*; 17 for *P. notoginseng*; 19 for *P. stipuleanatus,* and 36 for *P. trifolius*. *P. japonicus* did not exhibit any species-specific polymorphisms (Fig. S2-1). Thus, 90% (72 of 80) of species-specific polymorphisms were described in three species: *P. notoginseng*, *P. stipuleanatus*, and *P. trifolius*. The 18S-ITS1-5.8S-ITS2-26S unit region could not establish species-specific molecular markers for all *Panax* species.

The comparative analysis of 84 *Panax* plastomes (7 in this study) (Table S1-1) revealed 7789 nucleotide variant sites. These variants were distributed as follows: 5,061 in LSC, 1,178 in SSC, and 775 in each IR and categorized into 3007 SNPs and 491 InDels (1 bp to 292 bp) (Fig. 2-1A).

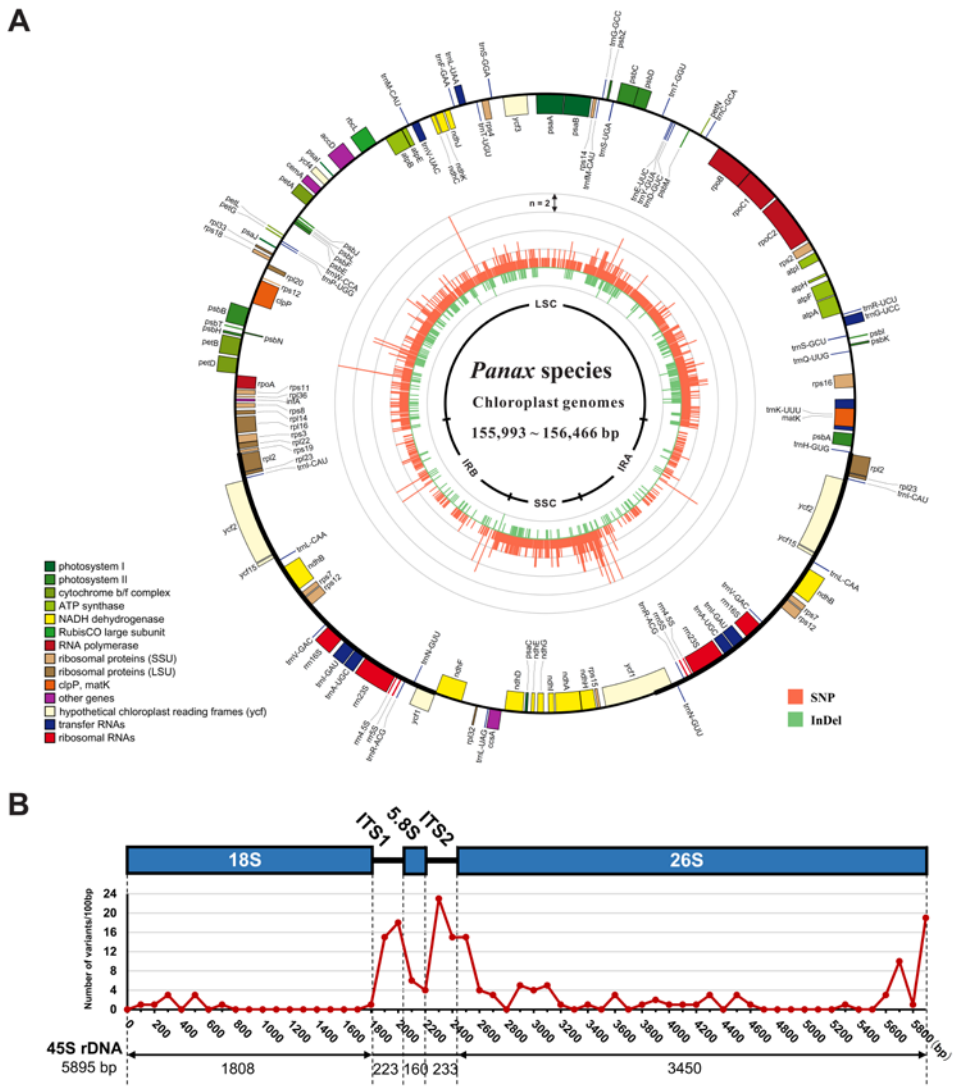

Figure 2-1. An overview of genome variation across the *Panax* plastomes and 45S nrDNA. (A) Circles display: (1) the outer circle is the gene map of the *Panax* chloroplast genomes, genes belonging to different groups are coded with different colors, genes transcribed clockwise are shown inside, and those transcribed counterclockwise are outside; (2) the middle circle is Circos histograms displays the counted number of SNPs (red) and InDels (green) per 10 bp along the 84 complete chloroplast genome sequences from 9 *Panax* species. (B) The graph illustrates the number of variations per 100bp based on their presence in the alignment of 76 *Panax* nuclear 18S-ITS1-5.8S-ITS2-26S unit accessions.

Phylogenetic analysis

Maximum likelihood (ML) phylogenetic analysis of plastome and 45S nrDNA sequences generated similar topologies for 33 *Panax* accessions (Table S2-2), in that they showed support to the monophyly of nine major clades: *P. ginseng* clade (1), *P. quinquefolius* clade (2), *P. japonicus* clade (3), *P. wangianus* clade (4), *P. vietnamensis* clade (5), *P. zingiberensis* clade (6), *P. notoginseng* clade (7), *P. stipuleanatus* clade (8), and *P. trifolius* clade (9) (Fig. 2-2). However, incongruences between the plastome-based tree and 45S nrDNA-based trees were detected, similar to the results deduced in previous *Panax* phylogenetic studies (Ji et al., 2019; Kim et al., 2017). The placement of three clades: (3), (4), and (7) in the plastomes ML phylogenetic tree differed from the tree based on 45S nrDNA sequences. In addition, the topological inconsistencies were also presented by accessions within the clade (3) and clade (6) (Fig. 2-2). Plastomes of *P. vietnamensis* var. *vietnamensis* and var. *langbianensis* were identical and shared the same clade (5), while the three plastomes of *P. vietnamensis* var. *fuscidiscus* (MT798587, MT798586, and MT798585) were clustered with *P. zingiberensis* in the clade (6) (Fig. 2-2).

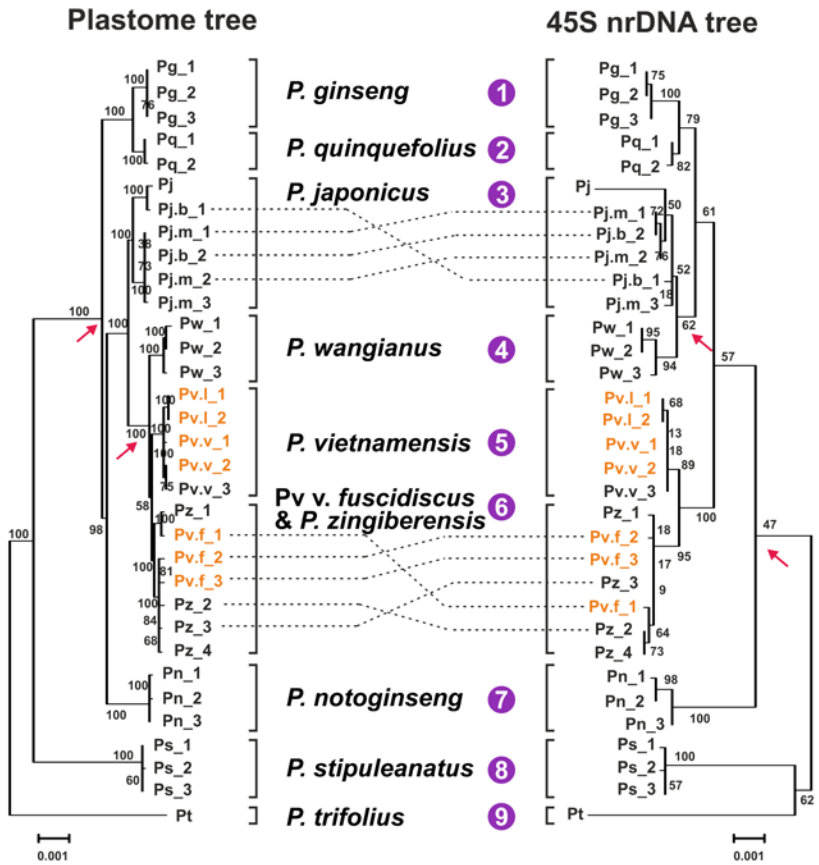

Figure 2-2. Phylogeny of *Panax* plastomes and 45S nrDNA. ML phylogenies with bootstrap support values summarizing relationships of 9 major *Panax* based on the analysis plastomes and 45S nrDNA of 33 *Panax* accessions. Numbered circles from 1 to 9 indicate 9 *Panax* clades: (1) Pg, *P. ginseng*; (2) Pq, *P. quinquefolius*; (3) Pj, *P. japonicus* with 2 varieties: b, *bipinnatifidus* and m, *major*; (4) Pw, *P. wangianus*; (5) Pv, *P. vietnamensis* with 3 varieties: v, *vietnamensis*; l, *langbianensis*; f, *fuscidiscus*; (6) Pz, *P. zingiberensis*; (7) Pn, *P. notoginseng*; (8) Ps, *P. stipuleanatus*; (9) Pt, *P. trifolius*. Dashed lines connected lineages between the two phylogenies. The seven *Panax* accessions sequenced and assembled in this study were highlighted in orange. Red arrows denoted the conflicting topologies obtained from plastomes and 45S nrDNA.

The matrix of pairwise SNP differences (Δ) and the ML phylogenetic analysis of *Panax* plastomes

Association analysis between Δ matrix and the ML phylogenetic tree was conducted using the set of 84 *Panax* plastomes (Fig. 2-3). The *Panax* plastomes were clustered into 9 well-established clades (Fig. 2-2). According to the ML phylogenetic classification, the intra-clade Δ (Δ_{intra}) ranged from 0 to 158, while the inter-clade Δ (Δ_{inter}) ranged from 111 to 1,573 (Fig. 2-3). Clade (8) and (9) were the most diverse, as evidenced by the high value of the minimum Δ_{inter} (1,112 and 1,469, respectively). The minimum Δ_{inter} of clades (1) ~ (7) was 111, 111, 270, 137, 134, 134 and 435, respectively. Clade (1) was sister to the clade (2), given that their minimum Δ_{inter} was equal. Clade (4) was sister to (clade (5) + clade (6)), and their minimum Δ_{inter} also were identical. While the maximum Δ_{intra} of 6 of 9 *Panax* clades (no Δ_{intra} for the clade (9)) were lower than 60, the clades (3) and (6) had the highest maximum Δ_{intra} values (158 and 95, respectively). Therefore, as illustrated in Figure S2-3, there was an overlap between intra- and interspecific SNP distances when 84 Panax plastomes were included. Thus, an upper limit for Δ_{intra} and a lower limit for Δ_{inter} could not be determined, prohibiting the establishment of threshold value-based plastomes for *Panax* species identification.

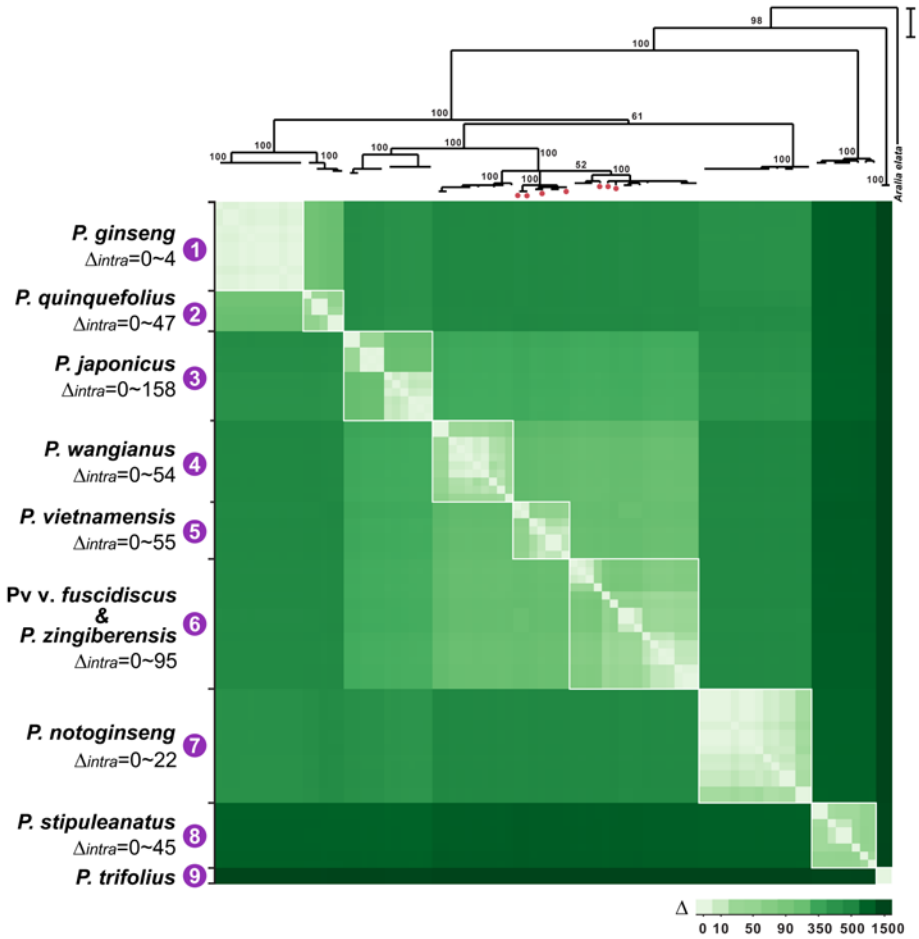

Figure 2-3. Combining the phylogenetic trees constructed by the *Panax* whole plastomes with the pattern of pairwise SNP differences. The number of SNPs (Δ) found in pairwise analyses of 84 plastome sequences was interpreted as a distance matrix for the heat map. Color in the heat map representing Δ corresponded to the values specified in the legend at the bottom right. The values Δ ranged from 0 to 1573. Δ_{intra} was estimated as the value of pairwise SNP distances within *Panax* clades (9). The seven chloroplast genomes sequenced and assembled in this study were highlighted by red dots in the ML phylogenetic tree shown to the top.

Nucleotide substitutions in plastid genomes of *Panax* species

Rates of nucleotide substitutions (Ka and Ks) were estimated by comparing 78 plastid-encoded genes (Fig. 2-4) shared among 84 plastomes from nine *Panax* clades. At first glance, nucleotide substitution rates greatly favored synonymous (silent) substitutions (Ks), and IR genes had significantly lower nucleotide substitution rates than single-copy regions (LSC and SSC) (Fig. 2-4). Mean of Ka values less than 0.0067 and mean of Ks values less than 0.0168. In terms of conservation of coding sequences across *Panax* species, the median value of Ka and Ks was 0.000 in 36 genes and four genes were completely conserved with 0.00 of both Ka and Ks, 3 photosystem II genes (*psbF*, *psbT,* and *psbN*), and 1 ribosomal protein (*rpl23*) (Table S2-3). Six genes were shown to have no sequence variation in a previous study: *petN*, *psaJ*, *psbF*, *psbN*, *psbT,* and *rpl23,* showed no sequence variation (Kim et al., 2017). However, sequence differences were discovered in *pet*N gene of *P. trifolius* and the *psa*J gene of *P. wangianus*, two species having recently submitted plastome sequences to GenBank. Finally, the *psbF*, *psbN*, *psbT*, and *rpl23* genes were conserved among *Panax* species. Six genes: *atpF*, *accD*, *rpoA*, *rpl2*, *ycf2*, and *rps15,* had median Ka values greater than Ks values. Five genes: *mat*K, *rps*4, *pet*L, *ndhE,* and *ndh*H, could be addressed as coding sequence hotspots of nucleotide substitutions and could be utilized to develop markers for *Panax* species identification (Fig. 2-4).

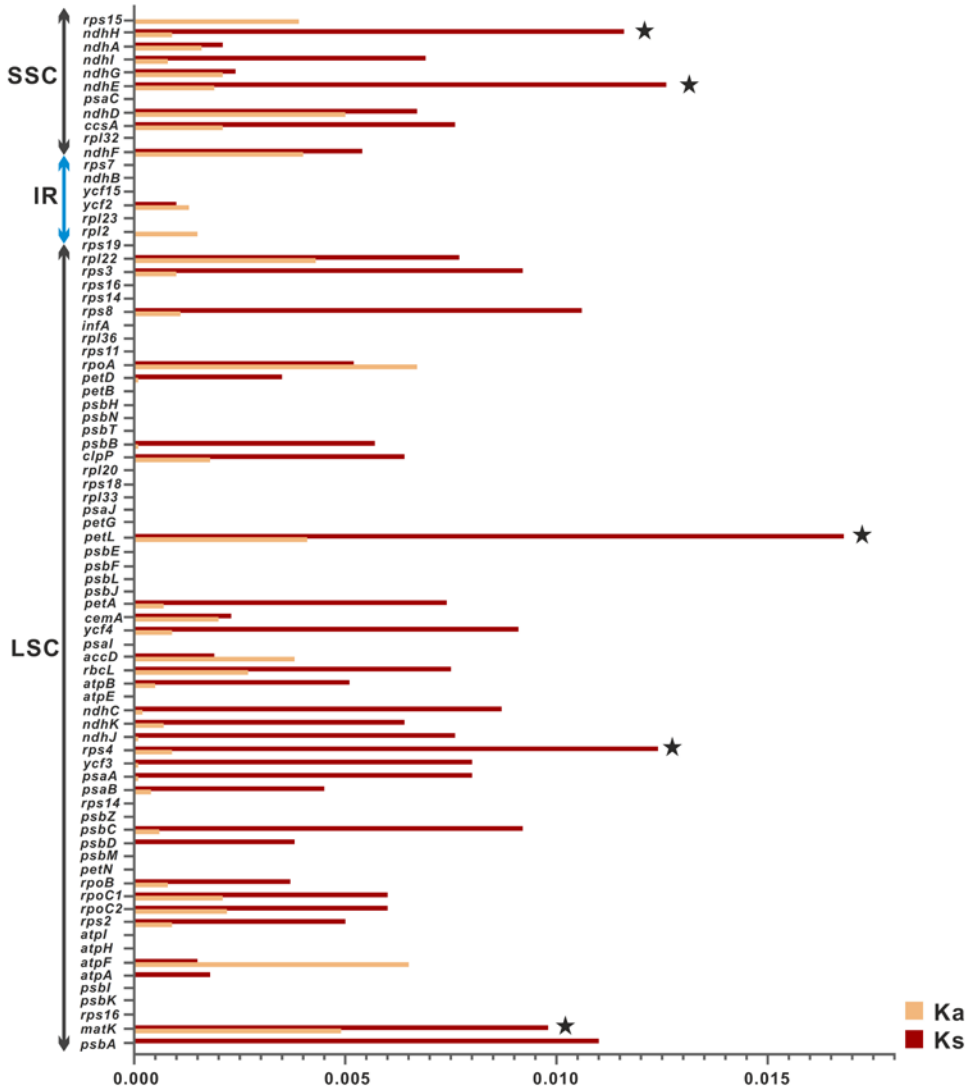

Figure 2-4. Nucleotide substitution rates. The bar graph presents the median of nonsynonymous substitutions (Ka) and synonymous (Ks) substitutions. Black stars correspond to the hotspots of nucleotide substitutions.

Clade-specific SNPs analysis in *Panax* genus

Among 3,007 SNPs identified in plastomes (Fig. 2-1), 1,790 SNPs (59.5%) satisfied the *Panax* clade-specific SNPs criteria. Their physical placement and distribution within the plastome sequences of 9 *Panax* clades were represented in Fig. 2-5. The *P. trifolius* clade had the highest number of clade-specific SNPs (866), followed by the *P. stipuleanatus* clade (497) and the *P. notoginseng* clade (185). In contrast, only a small number of clade-specific SNPs were found in *P. ginseng* (42), *P. quinquefolius* (44), *P. japonicus* (32), *P. wangianus* (50), *P. vietnamensis* (54), and *P. zingiberensis* (20) clades.

896 (50.1%) of the 1,790 *Panax* clade-specific SNPs were in coding regions (CDSs). However, 393 *Panax* clade-specific SNPs were identified in 26 genes (*psb*A, *mat*K, *psb*D, *psb*C, *atp*E, *atp*B, *rbc*L, *rps*12, *pet*D, *rpo*A, *rps*11, *rpl*36, *inf*A, *rps*8, *rpl*14, *rpl*16, *rps*3, *rpl*22, *rps*19, *rpl*2, *rpl*23, *ycf*2, *ycf*15, *ndh*B, *rps*7, and *ycf*1) located in the IR and the MTPT regions, where false-positive amplification may occur (Jang et al., 2020; Nguyen et al., 2018; Park et al., 2020).

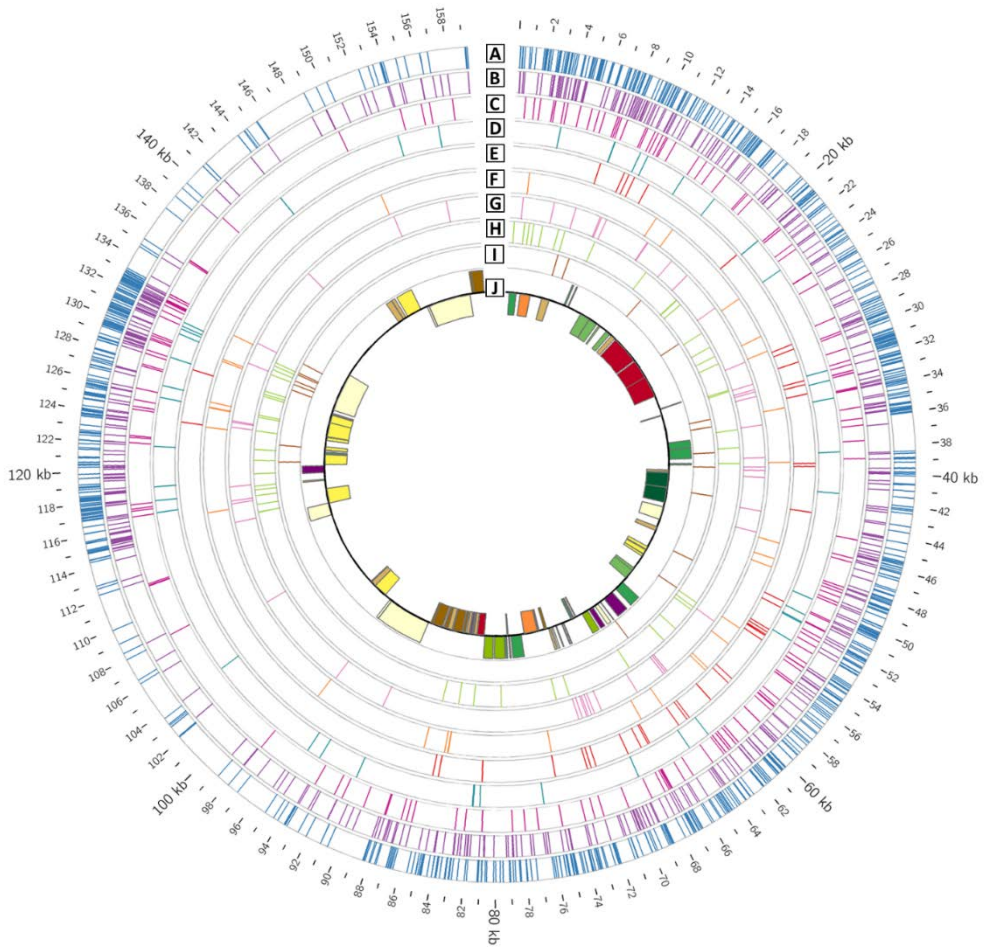

Figure 2-5. Circos plot distribution of 1,790 *Panax* clade-specific SNP identified from 84 *Panax* plastomes. (A – I) Distribution of unique SNPs recorded in the plastomes of *P. trifolius* clade (866), *P. stipuleanatus* clade (497), *P. notoginseng* clade (185), *P. quinquefolius* clade (44), *P. ginseng* clade (42), *P. japonicus* clade (32), *P. wangianus* clade (50), *P. vietnamensis* clade (54), and *P. zingiberensis* clade (20), respectively. The unique SNPs are represented by lines in different colors for each clade. (J) Plot of coding sequences (CDS).

Development and validation of KASP and dCAPS markers

Due to the release of additional plastomes of *Panax* species in 2019 ((Ji et al., 2019) and in this study, the 18 SNPs that were thoroughly investigated to generate dCAPS markers in the previous study (Nguyen et al., 2018) were re-evaluated. Among these, 15 SNPs were still appropriate for converting to KASP assays that can identify 6 *Panax* clades: (1), (2), (5), (7), (8), and (9). We ended up with 11 SNPs and added 3 more *Panax* clade-specific SNPs to develop a total of 14 KASP markers (Table 2) for the identification of 7 clades: PgKm1 and 2 for clade (1); PqKm3 and 4 for clade (2); PjKm5 and 6 for clade (3); PvKm7 and 8 for clade (5); PnKm9 and 10 for clade (7); PsKm11 and 12 for clade (8); and PtKm13 and 14 for clade (9). In terms of authentication of *P. vietnamensis* varieties, 2 dCAPS assays, Pv.fdm19 and 20 (Table 3) had been employed in primary examining genotypic distinctions of clade (6) (*P. zingiberensis* and *P. vietnamensis* var. *fuscidiscus* complex). 14 KASP and 2 dCAPS assays successfully distinguished 8 *Panax* clades (Table 2-4, Fig. 2-6).

Table 2-2 Primer sequences of 26 KASP markers used to authenticate seven *Panax* clades.

Gene	KASP_ID	Allele_FAM (5'-3') (F)	Allele_HEX (5'-3') (H)	Common/reverse primer	SNP	dCAPS_ID*
rpl20	**PgKm1**	GATTAAGAAGTAACTGTTTCTTATACAGATCA	AAGAAGTAACTGTTTCTTATACAGATCG	GGGGGGTATCCTATAGTTATAGTAGATTA	A/G	**Pgdm1**
ndhK	**PgKm2**	GAGAGTCTTGACCAATTTGAAAGATCA	AGAGTCTTGACCAATTTGAAAGATCG	TCCCTTACTTGATCGAACAACCCAAAATT	A/G	**Pgdm2**
rpoC1	**PqKm3**	GATACACTTCTTGATAATGGAATCCGT	ATACACTTCTTGATAATGGAATCCGC	CTTTATTATGACCGTCCCTCATCGGTT	A/G	**Pqdm4**
ndhA	**PqKm4**	AAATAACTTAGTAGAATTGATGTAACTGCTAT	ATAACTTAGTAGAATTGATGTAACTGCTAC	AGATACTCGTTTATTCAGTATCGGACCAT	T/C	**Pqdm5**
rpoC2	**PjKm5**	CAGGAGAATTAGTAATGTGTCAGGAA	CAGGAGAATTAGTAATGTGTCAGGAG	GTGTATCCACGGCTTCTTGTACCAA	T/C	
ndhH	**PjKm6**	GGATTTGCAAATTGATAAAACCCGGC	AGGATTTGCAAATTGATAAAACCCGGT	GGGGTGTTTTTCCTTGGAGATGGAA	C/T	
ndhH	**PvKm7**	TACATAAGGTAAATACTGTATAATTGTTCGA	ACATAAGGTAAATACTGTATAATTGTTCGG	CATAGAGGGATGGAAAAAATTGCGGAAAA	A/G	**Pvdm13**
ndhD	**PvKm8**	TCCGCCCATTTCGTCGAGATAAAT	CCGCCCATTTCGTCGAGATAAAC	CAGGAACGAGTTATGATAGAATACGTCTT	T/C	
rpoC1	**PnKm9**	ATTTACACAAATACCCCGACGAC	CCTATTTACACAAATACCCCGACGAT	CCAACTCAAGATATGCTTATTGGACTCTA	C/T	**Pndm7**
rpoC2	**PnKm10**	CTGAGTCAGAATAAATATGTTTTCGG	CCTCTGAGTCAGAATAAATATGTTTTCGA	CGCGCAGGAACATCCACTTTGAATT	G/A	**Pndm8**
rpoC1	**PsKm11**	GGAGCTTATCGGCAGAAACGAATT	GAGCTTATCGGCAGAAACGAATC	CACCGAAGCCACAAAGGACTATCTA	A/G	**Psdm15**
rpoB	**PsKm12**	AAGCTTCCTTCCTATTAATCTGGAAATTT	GCTTCCTTCCTATTAATCTGGAAATTC	GGCTCTGGAACTGAATCATTTCTTTGTAT	T/C	**Psdm16**
ndhA	**PtKm13**	CATTATTTGTAACGGTTCTTTACTTGGGA	ATTTGTAACGGTTCTTTACTTGGGG	CGGATATGTACGGAATAGAAAGATTCCAA	T/C	**Ptdm17**
rpoC1	**PtKm14**	CGCTCTATTTAGCAATACGGGATGT	GCTCTATTTAGCAATACGGGATGC	CGATTGTATGGGAAATACTTCAGGAAGTT	T/C	**Ptdm18**

Abbreviations: Pg, *P. ginseng*; Pq, *P. quinquefolius*; Pn, *P. notoginseng*; Pj, *P. japonicus*; Pv, *P. vietnamensis*; Ps, *P. stipuleanatus*; Pt, *P. trifolius*. * dCAPS markers developed in the previous study (Nguyen et al., 2018).

Table 2-3 Primer sequences of 2 dCAPS markers developed to authenticate *P. vietnamensis* var. *fuscidiscus*.

Gene	dCAPS_ID	Primer sequence (5'-3')	SNP [a]	Recognition site	Restriction enzyme	PCR product size (bp)
psaA	**Pv.fdm19**	F: GCACCTAGGAAGAAAAGTCCAT**C**TGC R: GAAGAGGGGGAACGTGTCAA	G/A	CTGCAG	PstI	291
ndhD	**Pv.fdm20**	F: AACTCTCTCGGTCCAGAATCA R: AGTGGTTTTGTTGCAGAATTGAT**GA**T	A/G	GATC	MboI	211

Abbreviations: Pv.f, *P.vietnamensis* var. *fuscidiscus*. Underline bold letters are mismatched base. [a] targeted genotype SNP/ non_targeted genotype SNP

Table 2-4 Results of KASP marker genotyping as a matrix.

SNPs markers	KASP														dCAPS	
Clades	**PgKm1**	**PgKm2**	**PqKm3**	**PqKm4**	**PjKm5**	**PjKm6**	**PvKm7**	**PvKm8**	**PnKm9**	**PnKm10**	**PsKm11**	**PsKm12**	**PtKm13**	**PtKm14**	**Pv.fdm19**	**Pv.fdm20**
(1) Pg	**A**	**A**	B	B	B	B	B	B	B	B	B	B	B	B	B	B
(2) Pq	B	B	**A**	**A**	B	B	B	B	B	B	B	B	B	B	B	B
(3) Pj	B	B	B	B	**A**	**A**	B	B	B	B	B	B	B	B	B	B
(5) Pv	B	B	B	B	B	B	**A**	**A**	B	B	B	B	B	B	B	B
(7) Pn	B	B	B	B	B	B	B	B	**A**	**A**	B	B	B	B	B	B
(8) Ps	B	B	B	B	B	B	B	B	B	B	**A**	**A**	B	B	B	B
(9) Pt	B	B	B	B	B	B	B	B	B	B	B	B	**A**	**A**	B	B
(6) Pvf_Pz	B	B	B	B	B	B	B	B	B	B	B	B	B	B	**A**	**A**

Abbreviations: A, targeted genotype and B, non_targeted genotype. Pg, *P. ginseng*; Pq, *P. quinquefolius*; Pn, *P. notoginseng*; Pj, *P. japonicus*; Pv, *P. vietnamensis* var. *vietnamensis* and var. *langbianensis*; Pv.f, *P.vietnamensis* var. *fuscidiscus*; Pz, *P. zingiberensis*; Ps, *P. stipuleanatus*; Pt, *P. trifolius*.

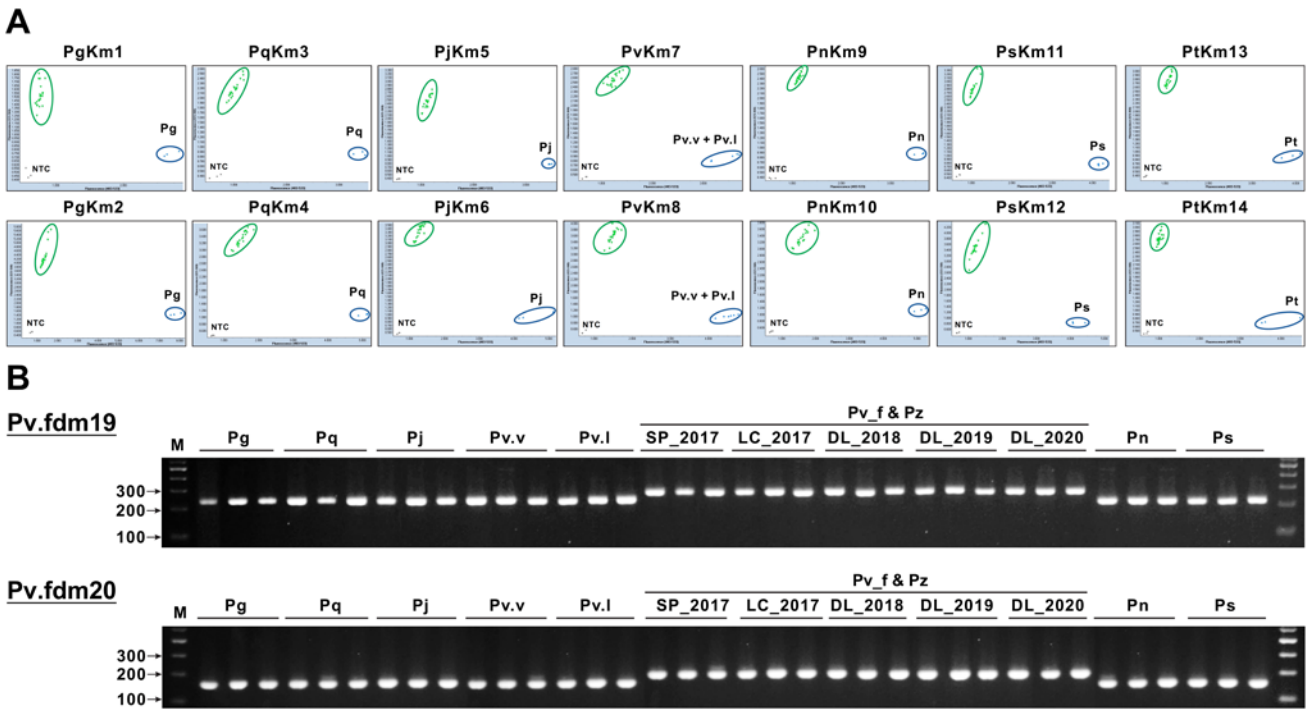

Figure 2-6. Results of KASP and dCAPS markers. (A) Screenshots of 14 KASP assay results. Marker names are shown at the top of screenshots. KASP assay results show differentiation between target *Panax* species (FAM-labeled, blue) with other *Panax* species (HEX-labeled, green). KASP assays examined in a set of *Panax* species: Pg, *P. ginseng*; Pq, *P. quinquefolius*; Pj, *P. japonicus*; Pv.v, *P. vietnamensis* var. *vietnamensis*; Pv.l, *P.vietnamensis* var. *langbianensis*; Pv.f, *P.vietnamensis* var. *fuscidiscus*; Pz, *P. zingiberensis*; Pn, *P. notoginseng*; Ps, *P.stipuleanatus*, and Pt, *P. trifolius*, respectively. NTC, no template control. (B) A proof of concept of the SNP-derived dCAPS markers to authenticate *P.vietnamensis* var. *fuscidiscus* collected in Sapa (SP), Lai Chau (LC), and Da Lat (DL) from 2017 to 2020. M, 100-bp DNA ladder.

Authentication of *Panax* collections in Vietnam and Southern China

The practical applicability of the established assays was evaluated by conducting 14 KASP and 2 dCAPS analyses on 114 *Panax* specimens collected from different locations in Vietnam. Overall, the assays analysis generated similar profiles for 114 *Panax* specimens, all of which were consistent with the initial identification of the specimens (Table S2-1). Four *Panax* clades were discovered in the collections from six different collecting locations (Fig. 2-7). 51 specimens were *P. vietnamensis* var. *vietnamensis* / var. *langbianensis*, clade (5); 41 specimens were *P. vietnamensis* var. *fuscidiscus* / *P. zingiberensis* clade (6); 2 specimens were *P. notoginseng* clade (7), and 20 specimens were *P. stipuleanatus* clade (8).

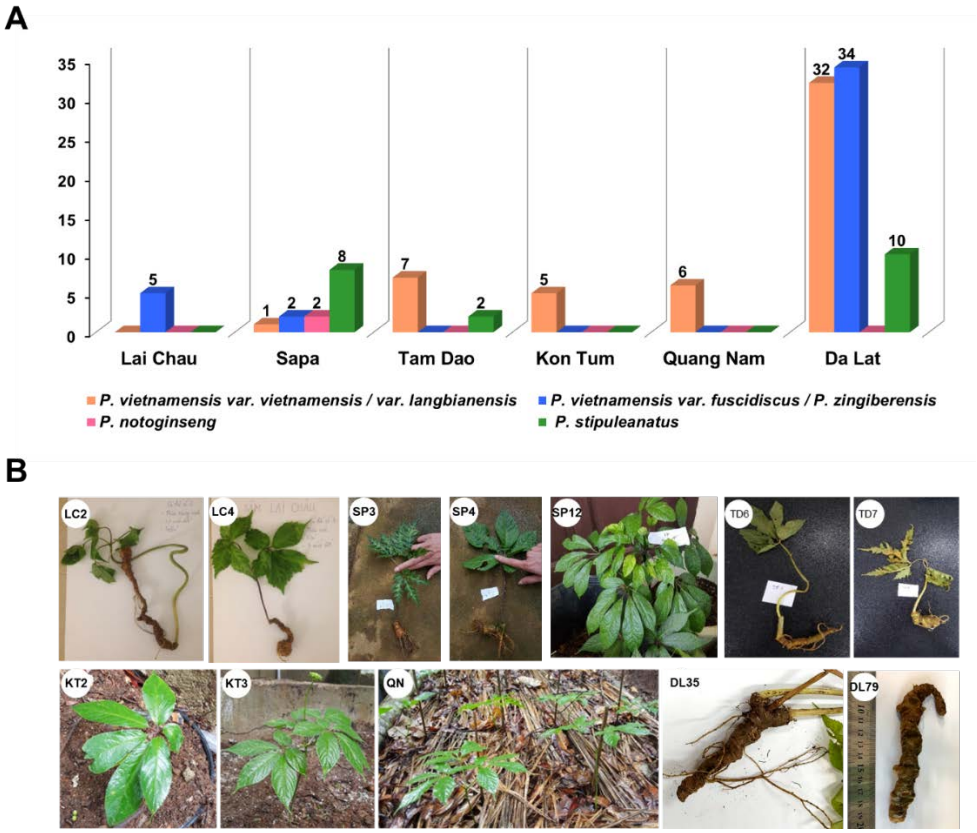

Figure 2-7. Authentication of 114 *Panax* collections using 14 KASP markers and 2 dCAPS markers. (A) The frequency distribution graph presents the number of *Panax* species successfully distinguished in 6 collection sites: Lai Chau province (LC), Sa Pa district (SP), Tam Dao district (TD), Kon Tum province (KT), Quang Nam province (QN) and cultivated in Da Lat city (DL). (B) Several pictures of *Panax* individuals were collected from six provinces in Vietnam.

DISCUSSION

Phylogenetic relationship in *Panax* species

The present work added NGS data for seven varieties of *P. vietnamensis*, thereby enriching the genomic reference databases for *Panax* species. Our phylogenetic analysis of the *Panax* genus utilizing 45S nrDNA and complete plastomes revealed nine well-supported clades (Fig. 2-2; Fig. 2-3), consistent with earlier phylogenetic results (Ji et al., 2019; Kim et al., 2015a; Kim et al., 2017; Kim et al., 2018; Liu et al., 2018; Nguyen et al., 2018; Zhou et al., 2018). According to Ji et al., 2019, the ancient hybridization and introgression (Wendel & Doyle, 1998) might be exhibited in the *Panax* population, and the interspecific introgression of plastome was unaccompanied by nuclear introgression. Consequently, the phylogenies generated by maternal inheritance (plastome) and biparental inheritance (45S nrDNA) of 9 *Panax* clades were incongruent. In addition, new species might be the products of interspecific hybridization (Comes & Kadereit, 1998). The bulk of hybrids probably accounted for the high biodiversity of the diploid *Panax* species in South-Eastern Asia, and these past introgression processes were also investigated in other plant species (Ito et al., 2013; Liu et al., 2020; Yi et al., 2015).

One noteworthy discovery was that Pv_f accessions were phylogenetically classified to the clade (6). For the first time, a strong link between *P. vietnamensis* var. *fuscidiscus* and *P. zingiberensis* was described. Within clades (6) and (3), the occurrence of interclade hybridizations was similar (Fig. 2-2). Ji et al., 2019 previously suggested that natural hybridization introgression occurs within the clade (3). Hence, it could infer that a complex process of lineage divergence was presented in these two clades evolution. The current study relied on DNA data and lacked association with phytochemical and morphological diagnostic characteristics; therefore,

additional research is necessary to ascertain the mechanisms underlying the complexity of these clades. However, when only clade names were evaluated rather than taxonomic positions, the phylogenetic incongruences did not affect the accuracy of *Panax* species identification.

Clade-specific SNPs in *Panax* plastomes

This research utilized the pairwise number of SNP differences between plastomes (Δ) as a monitoring tool to observe genetic variability across 84 *Panax* plastomes. Except for clade (9), at least five plastome accessions were obtained for each *Panax* clade that could capture Δ_{intra} and Δ_{inter}. Although Δ_{intra} were lower than Δ_{inter} within each clade (Fig. 2-3**)**, Δ_{intra} values were unstable among clades. When all plastomes were compared, the occurrence of high values of Δ_{intra} in two clades (3) and (6) results in the overlap between Δ_{intra} and Δ_{inter} values (Fig. S2-3). Therefore, the SNP difference-based plastomes were insufficient to define the upper limit for intraspecific divergences and the lower limit for interspecific divergences in *Panax* species. Estimation was challenging since numerous biological processes, including interspecific hybridization (Fehrer et al., 2007; Rieseberg & Willis, 2007; Soltis & Soltis, 2009), horizontal gene transfer (Bock, 2010; Gao et al., 2014), and incomplete lineage sorting (Charlesworth et al., 2005; Fehrer et al., 2007; Sousa & Hey, 2013). However, the data set used in this study did not allow for exploration of *Panax* species ambiguity, and additional research should be conducted to determine which processes lead to the abundant diversity of diploid *Panax* species in South-Eastern Asia.

Furthermore, the intra-clade SNPs remained a barrier to clade-level identification in *Panax* species. Maximum Δ_{intra} were greater than 0 in nearly all *Panax* clades. Except for clade-overlapping SNPs and intra-clade SNPs, only 59.5% of total SNPs were clade-specific SNPs suitable for uncovering

Panax clades' complexity. This research focused on the genetic basis of *Panax* plastome SNP diversity and successfully found 16 clade-specific SNPs that serve as effective ways for discriminating eight *Panax* clades (Fig. 2-6) (*P. wangianus* clade (4) was removed due to a lack of DNA samples).

Synonymous (Ks) and non-synonymous (Ka) substitution of plastid protein-coding genes (CDSs)

I performed the nucleotide substitutions analysis by employing the NGS data from 84 *Panax* plastomes. The results demonstrated that nucleotide substitutions were globally found in 74 of 78 CDSs while the CDS of 4 genes, *psbF*, *psbN*, *psbT*, and *rpl23*, were conserved across the tested plastomes. In addition, the maximum Ka value lower than that of Ks (Table S2-3) was observed in 65 CDSs. Therefore, non-synonymous substitution was less frequent in mutation of plastid genes during the evolution of *Panax* genus. It provided strong evidence of broad purifying selection throughout the plastid genes of *Panax* species due to the conservation of functional coding genes (Hurst, 2002). Although Ka values were lower than Ks values in most genes, six genes (*atpF*, *accD*, *rpoA*, *rpl2*, *ycf2*, and *rps15*) had median Ka values greater than Ks values, almost certainly due to positive Darwinian selection.

Authentication of *Panax* collections in Vietnam and Southern China

Fourteen KASP and two dCAPS markers (Table 2-4, Fig. 2-6) were effectively and rapidly classified 114 Panax specimens collected from six locations around Vietnam (Fig. 2-7). The established markers correctly categorized 41 specimens of Pv_f into the clade (6). To summarize, the analysis of ML phylogenies of the completed plastome and 45S nrDNA sequences in conjunction with the profile of clade-specific SNP markers demonstrated a clear separation between *P. vietnamensis* var. *fuscidiscus* and the other two *P. vietnamensis* varieties (var. *vietnamensis* and var.

langbianensis). In Vietnam, *P. vietnamensis* varieties are officially classified as endangered in the Vietnam Red Data Book 2007 (Nong et al., 2016b; Vu et al., 2020) and must be conserved equally (Decree 06/2019/ND-CP). Before appropriate reclassification for *P. vietnamensis* varieties can be achieved, the clade-specific SNP markers generated in this study provide prospective platforms for preventing overexploitation of *P. vietnamensis* var. *fuscidiscus* in commercial fraud of *P. vietnamensis* var. *vietnamensis*.

The utility and complementarity of the KASP assay

Previously, dCAPS markers were used to examine the genetic diversity of *Panax* species (Nguyen et al., 2018). However, this approach was tedious and time-consuming, as it required extra restriction endonuclease digestion and gel electrophoresis for scoring after amplification processes. In addition, this approach was dependent on site-specific restriction enzymes, which were not applicable to all *Panax* clade-specific SNPs. dCAPS was classified as a second-generation marker assumed unsuitable for automated or high-throughput analysis (Georgiev & Pavlov, 2017). Currently, KASP genotyping assays, also referred to as third-generation molecular markers, provide a rapid, highly accurate, cost-effective, and very flexible approach for SNP genotyping studies in many crops (Allen et al., 2013; Cheon et al., 2018; Ertiro et al., 2015; Moon et al., 2019; Semagn et al., 2014; Tan et al., 2017; Yang et al., 2019; Yu et al., 2017; Zhao et al., 2017). Thus, I converted *Panax* clade-specific SNPs to KASP markers (Table 2-2). KASP assays are gel-free, high-throughput singleplex SNP-genotyping platforms based on dual FRET (Fluorescent Resonance Energy Transfer) (Neelam et al., 2013). PCR products were easily detected by reading fluorescent signals after being amplified using a thermal cycler (Fig. 2-6A). The present study demonstrated that this single tube SNP-genotyping assay overcame limitations of gel-based genotyping methods and provided up-to-date status on the molecular tools for rapid and accurate identification of *Panax* species.

CONCLUSION

First-time reports of complete plastome and 45S nrDNA sequences of *P. vietnamensis* var. *fuscidiscus* and var. *langbianensis* were presented in this study. The phylogenetic analysis of 84 plastome sequences and 76 45S nrDNA revealed 9 *Panax* clades. Based on the survey of plastome SNPs, 16 clade-specific SNPs were converted to KASP and dCAPS assays that were sufficient to identify 8 of 9 Panax clades accurately. KASP assay is the high-throughput and high-resolution platform for rapid and accurate identification of *Panax* species in extensive collections. Transferring 14 KASP and two dCAPS markers to other laboratories worldwide would be highly advantageous to ongoing efforts to establish a high-throughput, sensitive, and reliable approach for tracking illegal ginseng adulterations. These assays also provide a framework for managing *Panax* species conservation and preventing them from overharvesting or habitat loss, which is particularly important in developing countries with high levels of biodiversity but limited resources to conduct DNA barcoding research, such as Vietnam. On the other hand, the instability of the intra SNP difference-based plastome pattern shows that additional research is necessary to fully comprehend the vast diversity of diploid *Panax* species in Southeast Asia.

REFERENCES

Allen, A. M., Barker, G. L., Wilkinson, P., Burridge, A., Winfield, M., Coghill, J., Uauy, C., Griffiths, S., Jack, P., & Berry, S. (2013). Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.). *Plant Biotechnology Journal*, *11*(3), 279-295.

Bai, L., Gao, J., Wei, F., Zhao, J., Wang, D., & Wei, J. (2018). Therapeutic potential of ginsenosides as an adjuvant treatment for diabetes. *Frontiers in Pharmacology*, *9*, 423.

Birky, C. W. (1995). Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proceedings of the National Academy of Sciences*, *92*(25), 11331-11338.

Bock, R. (2010). The give-and-take of DNA: horizontal gene transfer in plants. *Trends in plant science*, *15*(1), 11-22.

Bon, T. N., Tuyen, P. Q., Son, H. T., Anh, N. T. H., Tan, B. T., Son, N. T., Hung, N. Q., Van Anh, N. T., & Van Do, T. (2019). *Panax* sp. in Tuyen Quang, North Vietnam–A Potential Plant for Poverty Reduction. *Asian Journal of Research in Botany*, 1-10.

Brookes, A. J. (1999). The essence of SNPs. *Gene*, *234*(2), 177-186.

Charif, D., & Lobry, J. R. (2007). SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In *Structural approaches to sequence evolution* (pp. 207-232). Springer.

Charlesworth, B., Bartolomé, C., & NoëL, V. (2005). The detection of shared and ancestral polymorphisms. *Genetics Research*, *86*(2), 149-157.

Cheon, K.-S., Baek, J., Cho, Y.-i., Jeong, Y.-M., Lee, Y.-Y., Oh, J., Won, Y. J., Kang, D.-Y., Oh, H., & Kim, S. L. (2018). Single Nucleotide

Polymorphism (SNP) Discovery and Kompetitive Allele-Specific PCR (KASP) Marker Development with Korean Japonica Rice Varieties. *Plant Breeding and Biotechnology*, *6*(4), 391-403.

Coates, D. J., Byrne, M., & Moritz, C. (2018). Genetic diversity and conservation units: dealing with the species-population continuum in the age of genomics. *Frontiers in Ecology and Evolution*, *6*, 165.

Comes, H. P., & Kadereit, J. W. (1998). The effect of Quaternary climatic changes on plant distribution and evolution. *Trends in plant science*, *3*(11), 432-438.

Duc, N. M., Nham, N. T., Kasai, R., Ito, A., Yamasaki, K., & Tanaka, O. (1993). Saponins from Vietnamese ginseng, *Panax vietnamensis* Ha et Grushv. collected in central Vietnam. I. *Chemical and pharmaceutical bulletin*, *41*(11), 2010-2014.

Dung, H. T., & Grushvitski, I. (1985). A new species of the genus *Panax* (Araliaceae) from Vietnam. *Botanicheskii zhurnal*.

Ertiro, B. T., Ogugo, V., Worku, M., Das, B., Olsen, M., Labuschagne, M., & Semagn, K. (2015). Comparison of Kompetitive Allele Specific PCR (KASP) and genotyping by sequencing (GBS) for quality control analysis in maize. *BMC genomics*, *16*(1), 908.

Fehrer, J., Gemeinholzer, B., Chrtek Jr, J., & Bräutigam, S. (2007). Incongruent plastid and nuclear DNA phylogenies reveal ancient intergeneric hybridization in *Pilosella hawkweeds* (Hieracium, Cichorieae, Asteraceae). *Molecular phylogenetics and evolution*, *42*(2), 347-361.

Gao, C., Ren, X., Mason, A. S., Liu, H., Xiao, M., Li, J., & Fu, D. (2014). Horizontal gene transfer in plants. *Functional & integrative genomics*, *14*(1), 23-29.

Georgiev, V., & Pavlov, A. (2017). *Salvia Biotechnology*. Springer.

Goodwin, P., & Proctor, E. (2019). Molecular techniques to assess genetic variation within and between *Panax ginseng* and *Panax quinquefolius*. *Fitoterapia*, 104343.

Gouy, M., Guindon, S., & Gascuel, O. (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution*, *27*(2), 221-224.

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology*, *52*(5), 696-704.

Hall, T. (1999). A user-friendly biological sequence alignment editor and analysis program for Windows TM. *Bioedit Version*, *7*.

Ho, V. H., & Pham, Q. T. (2020). Development of Ngoc Linh ginseng in Nam Tra My district. E3S Web of Conferences,

Hou, M., Wang, R., Zhao, S., & Wang, Z. (2021). Ginsenosides in *Panax* genus and their biosynthesis. *Acta Pharmaceutica Sinica B*.

Hurst, L. D. (2002). The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends in genetics: TIG*, *18*(9), 486-486.

Ito, Y., Ohi-Toma, T., Murata, J., & Tanaka, N. (2013). Comprehensive phylogenetic analyses of the *Ruppia maritima* complex focusing on taxa from the Mediterranean. *Journal of plant research*, *126*(6), 753-762.

Jang, W., Lee, H. O., Kim, J.-U., Lee, J.-W., Hong, C.-E., Bang, K.-H., Chung, J.-W., & Jo, I.-H. (2020). Complete Mitochondrial Genome and a Set of 10 Novel Kompetitive Allele-Specific PCR Markers in Ginseng (*Panax ginseng* CA Mey.). *Agronomy*, *10*(12), 1868.

Ji, Y., Liu, C., Yang, Z., Yang, L., He, Z., Wang, H., Yang, J., & Yi, T. (2019). Testing and using complete plastomes and ribosomal DNA sequences

as the next generation DNA barcodes in *Panax* (Araliaceae). *Molecular ecology resources*, *19*(5), 1333-1345.

Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in bioinformatics*, *20*(4), 1160-1166.

Kim, K., Lee, S.-C., Lee, J., Lee, H. O., Joh, H. J., Kim, N.-H., Park, H.-S., & Yang, T.-J. (2015a). Comprehensive survey of genetic diversity in chloroplast genomes and 45S nrDNAs within *Panax ginseng* species. *PloS one*, *10*(6), e0117159. https://doi.org/https://doi.org/10.1371/journal.pone.0117159

Kim, K., Lee, S.-C., Lee, J., Yu, Y., Yang, K., Choi, B.-S., Koh, H.-J., Waminal, N. E., Choi, H.-I., & Kim, N.-H. (2015b). Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of *Oryza* AA genome species. *Scientific Reports*, *5*, 15655.

Kim, K., Nguyen, V. B., Dong, J., Wang, Y., Park, J. Y., Lee, S.-C., & Yang, T.-J. (2017). Evolution of the Araliaceae family inferred from complete chloroplast genomes and 45S nrDNAs of 10 *Panax*-related species. *Scientific Reports*, *7*(1), 4917. https://doi.org/10.1038/s41598-017-05218-y

Kim, N.-H., Jayakodi, M., Lee, S.-C., Choi, B.-S., Jang, W., Lee, J., Kim, H. H., Waminal, N. E., Lakshmanan, M., van Nguyen, B., Lee, Y. S., Park, H.-S., Koo, H. J., Park, J. Y., Perumal, S., Joh, H. J., Lee, H., Kim, J., Kim, I. S., . . . Yang, T.-J. (2018). Genome and evolution of the shade-requiring medicinal herb *Panax ginseng*. *Plant Biotechnology Journal*, *16*(11), 1904-1917. https://doi.org/10.1111/pbi.12926

Komatsu, K., Zhu, S., Fushimi, H., Qui, T. K., Cai, S., & Kadota, S. (2001). Phylogenetic analysis based on 18S rRNA gene and matK gene

sequences of *Panax vietnamensis* and five related species. *Planta medica*, *67*(05), 461-465.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome research*, *19*(9), 1639-1645.

Le, Q.-U., Lay, H.-L., Wu, M.-C., Nguyen, T. H.-H., & Nguyen, D.-L. (2018). Phytoconstituents and biological activities of *Panax vietnamensis* (Vietnamese Ginseng): A precious ginseng and call for further research-A systematic review. *Natural Product Communications*, *13*(10), 1934578X1801301036.

Liu, B.-B., Campbell, C. S., Hong, D.-Y., & Wen, J. (2020). Phylogenetic relationships and chloroplast capture in the Amelanchier-Malacomeles-Peraphyllum clade (Maleae, Rosaceae): Evidence from chloroplast genome and nuclear ribosomal DNA data using genome skimming. *Molecular phylogenetics and evolution*, *147*, 106784.

Liu, C., Yang, Z., Yang, L., Yang, J., & Ji, Y. (2018). The complete plastome of *Panax stipuleanatus:* Comparative and phylogenetic analyses of the genus *Panax* (Araliaceae). *Plant Diversity*, *40*(6), 265-276. https://doi.org/10.1016/j.pld.2018.11.001

Long, P. K., Van The, P., Thanh, T. T. V., Hien, D. P., Loc, P. K., Tam, N. M., & Duy, V. D. (2020). Morphological and molecular data of *Panax* population found in Phu Xai Lai Leng mountainous range of Nghe an province. *Research Journal of Biotechnology Vol*, *15*, 8.

Moon, J.-H., Son, D., Lee, J.-W., & Yoo, S.-C. (2019). Development of Kompetitive Allele Specific PCR Markers for Submergence Tolerant Gene Sub1 in Rice. *Plant Breeding and Biotechnology*, *7*(1), 62-66.

Neelam, K., Brown-Guedira, G., & Huang, L. (2013). Development and validation of a breeder-friendly KASPar marker for wheat leaf rust resistance locus Lr21. *Molecular breeding*, *31*(1), 233-237.

Ngoc, P. T., Huyen, P. T., Nga, N. Q., Van Truong, P., Khoi, N. M., & Long, D. D. (2020). A Molecular Phylogeny of *Panax* L. Genus (Araliaceae) Based on ITS-rDNA and matK Support for Identification of Panax Species in Vietnam. *VNU Journal of Science: Medical and Pharmaceutical Sciences*, *36*(2).

Nguyen, V. B., Giang, V. N. L., Waminal, N. E., Park, H.-S., Kim, N.-H., Jang, W., Lee, J., & Yang, T.-J. (2018). Comprehensive comparative analysis of chloroplast genomes from seven *Panax* species and development of an authentication system based on species-unique single nucleotide polymorphism markers. *Journal of Ginseng Research*. https://doi.org/https://doi.org/10.1016/j.jgr.2018.06.003

Nong, V. D., Le, N. T., Nguyen, D. C., & Tran, V. T. (2016a). A new variety of *Panax* (Araliaceae) from Lam Vien Plateau, Vietnam and its molecular evidence. *Phytotaxa*, *277*(1), 47-58.

Nong, V. D., Vu, T. C., & Tran, V. T. (2016b). Genetic diversity of *Panax vietnamensis* var. *fuscidiscus* K. Komatsu, S. Zhu & SQ cai population in western north of vietnam detected by inter simple sequence repeat markers. *Vietnam Journal of Biotechnology*, *14*(4), 619-627.

Park, H.-S., Jayakodi, M., Lee, S. H., Jeon, J.-H., Lee, H.-O., Park, J. Y., Moon, B. C., Kim, C.-K., Wing, R. A., & Newmaster, S. G. (2020). Mitochondrial plastid DNA can cause DNA barcoding paradox in plants. *Scientific Reports*, *10*(1), 1-12.

Phan, K., Le, T., Phan, K., Vo, D., & Phan, V. (2013). Lai Chau ginseng *Panax vietnamensis* var. *fuscidiscus* K. Komatsu, S. Zhu & SQ Cai I. morphology, ecology, distribution and conservation status. Proceeding

of the 2nd VAST-KAST Workshop on Biodiversity and Bio-active Compounds,

Qu, X.-J., Moore, M. J., Li, D.-Z., & Yi, T.-S. (2019). PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods*, *15*(1), 50.

Rieseberg, L. H., & Willis, J. H. (2007). Plant speciation. *Science*, *317*(5840), 910-914.

Secretariat, G. (2019). *Panax* L. in GBIF Secretariat. GBIF backbone taxonomy. *Checklist Dataset https://doi.org/10.15468/39omei accessed via GBIF.org on 2019-10-04*.

Semagn, K., Babu, R., Hearne, S., & Olsen, M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Molecular breeding*, *33*(1), 1-14.

Soltis, P. S., & Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annual review of plant biology*, *60*, 561-588.

Sousa, V., & Hey, J. (2013). Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics*, *14*(6), 404-414.

Tan, C.-T., Yu, H., Yang, Y., Xu, X., Chen, M., Rudd, J. C., Xue, Q., Ibrahim, A. M., Garza, L., & Wang, S. (2017). Development and validation of KASP markers for the greenbug resistance gene Gb7 and the Hessian fly resistance gene H32 in wheat. *Theoretical and Applied Genetics*, *130*(9), 1867-1884.

Vu, D. D., Shah, S. N. M., Pham, M. P., Nguyen, M. T., & Nguyen, T. P. T. (2020). De novo assembly and transcriptome characterization of an endemic species of Vietnam, *Panax vietnamensis* Ha et Grushv.,

including the development of EST-SSR markers for population genetics.

Wendel, J. F., & Doyle, J. J. (1998). Phylogenetic incongruence: window into genome history and molecular evolution. In *Molecular systematics of plants II* (pp. 265-296). Springer.

Yamasaki, K. (2000). Bioactive saponins in Vietnamese ginseng, *Panax vietnamensis*. *Pharmaceutical biology*, *38*(sup1), 16-24.

Yang, G., Chen, S., Chen, L., Sun, K., Huang, C., Zhou, D., Huang, Y., Wang, J., Liu, Y., & Wang, H. (2019). Development of a core SNP arrays based on the KASP method for molecular breeding of rice. *Rice*, *12*(1), 21.

Yi, T.-S., Jin, G.-H., & Wen, J. (2015). Chloroplast capture and intra-and inter-continental biogeographic diversification in the Asian–New World disjunct plant genus *Osmorhiza* (Apiaceae). *Molecular phylogenetics and evolution*, *85*, 10-21.

Yu, L.-X., Chao, S., Singh, R. P., & Sorrells, M. E. (2017). Identification and validation of single nucleotide polymorphic markers linked to Ug99 stem rust resistance in spring wheat. *PloS one*, *12*(2), e0171963.

Zhang, G.-H., Ma, C.-H., Zhang, J.-J., Chen, J.-W., Tang, Q.-Y., He, M.-H., Xu, X.-Z., Jiang, N.-H., & Yang, S.-C. (2015). Transcriptome analysis of *Panax vietnamensis* var. *fuscidicus* discovers putative ocotillol-type ginsenosides biosynthesis genes and genetic markers. *BMC genomics*, *16*(1), 1-20.

Zhao, S., Li, A., Li, C., Xia, H., Zhao, C., Zhang, Y., Hou, L., & Wang, X. (2017). Development and application of KASP marker for high throughput detection of AhFAD2 mutation in peanut. *Electronic Journal of Biotechnology*, *25*, 9-12.

Zhengyi, W. (1983). On the significance of Pacific intercontinental discontinuity. *Annals of the Missouri Botanical Garden*, *70*(4), 577-590.

Zhou, M., Gong, X., & Pan, Y. (2018). *Panax species* identification with the assistance of DNA data. *Genetic Resources and Crop Evolution*, *65*(7), 1839-1856.

Zuo, Y.-J., Wen, J., & Zhou, S.-L. (2017). Intercontinental and intracontinental biogeography of the eastern Asian–eastern North American disjunct *Panax* (the ginseng genus, Araliaceae), emphasizing its diversification processes in eastern Asia. *Molecular phylogenetics and evolution*, *117*, 60-74.

APPENDIX

Table S1-1. List of plastome sequences of Araliaceae

No.		GeneBank NGS_ID	Species	References
1	*	KM088019.1	*Panax ginseng* cultivar Chunpoong	Kim et al. 2015
2	*	KM067394.1	*Panax ginseng* landrace Hwangsook	Kim et al. 2015
3	*	KM067386.1	*Panax ginseng* cultivar Cheongsun	Kim et al. 2015
4	*	KM067388.1	*Panax ginseng* cultivar Gumpoong	Kim et al. 2015
5	*	KM067387.1	*Panax ginseng* cultivar Gopoong	Kim et al. 2015
6	*	KM067389.1	*Panax ginseng* cultivar Jakyung	Kim et al. 2015
7	*	KM067391.1	*Panax ginseng* cultivar Sunpoong	Kim et al. 2015
8	*	KM067392.1	*Panax ginseng* cultivar Sunun	Kim et al. 2015
9	*	KM088020.1	*Panax ginseng* cultivar Yunpoong	Kim et al. 2015
10	*	KM067390.1	*Panax ginseng* cultivar Sunone	Kim et al. 2015
11	*	KM067393.1	*Panax ginseng* cultivar Sunhyang	Kim et al. 2015
12		KF431956.1	*Panax ginseng*	Zhao et al. 2015
13		KC686332.1	*Panax ginseng* isolate ermaya	Zhao et al. 2015
14		KC686333.1	*Panax ginseng* isolate gaolishen	Zhao et al. 2015
15		KC686331.1	*Panax ginseng* isolate damaya	Zhao et al. 2015
16		MH049735.2	*Panax ginseng*	Wang et al. 2018
17		MK408938.1	*Panax ginseng* isolate JYH2016473	Ji et al. 2019
18		AY582139.1	*Panax ginseng*	Kim and Lee 2004
19	*	KM088018.1	*Panax quinquefolius*	Kim et al. 2015
20		KT028714.1	*Panax quinquefolius*	Unpublished
21		MK408923.1	*Panax quinquefolius* isolate JYH2016493	Ji et al. 2019
22		MK408953.1	*Panax quinquefolius* isolate JYH2016494	Ji et al. 2019
23	*	KP036469.1	*Panax japonicus*	Kim et al. 2015
24		MK308678.1	*Panax japonicus*	Ji et al. 2019
25		MK408957.1	*Panax japonicus* var. *bipinnatifidus* isolate JYH2016547	Ji et al. 2019
26		MK408918.1	*Panax japonicus* var. *bipinnatifidus* isolate JYH2016548	Ji et al. 2019
27		MK408959.1	*Panax japonicus* var. *bipinnatifidus* isolate JYH2016486	Ji et al. 2019
28		MK408944.1	*Panax japonicus* var. *bipinnatifidus* isolate JYH2016488	Ji et al. 2019
29		MK408926.1	*Panax japonicus* var. *bipinnatifidus* isolate JYH2016481	Ji et al. 2019
30		MK408962.1	*Panax japonicus* var. *bipinnatifidus* isolate JYH2016483	Ji et al. 2019
31		MK408940.1	*Panax japonicus* var. *bipinnatifidus* isolate JYH2016485	Ji et al. 2019
32		MK408939.1	*Panax* sp. 'sinensis' isolate JYH2016476	Ji et al. 2019
33		MK408967.1	*Panax* sp. 'sinensis' isolate JYH2016474	Ji et al. 2019
34		MK408956.1	*Panax* sp. 'sinensis' isolate JYH2016475	Ji et al. 2019
35		MK408924.1	*Panax* sp. 'sinensis' isolate JYH2016612	Ji et al. 2019
36		MK408932.1	*Panax* sp. 'sinensis' isolate JYH2016608	Ji et al. 2019
37		MK408958.1	*Panax* sp. 'sinensis' isolate JYH2016610	Ji et al. 2019
38		MK408919.1	*Panax* sp. 'sinensis' isolate JYH2016611	Ji et al. 2019

39		MK408961.1	*Panax* sp. 'sinensis' isolate JYH2016609	Ji et al. 2019
40		MN496312.1	*Panax major*	Ji et al. 2019
41		MK408941.1	*Panax wangianus* isolate JYH2016541	Ji et al. 2019
42		MK408943.1	*Panax wangianus* isolate JYH2016543	Ji et al. 2019
43		MK408963.1	*Panax wangianus* isolate JYH2016551	Ji et al. 2019
44		MK408934.1	*Panax wangianus* isolate JYH2016544	Ji et al. 2019
45		MK408921.1	*Panax wangianus* isolate JYH2016545	Ji et al. 2019
46		MK408951.1	*Panax wangianus* isolate JYH2016552	Ji et al. 2019
47		MK408935.1	*Panax wangianus* isolate JYH2016550	Ji et al. 2019
48		MK408964.1	*Panax wangianus* isolate JYH2016540	Ji et al. 2019
49		MK408930.1	*Panax wangianus* isolate JYH2016539	Ji et al. 2019
50	*	KP036471.1	*Panax vietnamensis* isolate 38	Kim et al. 2015
51	*	KU059178.1	*Panax vietnamensis*	Unpublished
52	*	KP036470.1	*Panax vietnamensis* isolate 37	Unpublished
53		MF377623.1	*Panax vietnamensis*	Manzanilla et al. 2018
54		MF377621.1	*Panax* sp. VM2017 isolate 38	Manzanilla et al. 2018
55	*	MT798584	*Panax vietnamensis* var. *langbianensis*	Unpublished
56	*	MT798583	*Panax vietnamensis* var. *langbianensis*	Unpublished
57	*	MT798585	*Panax vietnamensis* var. *fuscidiscus*	Unpublished
58	*	MT798586	*Panax vietnamensis* var. *fuscidiscus*	Unpublished
59	*	MT798587	*Panax vietnamensis* var. *fuscidiscus*	Unpublished
60		MK408966.1	*Panax zingiberensis* isolate JYH2016439	Ji et al. 2019
61		MK408950.1	*Panax zingiberensis* isolate JYH2016471	Ji et al. 2019
62		MK408947.1	*Panax zingiberensis* isolate JYH2016440	Ji et al. 2019
63		MK408929.1	*Panax zingiberensis* isolate JYH2016441	Ji et al. 2019
64		MK408922.1	*Panax zingiberensis* isolate JYH2016472	Ji et al. 2019
65		MK408949.1	*Panax zingiberensis* isolate JYH2016467	Ji et al. 2019
66		MK408933.1	*Panax zingiberensis* isolate JYH2016465	Ji et al. 2019
67		MK408942.1	*Panax zingiberensis* isolate JYH2016436	Ji et al. 2019
68		MK408952.1	*Panax zingiberensis* isolate JYH2016438	Ji et al. 2019
69		MK408969.1	*Panax zingiberensis* isolate JYH2016463	Ji et al. 2019
70		MK408960.1	*Panax zingiberensis* isolate JYH2016468	Ji et al. 2019
71		MK408968.1	*Panax zingiberensis* isolate JYH2016464	Ji et al. 2019
72		KP036468.1	*Panax notoginseng*	Kim et al. 2015
73		MK408927.1	*Panax notoginseng* isolate LCK1	Ji et al. 2019
74		MK408931.1	*Panax notoginseng* isolate LCK7	Ji et al. 2019
75		MK408946.1	*Panax notoginseng* isolate LCK4	Ji et al. 2019
76		MK408954.1	*Panax notoginseng* isolate LCK10	Ji et al. 2019
77		MK408955.1	*Panax notoginseng* isolate LCK9	Ji et al. 2019
78		MK408925.1	*Panax notoginseng* isolate LCK5	Ji et al. 2019

79		MK408928.1	*Panax notoginseng* isolate LCK8	Ji et al. 2019
80		MK408937.1	*Panax notoginseng* isolate LCK3	Ji et al. 2019
81		MK408945.1	*Panax notoginseng* isolate LCK2	Ji et al. 2019
82		KR021381.1	*Panax notoginseng*	Unpublished
83		KJ566590.1	*Panax notoginseng*	Dong et al. 2014
84		KT001509.1	*Panax notoginseng*	Unpublished
85	*	KX247147.1	*Panax stipuleanatus*	Kim et al. 2018
86	*	KX247146.1	*Panax japonicus* var. *bipinnatifidus*	Unpublished
87		MF377620.1	*Panax japonicus* var. *bipinnatifidus*	Manzanilla et al. 2018
88		MF377622.1	*Panax stipuleanatus*	Manzanilla et al. 2018
89		KY379906.1	*Panax stipuleanatus*	Liu et al. 2018
90		MK408936.1	*Panax stipuleanatus* isolate JYH2016435	Ji et al. 2019
91		MK408965.1	*Panax stipuleanatus* isolate JYH2016437	Ji et al. 2019
92		MK408920.1	*Panax stipuleanatus* isolate JYH2016466	Ji et al. 2019
93		MF100782.1	*Panax trifolius*	Kim et al. 2018
94	*	KT153023.1	*Aralia elata*	Kim et al. 2017
95	*	**IM180813_55**	*Aralia elata*	Unpublished
96		KC456163.1	*Aralia undulata*	Li et al. 2013
97	*	**IM180813_32**	*Aralia cordata*	Unpublished
98		MH778959.1	*Aralia cordata*	Kim and Kim 2019
99		MG914654.1	*Aralia continentalis*	Unpublished
100		MK809524.1	*Aralia atropurpurea* isolate JL729ok	Liu and Wen 2019
101		MK778455.1	*Aralia atropurpurea*	Liu and Wen 2019
102		MK943809.1	*Sciadodendron excelsum* voucher Wen 6779	Valcárcel and Wen 2019
103		MK943807.1	*Osmoxylon novoguineense* voucher Wen 10706	Valcárcel and Wen 2019
104	*	KR136270.1	*Dendropanax morbiferus*	Kim et al. 2016b
105	*	**IM180813_56**	*Dendropanax morbiferus*	Unpublished
106		KP271241.1	*Dendropanax dentiger* voucher DDENT20141207	Wang et al. 2016
107		MT909827.1	*Dendropanax oligodontus*	Unpublished
108		MK943797.1	*Dendropanax nutans* voucher Wen 118783	Valcárcel and Wen 2019
109		MK930365.1	*Chengiopanax sciadophylloides*	Valcárcel and Wen 2019
110		MK943799.1	*Gamblea ciliata* var. *evodiifolia* voucher Wen 11199	Valcárcel and Wen 2019
111		MG397138.1	*Eleutherococcus brachypus*	Zhang et al. 2019
112		MN527993.1	*Eleutherococcus brachypus*	Unpublished
113		MN727298.1	*Eleutherococcus trifoliatus* voucher 2016010149	Chen et al. 2020
114		MT754220.1	*Eleutherococcus trifoliatus*	Li et al. 2020
115	*	KT153020.1	*Eleutherococcus gracilistylus*	Kim et al. 2016
116		KY085901.1	*Eleutherococcus senticosus*	Unpublished
117		JN637765.1	*Eleutherococcus senticosus*	Yi et al. 2012
118	*	KT153019.1	*Eleutherococcus sessiliflorus*	Kim et al. 2016

119		KC456164.1	*Brassaiopsis hainla*	Li et al. 2013
120		MK943811.1	*Trevesia sundaica* voucher Wen 10669	Valcárcel and Wen 2019
121		KC456165.1	*Metapanax delavayi*	Li et al. 2013
122		MK943802.1	*Macropanax dispermus* voucher Wen 10862	Valcárcel and Wen 2019
123		KC456167.1	*Kalopanax septemlobus*	Li et al. 2013
124		MK130890.1	*Hedera nepalensis* var. *sinensis*	Wu et al. 2019
125	*	**IM180813_57**	*Hedera rhombea*	Unpublished
126		MK943800.1	*Hedera helix* voucher Wen 12871	Valcárcel and Wen 2019
127	*	**IM180813_34**	*Hedera helix*	Unpublished
128		MK943803.1	*Merrilliopanax listeri* voucher Wen 5065	Valcárcel and Wen 2019
129		MK943804.1	*Merrilliopanax listeri* voucher Wen 5038	Valcárcel and Wen 2019
130		MK943798.1	*Fatsia polycarpa* voucher Wen 9426	Valcárcel and Wen 2019
131		KR021045.1	*Fatsia japonica*	Chen et al. 2016
132	*	**IM180813_33**	*Fatsia japonica*	Unpublished
133		MK943806.1	*Oreopanax obtusifolius* voucher Wen 8736	Valcárcel and Wen 2019
134		MK943808.1	*Schefflera morototoni* voucher Nee and Wen 53964	Valcárcel and Wen 2019
135		KT748629.1	*Schefflera heptaphylla*	Zong et al. 2016
136	*	**IM180813_58**	*Schefflera arboricola*	Unpublished
137		MT385083.1	*Schefflera actinophylla* voucher CANB874342	Maurin 2020
138		KC456166.1	*Schefflera delavayi*	Li et al. 2013
139		MK943801.1	*Heteropanax fragrans* voucher Li Rong1318	Valcárcel and Wen 2019
140		MK943810.1	*Tetrapanax papyrifer* voucher Wen 9431	Valcárcel and Wen 2019
141	*	**IM180813_35**	*Tetrapanax papyriferus*	Unpublished
142		MK943805.1	*Oplopanax horridus* voucher Taylor s.n.	Valcárcel and Wen 2019
143	*	**IM180813_36**	*Oplopanax elatus*	Unpublished
144	*	**IM1907233**	*Polyscias fruticosa*	Unpublished
145		MT385080.1	*Raukaua anomalus* voucher CHR649673	Maurin 2020
146		MT385082.1	*Raukaua simplex* voucher CHR437312	Maurin 2020
147		MT385081.1	*Raukaua edgerleyi* voucher CHR655508	Maurin 2020
148		MT385071.1	*Cheirodendron bastardianum* voucher P02800554	Maurin 2020
149		MK922468.1	*Harmsiopanax ingens*	Valcárcel and Wen 2019
150		MT561038.1	*Hydrocotyle nepalensis*	Unpublished
151		KT589392.1	*Hydrocotyle sibthorpioides*	Ge et al. 2017
152		HM596070.1	*Hydrocotyle verticillata*	Ge et al. 2017
153	*	**IM180813_54**	*Centella asiatica*	Unpublished

* Works of Lab. of Functional Plants: 38 plastomes; 26 Published

Table S1-2. List of genes in the Araliaceae plastomes.

Category for genes	Group of gene	Name of gene
Photosynthesis related genes (47 genes)	Rubisco	*rbcL*
	Photosystem I	*psaA, psaB, psaC, psaI, psaJ*
	Assembly/stability of photosystem I	*ycf3**, ycf4*
	Photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ*
	ATP synthesis	*atpA, atpB, atpE, atpF*, atpH, atpI*
	Cytochrome b/f compelx	*petA, petB*, petD*, petG, petL, petN*
	Cytochrome c synthesis	*ccsA*
	NADPH dehydrogenase	*ndhA*, ndhB*(x2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
Transcription and translation related genes (26)	Trascription	*rpoA, rpoB, rpoC1*, rpoC2*
	Ribosomal proteins	*rps2, rps3, rps4, rps7(x2), rps8, rps11, rps12**(x2), rps14, rps15, rps16*, rps18, rps19, rpl2*(x2), rpl14, rpl16*, rpl20, rpl22, rpl23(x2), rpl32, rpl33, rpl36*
	Translation initiation factor	*infA*
RNA genes (35)	Ribosomal RNA	*rrn4.5(x2), rrn5(x2), rrn16(x2), rrn23(x2)*
	Transfer RNA	*trnA-UGC*(x2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-UCC*, trnH-GUG, trnI-CAU(x2), trnI-GAU*(x2), trnK-UUU*, trnL-CAA(x2), trnL-UAA*, trnL-UAG, trnfM-CAU, trnM-CAU, trnN-GUU(x2), trnP-UGG, trnT-GGU, trnT-UGU, trnV-GAC(x2), trnV-UAC*, trnW-CCA, trnY-GUA, trnQ-UUG, trnR-ACG(x2), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnG-UCC, trnG-GCC*
Other genes (4)	RNA processing	*matK*
	Carbon metabolism	*cemA*
	Fatty acid synthesis	*accD*
	Proteolysis	*clpP**
Genes of unknown function (3)	Conserved reading frames	*ycf1(x2), ycf2(x2), ycf15*

* indicate intron containing genes, and genes located in the IR regions are indicated by (2x) after the gene name.

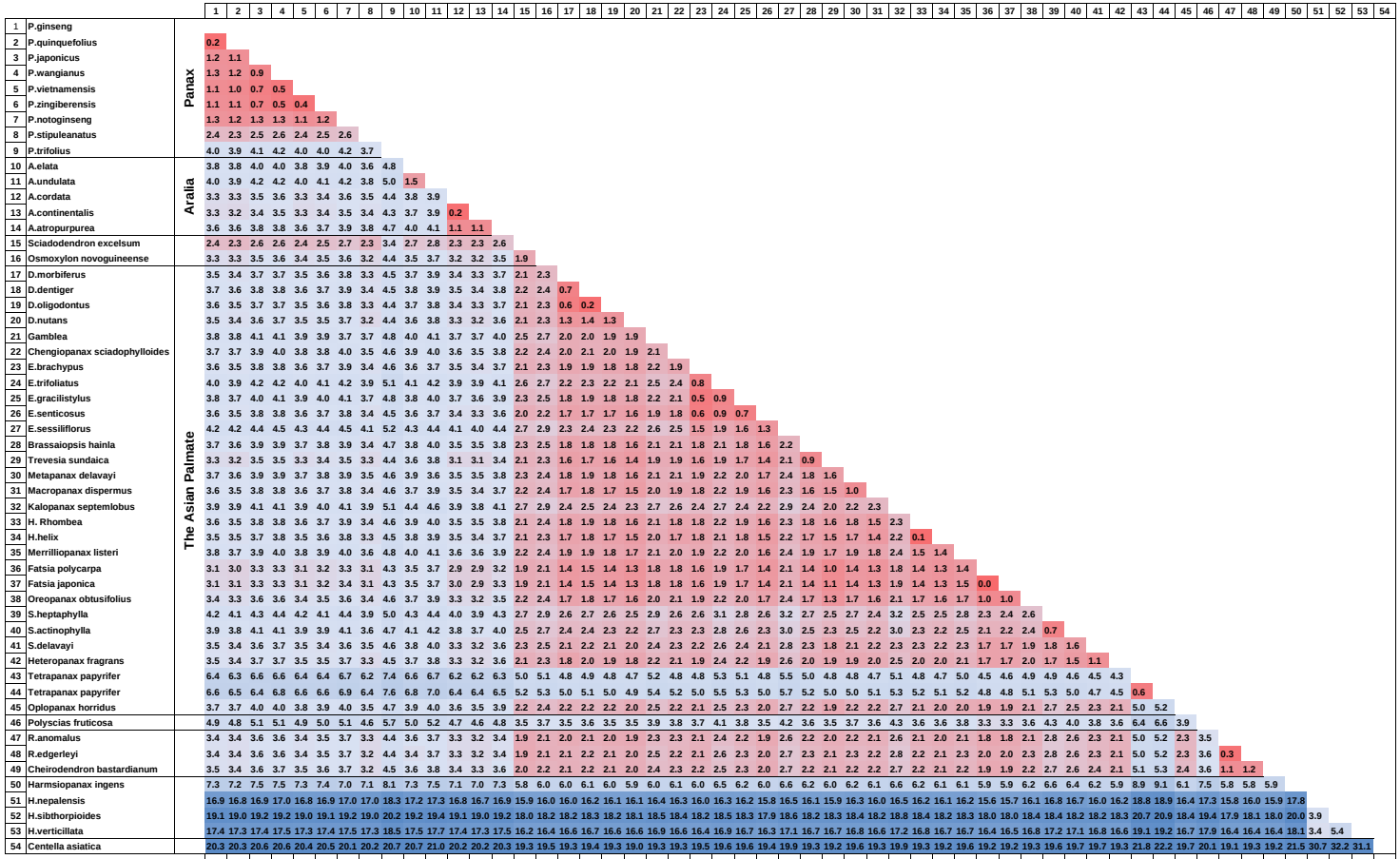

Figure S1-1. Divergence time estimated by the formula $T = Ks/2\lambda \times 10^{-6}$ MYA. $\lambda = 2 \times 10^{-9}$

Table S2-1. List of plastome and 45S rDNA sequences used for phylogenetic analysis (Supplementary Figure 2-2)

No	Sample ID	Primary Testing	Species	Locality	Collector(s)	Coll. Year
1	LC1	M	Pv.f	VN, Lai Chau Province	T. J. Yang et al.	2017
2	LC2	M	Pv.f	VN, Lai Chau Province	T. J. Yang et al.	2017
3	LC3	M	Pv.f	VN, Lai Chau Province	T. J. Yang et al.	2017
4	LC4	M	Pv.f	VN, Lai Chau Province	T. J. Yang et al.	2017
5	LC5	M	Pv.f	VN, Lai Chau Province	T. J. Yang et al.	2017
6	SP1	M	Ps	VN, Sa Pa District	T. J. Yang et al.	2017
7	SP2	M	Ps	VN, Sa Pa District	T. J. Yang et al.	2017
8	SP3	M	Ps	VN, Sa Pa District	T. J. Yang et al.	2017
9	SP4	M	Ps	VN, Sa Pa District	T. J. Yang et al.	2017
10	SP5	M	Ps	VN, Sa Pa District	T. J. Yang et al.	2017
11	SP6	M	Ps	VN, Sa Pa District	T. J. Yang et al.	2017
12	SP7	M	Ps	VN, Sa Pa District	T. J. Yang et al.	2017
13	SP8	M	Ps	VN, Sa Pa District	T. J. Yang et al.	2017
14	SP9	M	Pv.f	VN, Sa Pa District	T. J. Yang et al.	2017
15	SP10	M	Pv.f	VN, Sa Pa District	T. J. Yang et al.	2017
16	SP11	M	*Pn*	VN, Sa Pa District	T. J. Yang et al.	2017
17	SP12	M	*Pn*	VN, Sa Pa District	T. J. Yang et al.	2017
18	SP13	M	Pv	VN, Sa Pa District	T. J. Yang et al.	2017
19	TD1	M	Pv	VN, Tam Dao District	T. J. Yang et al.	2017
20	TD2	M	Pv	VN, Tam Dao District	T. J. Yang et al.	2017
21	TD3	M	Pv	VN, Tam Dao District	T. J. Yang et al.	2017
22	TD4	M	Pv	VN, Tam Dao District	T. J. Yang et al.	2017
23	TD5	M	Pv	VN, Tam Dao District	T. J. Yang et al.	2017
24	TD6	M	Ps	VN, Tam Dao District	T. J. Yang et al.	2017
25	TD7	M	Ps	VN, Tam Dao District	T. J. Yang et al.	2017
26	TD8	M	Pv	VN, Tam Dao District	T. J. Yang et al.	2017
27	TD9	M	Pv	VN, Tam Dao District	T. J. Yang et al.	2017
28	KT1	M	Pv	VN, Kon Tum Province	T. J. Yang et al.	2017
29	KT2	M	Pv	VN, Kon Tum Province	T. J. Yang et al.	2017
30	KT3	M	Pv	VN, Kon Tum Province	T. J. Yang et al.	2017
31	KT4	M	Pv	VN, Kon Tum Province	T. J. Yang et al.	2017
32	KT5	M	Pv	VN, Kon Tum Province	T. J. Yang et al.	2017
33	QN1	M	Pv	VN, Quang Nam Province	T. J. Yang et al.	2017
34	QN2	M	Pv	VN, Quang Nam Province	T. J. Yang et al.	2017
35	QN3	M	Pv	VN, Quang Nam Province	T. J. Yang et al.	2017
36	QN4	M	Pv	VN, Quang Nam Province	T. J. Yang et al.	2017
37	QN5	M	Pv	VN, Quang Nam Province	T. J. Yang et al.	2017
38	QN6	M	Pv	VN, Quang Nam Province	T. J. Yang et al.	2017
39	DL1	M	Pv.f	VN, cv in Da Lat City	T. J. Yang et al.	2019
40	DL2	M	Pv.f	VN, cv in Da Lat City	T. J. Yang et al.	2019
41	DL3	M	Pv.f	VN, cv in Da Lat City	T. J. Yang et al.	2019
42	DL4	M	Pv.f	VN, cv in Da Lat City	T. J. Yang et al.	2019
43	DL5	M	Pv.f	VN, cv in Da Lat City	T. J. Yang et al.	2019
44	DL6	M	Pv.f	VN, cv in Da Lat City	T. J. Yang et al.	2019
45	DL7	M	Pv	VN, cv in Da Lat City	T. J. Yang et al.	2019
46	DL8	M	Pv.f	VN, cv in Da Lat City	T. J. Yang et al.	2019
47	DL9	M	Pv.f	VN, cv in Da Lat City	T. J. Yang et al.	2019
48	DL10	M	Pv.f	VN, cv in Da Lat City	T. J. Yang et al.	2019
49	DL11	M	Pv.f	VN, cv in Da Lat City	T. J. Yang et al.	2019
50	DL12	M	Pv.f	VN, cv in Da Lat City	T. J. Yang et al.	2019
51	DL13	M	Pv.f	VN, cv in Da Lat City	T. J. Yang et al.	2019
52	DL14	M	Pv.f	VN, cv in Da Lat City	T. J. Yang et al.	2019
53	DL15	M	Pv.f	VN, cv in Da Lat City	T. J. Yang et al.	2019
54	DL16	M	Pv.f	VN, cv in Da Lat City	T. J. Yang et al.	2019
55	DL17	M	Pv.f	VN, cv in Da Lat City	T. J. Yang et al.	2019

56	DL18	M	Pv.f	VN, cv in Da Lat City	T. J. Yang et al.	2019
57	DL19	M	Ps	VN, cv in Da Lat City	T. J. Yang et al.	2019
58	DL20	M	Ps	VN, cv in Da Lat City	T. J. Yang et al.	2019
59	DL21	M	Ps	VN, cv in Da Lat City	T. J. Yang et al.	2019
60	DL22	M	Ps	VN, cv in Da Lat City	T. J. Yang et al.	2019
61	DL23	M	Ps	VN, cv in Da Lat City	T. J. Yang et al.	2019
62	DL24	M	Pv.l	VN, cv in Da Lat City	T. J. Yang et al.	2019
63	DL25	M	Pv.l	VN, cv in Da Lat City	T. J. Yang et al.	2019
64	DL26	M	Pv.l	VN, cv in Da Lat City	T. J. Yang et al.	2019
65	DL27	M	Pv.l	VN, cv in Da Lat City	T. J. Yang et al.	2019
66	DL28	M	Pv.l	VN, cv in Da Lat City	T. J. Yang et al.	2019
67	DL29	M	Ps	VN, cv in Da Lat City	T. J. Yang et al.	2019
68	DL30	M	Ps	VN, cv in Da Lat City	T. J. Yang et al.	2019
69	DL31	M	Ps	VN, cv in Da Lat City	T. J. Yang et al.	2019
70	DL32	M	Ps	VN, cv in Da Lat City	T. J. Yang et al.	2019
71	DL33	M	Ps	VN, cv in Da Lat City	T. J. Yang et al.	2019
72	DL34	M	Pv	VN, cv in Da Lat City	T. J. Yang et al.	2019
73	DL35	M	Pv	VN, market in Da Lat	T. J. Yang et al.	2019
74	DL36	M & Phy	Pv.f	VN, cv in Da Lat City	J. H. Park et al.	2019
75	DL37	M & Phy	Pv	VN, cv in Da Lat City	J. H. Park et al.	2019
76	DL38	M & Phy	Pv	VN, cv in Da Lat City	J. H. Park et al.	2019
77	DL39	M & Phy	Pv.f	VN, cv in Da Lat City	J. H. Park et al.	2019
78	DL40	M & Phy	Pv.f	VN, cv in Da Lat City	J. H. Park et al.	2019
79	DL41	M & Phy	Pv.f	VN, cv in Da Lat City	J. H. Park et al.	2019
80	DL42	M & Phy	Pv.f	VN, cv in Da Lat City	J. H. Park et al.	2019
81	DL43	M & Phy	Pv.f	VN, cv in Da Lat City	J. H. Park et al.	2019
82	DL44	M & Phy	Pv.f	VN, cv in Da Lat City	J. H. Park et al.	2019
83	DL45	M & Phy	Pv.f	VN, cv in Da Lat City	J. H. Park et al.	2019
84	DL46	M & Phy	Pv.l	VN, cv in Da Lat City	J. H. Park et al.	2019
85	DL47	M & Phy	Pv.f	VN, cv in Da Lat City	J. H. Park et al.	2019
86	DL48	M & Phy	Pv.l	VN, cv in Da Lat City	J. H. Park et al.	2019
87	DL49	M & Phy	Pv.l	VN, cv in Da Lat City	J. H. Park et al.	2019
88	DL50	M & Phy	Pv	VN, cv in Da Lat City	J. H. Park et al.	2020
89	DL51	M & Phy	Pv	VN, cv in Da Lat City	J. H. Park et al.	2020
90	DL52	M & Phy	Pv.f	VN, cv in Da Lat City	J. H. Park et al.	2020
91	DL53	M & Phy	Pv.f	VN, cv in Da Lat City	J. H. Park et al.	2020
92	DL54	M & Phy	Pv	VN, cv in Da Lat City	J. H. Park et al.	2020
93	DL55	M & Phy	Pv	VN, cv in Da Lat City	J. H. Park et al.	2020
94	DL56	M & Phy	Pv.f	VN, cv in Da Lat City	J. H. Park et al.	2020
95	DL57	M & Phy	Pv	VN, cv in Da Lat City	J. H. Park et al.	2020
96	DL58	M & Phy	Pv	VN, cv in Da Lat City	J. H. Park et al.	2020
97	DL59	M & Phy	Pv	VN, cv in Da Lat City	J. H. Park et al.	2020
98	DL60	M & Phy	Pv	VN, cv in Da Lat City	J. H. Park et al.	2020
99	DL61	M & Phy	Pv	VN, cv in Da Lat City	J. H. Park et al.	2020
100	DL62	M & Phy	Pv	VN, cv in Da Lat City	J. H. Park et al.	2020
101	DL63	M & Phy	Pv	VN, cv in Da Lat City	J. H. Park et al.	2020
102	DL64	M & Phy	Pv	VN, cv in Da Lat City	J. H. Park et al.	2020
103	DL65	M & Phy	Pv	VN, cv in Da Lat City	J. H. Park et al.	2020
104	DL66	M & Phy	Pv.f	VN, cv in Da Lat City	J. H. Park et al.	2020
105	DL67	M & Phy	Pv	VN, cv in Da Lat City	J. H. Park et al.	2020
106	DL68	M & Phy	Pv	VN, cv in Da Lat City	J. H. Park et al.	2020
107	DL69	M & Phy	Pv.f	VN, cv in Da Lat City	J. H. Park et al.	2020
108	DL70	M & Phy	Pv	VN, cv in Da Lat City	J. H. Park et al.	2020
109	DL71	M & Phy	Pv	VN, cv in Da Lat City	J. H. Park et al.	2020
110	DL72	M & Phy	Pv.f	VN, cv in Da Lat City	J. H. Park et al.	2020
111	DL73	M & Phy	Pv.f	VN, cv in Da Lat City	J. H. Park et al.	2020
112	DL74	M & Phy	Pv.f	VN, cv in Da Lat City	J. H. Park et al.	2020
113	DL75	M & Phy	Pv	VN, cv in Da Lat City	J. H. Park et al.	2020
114	DL76	M & Phy	Pv	VN, cv in Da Lat City	J. H. Park et al.	2020

Abbreviations: Pv, *P. vietnamensis* 3 var: v, *vietnamensis*; l, *langbianensis*; f, *fuscidiscus*; Ps, *P. stipuleanatus*. M, morphology; M&Phy, morphology & phytochemicals.

Table S2-2. List of plastome and 45S rDNA sequences used for phylogenetic analysis (Supplementary Figure 2-2)

		Genebank	
No	**ID**	**Chloroplast genome**	**45S nrDNA**
1	**Pg_1**	KM088019.1	KM036295.1
2	**Pg_2**	KM067394.1	KM207672.1
3	**Pg_3**	KM067388.1	KM207667.1
4	**Pq_1**	KM088018.1	KM036297.1
5	**Pq_2**	MK408923.1	MK408799.1
6	**Pj**	KP036469.1	KT380920.1
7	**Pj.b_1**	MK408918.1	MK408792.1
8	**Pj.m_1**	MK408962.1	MK408761.1
9	**Pj.b_2**	MK408926.1	MK408776.1
10	**Pj.m_2**	MK408940.1	MK408800.1
11	**Pj.m_3**	MK408948.1	MK408782.1
12	**Pw_1**	MK408921.1	MK408797.1
13	**Pw_2**	MK408963.1	MK408805.1
14	**Pw_3**	MK408964.1	MK408801.1
15	**Pv.l_1**	MT798583	MW374462.1
16	**Pv.l_2**	MT798584	MW374329.1
17	**Pv.v_1**	KP036471.1	MW374467.1
18	**Pv.v_2**	KU059178.1	MW374330.1
19	**Pv.v_3**	KP036470.1	KT380922.1
20	**Pv.f_1**	MT798587	MW374471.1
21	**Pv.f_2**	MT798586	MW374464.1
22	**Pv.f_3**	MT798585	MW374463.1
23	**Pz_1**	MK408960.1	MK408808.1
24	**Pz_2**	MK408947.1	MK408783.1
25	**Pz_3**	MK408933.1	MK408768.1
26	**Pz_4**	MK408922.1	MK408794.1
27	**Pn_1**	MK408931.1	MK408766.1
28	**Pn_2**	KP036468.1	KT380921.1
29	**Pn_3**	MK408945.1	MK408777.1
30	**Ps_1**	KX247147.1	MF091695.1
31	**Ps_2**	MK408920.1	MK408807.1
32	**Ps_3**	MK408936.1	MK408811.1
33	**Pt**	MF100782.1	MF099781.1

Abbreviations: Pg, *P. ginseng*; Pq, *P. quinquefolius*; Pj, *P. japonicus*: b, *bipinnatifidus* and m, *major*; Pw, *P. wangianus*; Pv, *P. vietnamensis*: v, *vietnamensis*; l, *langbianensis*; f, *fuscidiscus*; Pz, *P. zingiberensis*; Pn, *P. notoginseng*; Ps, *P. stipuleanatus*; Pt, *P. trifolius*. New sequences were highlighted.

Table S2-3. Tukey's five-number summary of Ka and Ks

Gene	Ka					Ks				
	min	**Q1**	**med**	**Q3**	**max**	**min**	**Q1**	**med**	**Q3**	**max**
psbA	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.03
matK	0.00	0.00	0.00	0.01	0.02	0.00	0.01	0.01	0.02	0.05
rps16	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.07
psbK	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.06
psbI	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08
atpA	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.03
atpF	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.01	0.02
atpH	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.02
atpI	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.06
rps2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.04
rpoC2	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.03
rpoC1	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.03
rpoB	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02
petN	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10
psbM	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.07	0.14
psbD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
psbC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.03
psbZ	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.02	0.05
rps14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.05
psaB	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03
psaA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.03
ycf3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.04
rps4	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.02	0.02	0.04
ndhJ	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.04
ndhK	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.06
ndhC	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.02	0.03
atpE	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.01	0.03
atpB	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.04
rbcL	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.01	0.02	0.03
accD	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.03
psaI	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03
ycf4	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.03	0.06
cemA	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.02
petA	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.02	0.03
psbJ	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.03
psbL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03
psbF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
psbE	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.02
petL	0.00	0.00	0.00	0.02	0.02	0.00	0.00	0.03	0.03	0.06
petG	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.09
psaJ	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05
rpl33	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.01
rps18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02
rpl20	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.03
clpP	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.02	0.06
psbB	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.03
psbT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
psbN	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
psbH	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.03
petB	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
petD	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.03

rpoA	0.00	0.00	0.01	0.01	0.02	0.00	0.00	0.00	0.01	0.02
rps11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05
rpl36	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08
infA	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.05	0.08
rps8	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.02	0.05
rpl14	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.06
rpl16	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.02
rps3	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.03
rpl22	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.01	0.02	0.08
rps19	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.09
rpl2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
rpl23	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ycf2	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01
ycf15	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
ndhB	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
rps7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
ndhF	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.01	0.03
rpl32	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.10
ccsA	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.01	0.01	0.03
ndhD	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.01	0.01	0.02
psaC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.09
ndhE	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.01	0.02	0.08
ndhG	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.04
ndhI	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.01	0.04
ndhA	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.02
ndhH	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.03	0.07
rps15	0.00	0.00	0.01	0.01	0.03	0.00	0.00	0.00	0.00	0.04

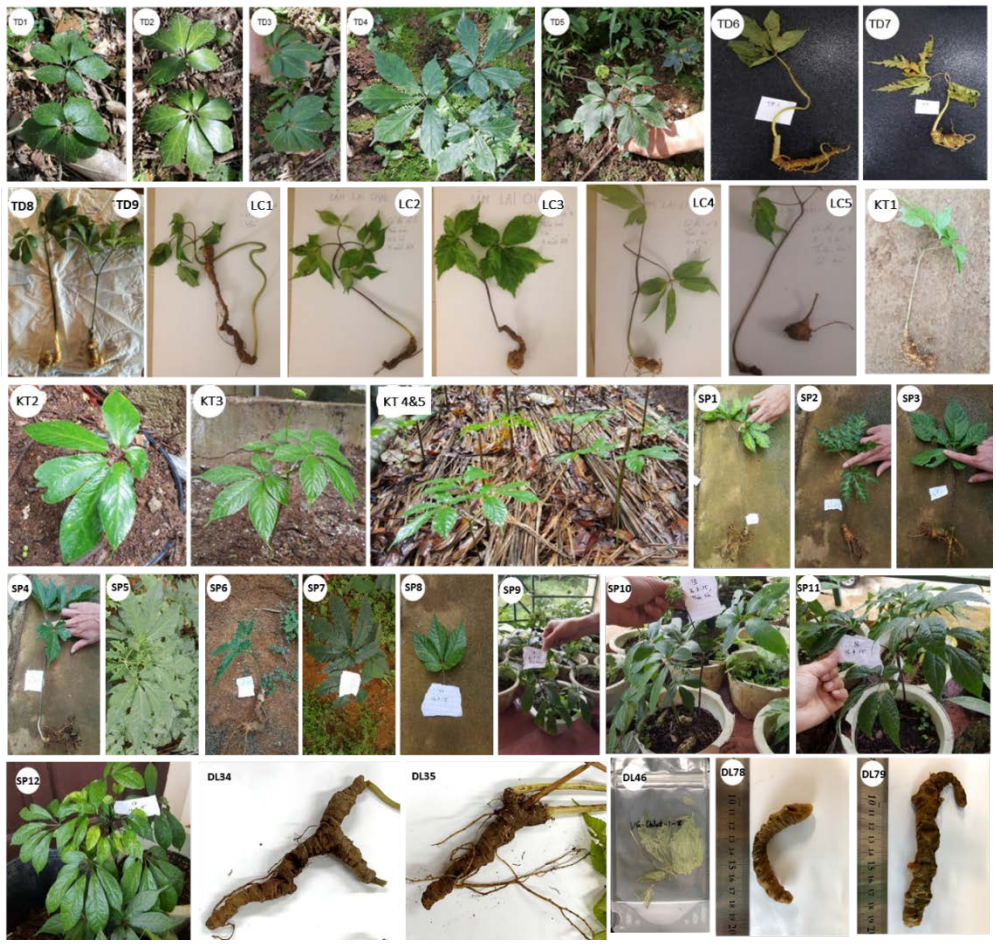

Figure. S2-1. Panax specimens collected from six provinces in Vietnam. Sample IDs are followed Supplementary Table S2-1.

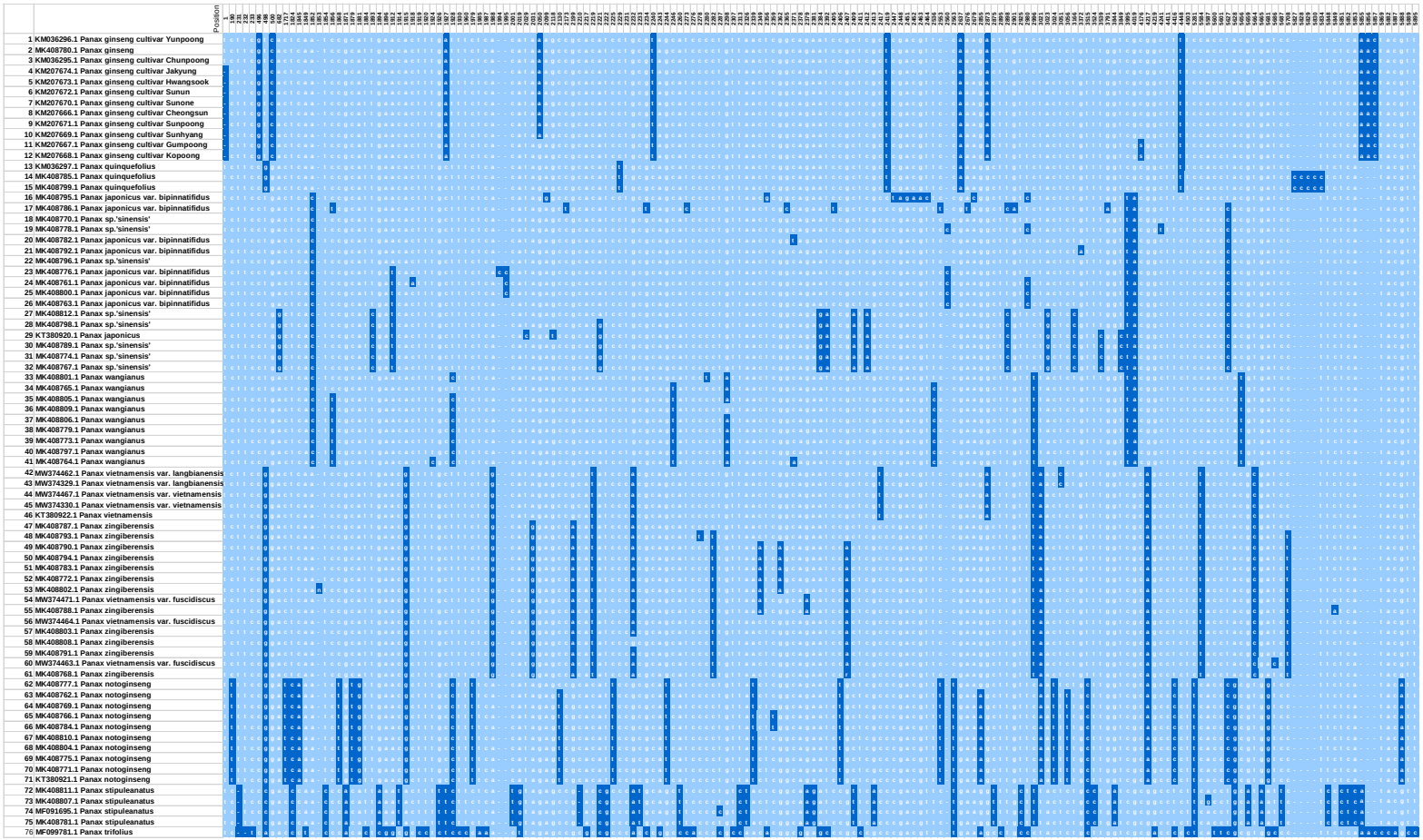

Figure S2-2. 179 nucleotide variant sites in 45S rDNA coding units of 76 Panax accessions

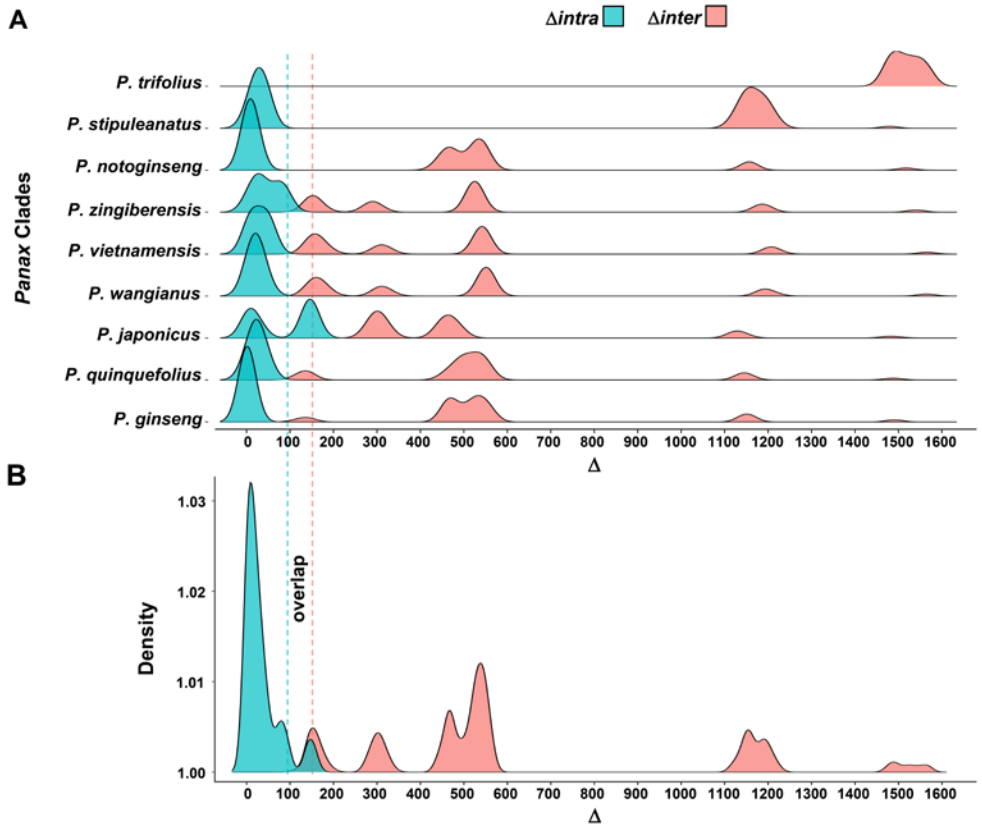

Figure S2-3. Ridgeline plot of smoothed density estimates of Δ_{intra} and Δ_{inter}.

(A) Each clade. (B) All clades.

ACKNOWLEDGEMENTS

I have many people to thank for their significant contributions to the successful accomplishment of my dissertation research. First, I want to express my sincere appreciation to my advisor, Prof. Dr. Tae-Jin Yang, for his supervision and wholehearted support over the last five years. He patiently (and in the end successfully) directed my attention to the super barcoding of Araliaceae. Meanwhile, I appreciate fruitful and critical discussions and a wide range of professional comments on the various chapters of my dissertation. I am grateful for his extraordinary compassion and trust throughout my extended visits to my hometown around the holidays and for offering the space and freedom to explore and pursue my ideas and research interests.

I would like to thank my committee members Prof. Dr. Paek Nam Chon, Prof. Dr. Kim Do Soon, Prof. Dr. Park Jeong Hill, Prof. Dr. Nguyen Van Binh, for serving as my committee members. All of them had a fundamental impact during my Ph.D. study. Your teachings have continually provided me with knowledge, motivation, and stability throughout my life. I can't describe how grateful I am for everything you've taught me. Furthermore, I want to thank you for making my defense a pleasurable experience, as well as for your wonderful comments and suggestions.

I'm also in debt to several doctoral colleagues who have kept track of my Ph.D. research, commenting on the progress and offering ideas for improvement. Dr. Park Hyun-Seung has been a fantastic mentor for me, and without his help, my job would have been a lot more difficult. I learned a lot from him because of his scientific knowledge, his ability to solve seemingly impossible practical problems, and his ability to put complicated ideas into simple terms. Dr. Park Hyun Seung and Dr. Kang Jong Soo significantly impacted my Ph.D. research and provided essential advice and

recommendations, critical discussion and comments on many areas of dissertations, last-minute revisions, and most significantly, patient and consistent support. Many thanks also to Dr. Park Jee Young, who is in charge of lab managing. I would like to thank Dr. Padmanaban Mohanan for sharing his great experience, comments and suggestions, and friendly hours inside and outside the lab.

Every result described in this thesis was accomplished with the help and support of fellow labmates: Dr. 유홍섭, Joh Ho Jun, Dr. Gong Haiguang, Koo Hyunjin, Lee Sae Hyun, Shim Hyeonah, Park Young Sang, Cho Woohyeon, Kim JinTae, Kim Jiseok, Kim Eunbi, Lee Yeonjeong, Jeong Yonghyeok, Jeong Seonheui, Lee Minyoung, Ms. 최화춘, Mr. 박세원, Mr 변찬우. They provided a friendly and cooperative atmosphere at work, helpful feedback, and insightful comments on my work.

I would like to acknowledge friends and family who supported me during my time here.

This work is dedicated to my deceased parents, who always encouraged me to pursue higher education. Although you are no longer with us, but your belief in me has made this journey possible.