A DISSERTATION FOR

THE DEGREE OF DOCTOR OF PHIILOSOPHY

Dynamic evolution of plastome structure in Apioideae

산형아과 식물의 엽록체 게놈 진화 연구

By

HO JUN JOH

February 2022

INTERDISCIPLINARY PROGRAM IN

AGRICULTURAL GENOMICS

COLLEGE OF AGRICULTURE AND LIFE SCIENCES

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

DYNAMIC EVOLUTION OF PLASTOME STRUCTURE IN APIOIDEAE

UNDER THE DIRECTION OF DR. TAE-JIN YANG

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

By

HO JUN JOH

Interdisciplinary Program in Agricultural Genomics

FEBRUARY, 2022

APPROVED AS QUALIFIED DISSERTATION OF HO JUN JOH FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

BY THE COMMITTEE MEMBERS

CHAIRMAN _____(Seal)

Byoung-Cheorl Kang, Ph.D.

VICE-CHAIRMAN _____(Seal)

Tae-Jin Yang, Ph.D.

MEMBER _____(Seal)

Jin Hoe Huh, Ph.D.

MEMBER _____(Seal)

Myounghai Kwak, Ph.D.

MEMBER _____(Seal)

Ki-Oug Yoo, Ph.D.

Dynamic evolution of plastome structure in Apioideae

Ho Jun Joh

Interdisciplinary Program in Agricultural Genomics

The Graduate School Seoul National University

GENERAL ABSTRACT

Apioideae, one of the largest sub-families under the Apiaceae family, comprises economic, herbal, and medicinally important plants distributed worldwide. Plastid gene and nuclear ribosomal gene sequences have been used for classification of the Apioideae sub-family. However, previous phylogenetic studies showed an ambiguous relationship associated with monophyly of genera or clades. To overcome the limitation, the complete plastid genomes (plastomes) and nuclear ribosomal DNA (nrDNA) sequences were used for phylogenetic reconstruction of the sub-family and investigation of inter- and intraspecific patterns within the Apioideae plastomes.

In Chapter I, the interspecific relationship among Apioideae species was investigated through the phylogenetical analysis. The complete plastomes of 31 Apioideae species distributed in Korea were newly assembled in this study, and then

total of 60 Apioideae species were selected to infer phylogenetic relationships of the sub-family using the Maximum Likelihood (ML) method. The phylogenetic relationships within the Apioideae were similar to other studies. The plastome sizes observed in Apioideae ranged from 147 kbp ~ 158 kbp. To understand the plastome evolution of the Apioideae, plastome structure, especially the junction between large single copy (LSC) and inverted repeat (IR) regions was investigated. It confirmed that Apioideae species could be classified into three major groups based on large-scale IR length changes. Within the three groups, four different dynamic events (two significant IR reductions in groups 2 and 3 and two random species level or intraspecies level expansions in group 3) related to the changes of IR junction were discovered. The IR lengths of Group 1, 2, and 3 were approximately around 25, 24, and 18 kbp. Group 1 contained the original IR junction near the *rps19* gene, similar to the neighbor family Araliaceae species containing ginseng. Species in Group 2 possessed intermediated IR junctions around the *rps2* gene. Although four major tribes belong to Group 3, they had the most reduced IR junction near the *ycf2* gene by an IR contraction. Random and recent dynamic IR expansions up to 27 kbp were also identified in some species of Group 3. Furthermore, the most surprising intraspecific IR junction expansion up to 17 kbp was discovered in *Peucedanum japonicum*. Finally, the dynamic plastome structure changes in the Apioideae provides an evolutionary model for the plastome evolution of angiosperms.

In Chapter II, the most recent dynamic intraspecific IR expansion in the plastome of *Peucedanum japonicum* was investigated. *P. japonicum* is a native plant in Korea with promising beneficial properties as a medicinal product. To understand the

genetic diversity in *P. japonicum* at the intraspecific level, the plastomes of nine *P. japonicum* accessions among 38 *P. japonicum* germplasm collected from various places in Korea were compared. Nine *P. japonicum* plastome could be sorted into two types differing in by the length of IR region, long (L: 35,760 bp) and short (S: 18,606 bp) types, which showed 17,154 bp IR length difference. Phylogenetic analysis including nine accessions of *P. japonicum*, and Selineae species illustrated that the S-type plastomes of *P. japonicum* accession had an identical IR junction with other Selineae species, suggesting that the L-types plastomes have diverged very recently from the S-type plastomes in *P. japonicum*. The analysis of developed molecular markers in this study such as one L-and S-type classification markers, 17 InDels, and eight KASPs showed that the L-type possessed a low level of intraspecific polymorphisms whereas the S-type had higher diversity. SNP-based KASP marker analysis distinguished each type into sub-groups; the L-type into two and the S-type into four groups. Categories on the additional types of *P. japonicum* accessions through intraspecific polymorphisms provided an extensive genetic diversity of *P. japonicum*. Molecular tools to identify intraspecific diversity will benefit managing the genetic resources and breeding of the wild resource plants.

In summary, the diversification of the plastomes, both inter and intraspecific analysis on Apioideae species, provides a valuable resource for understanding the dynamic evolution of the plastomes, and the phylogenetic relationship in the Apioideae. Furthermore, in-depth analysis of plastomes and nrDNAs and development of molecular tools provide useful and essential tools for managing diverse wild germplasm and breeding of *P. japonicum*.

Keywords: *Apioideae, Peucedanum japonicum, Phylogenetic analysis, Molecular marker*

Student number: 2014-22947

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

LSC	Large single copy
SSC	Small single copy
SC	Single copy
IR	Inverted repeat
TR	Tandem repeat
InDel	Insertion Deletion
ITS	Intergenic spacer region
CDS	Protein Coding Sequence
KASP	Kompetitive Allele Specific PCR
MAF	Major allele frequency
MAS	Marker assisted selection
Mya	Million years ago
PCR	Polymerase chain reaction
SNP	Single nucleotide polymorphism
TR	Tandem repeat

GENERAL INTRODUCTION

Apiaceae is one of the largest groups of flowering plants under Apiales order, comprised of 434 genera and 3700 species [1-3]. Species under the Apiaceae family contain economic, herbal, and medicinally important species like carrots, parsley, caraway. The most renowned morphological trait representing the Apiaceae species is the flowering trait 'umbel,' an umbrella shape of a flower [1]. Numerous morphological-based classification were applied for the Apiaceae [4-7]. Around the year 1896, Drude first announced three sub-families of Apiaceae, including Apioideae; the largest sub-family containing 41 major clades and 21 tribes [6]. However, the ambiguity in morphological-based classification led to the replacement into molecular-based classification near the end of the 20th century. Numerous studies on the phylogenetic relationships among the Apiaceae through universal barcoding markers have proposed the classification on sub-family and species level [8-10].

Plastid (or chloroplast) is a key organelle of photosynthesis and other biomechanical pathways in the plants converting light into an essential energy source for a plant to survive. Plastid genome (plastome) is a circular structure with a quadripartite structure comprising two single-copy (SC) regions and two inverted repeat (IR) regions. Plastome range between 120~217 kbp in most plants [8~10]. Plastome represents very low intra-species polymorphism than inter-species polymorphism provided by the conserve characteristics from circular genome structure, uni-parental inheritance, and less heteroplasmy and recombination, compared to nuclear and mitochondria genomes [8, 11]. In addition, plastomes present a small genome size with

a high copy number in a single cell and conserved characteristics with a moderate substitution rate, and plastomes are most often used as a target material on comparative analysis in plant species [12]. Before sequencing technology advanced, plastid gene sequences as DNA barcoding regions were primarily used in plant authentication, phylogenetic, and evolutionary studies [8, 13, 14]. However, high-throughput sequencing technologies are rapidly developing, gradually reducing costs. These advances the environment for large-scale plastome analysis to resolve ambiguous relationships associated within Apioideae species.

Plastid genome sizes are different around plant species. The genome sizes are greatly affected by the changes of inverted repeat (IR) region in length. The IR region on the plastome is around 25 kbp in most plants and typically contains ribosomal operons that include four rRNA encoding 4.5S, 5S, 16S, and 23S rRNAs [15]. Apart from the ribosomal operons, gene contents within the IR region are determined by the boundaries between SC regions. The function of the IR region is recognized as the maintenance of the stability of the plastome structure. The stability of IR comes from the two identical copies located between the SC region and providing error correction from the mutations. Such properties allow the nucleotide substitution rate within the IR region is lower than that within the SC regions [16, 17]. The expansion or contraction of the IR region is often observed in comparative analysis of plastomes on an interspecific level. Although it is different by the species, the variation of IR lengths is from 13 kbp to 75 kbp. On the other hand, plant species that acquire energy without photosynthesis dependent, such as parasite plants, show small plastome size with

missing a single-copy of IR regions referred to as Inverted Repeat Lacking Clade (IRLC). [18, 19]. Variations in length and gene contents on the IR region provide an information how the genome has been evolved. Plastome based studies have mainly addressed the complete assembly of the plastome for target species and comparative with closely related species on an interspecific level. On the other hand, comparative studies on an intraspecific level are rare due to the conserved feature provided by the plastid genome. On the intraspecific level, plastome would provide less than 1 kbp difference in plastome size mostly from polymorphism composed of repeated sequences discovered within SC regions rather than IR regions. For such reason, no studies have focused on analyzing the genetic variation at the intraspecific level.

Peucedanum japonicum Thunberg is a member of the *Peucedanum* genus under the Selineae tribe of the Apioideae. Its habitat is near the coastal region around the cliff area, distributed in the East Asian region The root of *P. japonicum* has been used as a traditional oriental medicine, and the aerial parts of this species such as leaves and flowers, are widely used for manufacture herbal tea, diet product, and vegetable food. A previous study have reported a complete assembly on two *P. japonicum* plastomes with a genome size of 164,653 bp without any intraspecific polymorphism [20]. Despite the low variation rate of the plastome, plastome sizes are generally differ by species. Given the situation that both two samples in the previous study were grown collections of *P. japonicum* germplasms from different institutions, the genetic diversity of the two cultivated *P. japonicum* germplasm was estimated to be lower than the wild type [20].

This study assembled a complete plastid genome of various Apioideae species,

including rare species distributed in Korea, to understand the phylogenetic relationship within Apioideae species. In addition, a comparative analysis of the complete plastome sequence examined how the plastome structure, particularly IR junction, has been changed from sister clade Araliaceae toward the Selineae tribe. From this study, another three more *P. japonicum* species collected from both wild and cultivated regions was sequenced to analyze the changes over the intraspecific level of the plasomes and developed molecular marker to apply to germplasms to uncover extensive genetic diversity. The results in this study would further advance current understanding of Apioideae and demonstrate that the plastome sequence is an efficient tool for future breeding research on Apioideae species.

REFERENCES

1. Drude, O. 1897. "Umbelliferae." *Die naturlichen Pflanzenfamilien*, 3:63-150.

2. Cronquist, A., and Takhtadzhiãn, AL. 1981. *An integrated system of classification of flowering plants*: Columbia university press.

3. Pimenov, MG., and Leonov, MVE. 1993. *The genera of the Umbelliferae: a nomenclator*: Royal Botanic Gardens, Kew.

4. Koch, WDJ. 1825. *Generum tribuumque plantarum umbelliferarum nova dispositio: Cum tab. tribus lithograptis*.

5. Bentham, G., and Hooker, JD. 1867. Umbelliferae. *Genera plantarum* 1(3), 859-931.

6. Drude, O. 1897. Umbelliferae. *Die naturlichen Pflanzenfamilien*, 3, 63-150.

7. Koso-Poljansky, BM. 1916. "Sciadophytorum systematis lineamenta." *Bull. Soc. Imp. Naturalistes Moscou* 29, 93-221.

8. Downie, SR., Plunkett, GM., Watson, MF., Spalik, K, Katz-Downie, DS., Valiejo-Roman, CM., Terentieva, EI., Troitsky, AV., Lee, BY., … and Lahham, J. 2001. Tribes and clades within Apiaceae subfamily Apioideae: the contribution of molecular data. *Edinburgh Journal of Botany* 58(2), 301-330.

9. Downie, SR., Krzysztof, S., Katz-Downie, DS., and Reduron, JP. 2010. Major clades within Apiaceae subfamily Apioideae as inferred by phylogenetic analysis of nrDNA ITS sequences. Plant Diversity and Evolution 128(1), 111.

10. Plunkett, GM. 2001. Relationship of the order Apiales to subclass Asteridae: a re-evaluation of morphological characters based on insights from molecular data.

Edinburgh Journal of Botany 58(2), 183-200.

11. Sugiura, M. 1992. The chloroplast genome. *Plant molecular biology* 19(1), 149-168.

12. Wicke, S., Schneeweiss, GM., Depamphilis, CW., Müller, KF., and Quandt, D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology* 76(3), 273-297.

13. Spalik, K., Reduron, JP., and Downie, SR. 2004. The phylogenetic position of *Peucedanum* sensu lato and allied genera and their placement in tribe Selineae (Apiaceae, subfamily Apioideae). *Plant Systematics and Evolution* 243(3), 189-210.

14. Hollingsworth, PM., Graham, SW., and Little, DP. 2011. Choosing and using a plant DNA barcode. *PloS one* 6(5), e19254.

15. Ruhlman, TA., and Jansen, RK. 2014. *The plastid genomes of flowering plants. In Chloroplast Biotechnology*, 3-38. Totowa, NJ: Humana Press.

16. Perry, AS., and Wolfe, KH. 2002. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *Journal of Molecular Evolution* 55(5), 501-508.

17. Zhu, A., Guo, W., Gupta, S., Fan, W., and Mower, JP. 2016. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytologist* 209(4), 1747-1756.

18. Chumley, TW., Palmer, JD., Mower, JP., Fourcade, HM., Calie, PJ., Boore, J. L., and Jansen, RK. 2006. The complete chloroplast genome sequence of Pelargonium× hortorum: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Molecular Biology and Evolution* 23(11), 2175-2190.

19. Sabir, J., Schwarz, E., Ellison, N., Zhang, J., Baeshen, NA., Mutwakil, M., … and Ruhlman, T. (2014). Evolutionary and biotechnology implications of plastid genome variation in the inverted-repeat-lacking clade of legumes. *Plant Biotechnology Journal* 12(6), 743-754.

20. Lee, HO., Joh, HJ., Kim, K., Lee, SC., Kim, NH., Park, JY., … and Yang, TJ. 2019. Dynamic chloroplast genome rearrangement and DNA barcoding for three apiaceae species known as the medicinal herb "Bang-Poong". *International Journal of Molecular Sciences* 20(9), 2196.

CHAPTER I

Dynamic plastome structural evolution of Apioideae species distributed in Korea

ABSTRACT

Sub-family Apioideae is one of the largest groups within the Apiaceae. Classifying the clade and tribe within Apioideae was one of the most prominent debates among scientists due to the lack of morphological discontinuity in many morphological characters. To perform phylogenetic analysis on Apioideae, molecular phylogeny employing plastid gene and nuclear ribosomal internal transcribed spacer or DNA barcoding was established around the turn of the century. Despite this, DNA barcoding using universal markers were restricted due to few of loci and their parsimony informative sites. As a result, phylogenetic researches focus on comprehensive plastome analysis, known as "super-barcoding." Approximately 100 Apioidese species are distributed in Korea, including 13 endemic and rare species, but there are very few genetic studies accomplished on the species.

As a part of genetic studies on Apioideae, in this study, the complete plastomes from 31 Apioideae species, including three endemic and rare species distributed in Korea, were assembled. A total of 60 Apioideae plastomes, including a sister family Araliaceae are used in this study, and are divided into three groups related to the IR junctions (*rps19*, *rpl2*, and *ycf2*) of plastome structure. Among the three groups, four events related with IR contraction have been illustrated throughout the comparative analysis based on phylogenetic relationship of Apioideae plastome. Moreover, Apioideae also illustrated a recent IR expansion including *trn*H-GUG - partial *psbA* gene insertion in IRa region along with an intraspecific IR expansion and tRNA gene order inversion. Compared to the sister family Araliaceae, Apioideae plastomes

illustrated a dynamic evolution in plastome structure. The valuable information provided from this study will be helpful for the understanding of relationship of Apioideae species.

Keywords: Apioideae, Plastome, Phylogenetic Analysis

INTRODUCTION

Apiaceae is one of the largest groups of flowering plant under Apiales order, comprised of sub-families including Apioideae, Saniculoideae, Azorelloideae, and Makinlayoideae [1]. Among the four sub-families, Apioideae is composed of 41 major clades and 21 tribes [2]. The complexity of the morphological similarities, including the renowned flowering trait 'umbel,' an umbrella shape of a flower, for classification of Apioideae was still unclear. Numerous morphological-based taxonomic studies were conducted on Apioideae, over the defining classifications on Apioideae [3~6]. Different methods were utilized to define the classification of the family Apiaceae so far. Around 1896, Drude proposed the first classification for the family Apiaceae that three sub-families with the crystallized form of calcium oxalate in the pericarp and other separated tribes and clades using morphological traits such as the shape of endosperm and flower [5]. Drude proposed the classification based on the Apiaceae species distributed in Europe [5]. However, the consideration of identifying trait-related mutation aspects and monophyletic relationships were ignored in the classification proposed by Drude [5]. Koso-Poljansky redefined the subfamilies of Apiaceae into two sub-families (sub-familes Hydrocotyloideae and Ligusticoideae) with morphological traits such as a number of canel and mericarp surface [6]. Cerceau-Larrival redefined the Apiaceae into five sub-families Bupleuroideae, Endressioideae, Azorelloideae, Eryngiooideae, and Apioideae with morphology traits on pollen, cotyledon, and flowering shape [7].

Near the end of the 20th century, molecular-based phylogenetic studies have widely performed for understanding phylogenetic relationship within Apioideae species

using plastid genes and nuclear ribosomal ITS region. The phylogenetic studies using molecular markers have revealed numerous sub-tribes within the Apioideae [8~16]. However, the studies have mentioned that the phylogenetic analysis through the ITS sequence is inadequate to investigate intertribal and deep relationships within Apioideae due to short sequence length and demonstrating high homoplastic within the phylogeny [2, 17, 18]. As for the plastid genome, the availability of limited loci (*rbcl*, *matK*, *rpo*C2) along with limited parsimony information still hinders our understanding of the ambiguous relationship between morphological features and molecular phylogenetic investigation [10, 19, 20]. To overcome the limit regarding small parsimony information provided from plastid gene sequence, whole plastome comparative analysis known as 'super barcoding' method is implemented as a new trend for revealing the ambiguous relationship. With the rapid development of sequencing technology, about 100 Apioideae species are available for complete plastome sequences on the NCBI Genbank database (https://www.ncbi.nlm.nih.gov/genome/organelle). Using super barcoding method would provides enough parsimony informative sites and genomic features for understanding the plastome evolution.

Whole plastome comparative analysis on Apioideae was focused on analyzing either the changes in plastome structure or the resolution of the ambiguous phylogenetic relationship between intra-genus or inter-species level of Apioideae on a small scale. For example, a study by Downie et al. has analyzed the relationship of Apiales species, including a small number of Apioideae species plastomes, and indicated a change in the IR junction with a reference that IR can be used as a phylogenetic marker [21]. While, the study by Spooner et al. focused on investigating the phylogenetic analysis of

complete plastome on the *Daucus* genus and discovered an incongruent phylogenetic relationship between plastome and nuclear genes [22]. Large scale phylogenetic studies using complete plastome within Apioideae studies remained poorly until a recent study by Wen et al. proposed a study of comparative analysis on Apioideae species to resolve the monophyletic relationship under Apioideae species [23].

About 34 genera with 75 Apioideae species are distributed in Korea, and most of the studies were primarily focused on phylogenetic-related studies to understand the evolutionary history using ITS sequences rather than plastome data. Other previous studies using plastome data elucidated the monophyletic relationship under Apioideae, but a polyphyletic relationship remains within the genus level. Therefore, to better understand the Apioideae species, phylogenetic studies focusing on other areas such as IR junction as a target field are necessary. Therefore, the focus of this study is to complete the plastid genome on 31 Apioideae species distributed in Korea and investigate the plastome variation around IR junctions within Apioideae species through comparative and phylogenetic analysis to understand the evolutionary history of the plastid genome in the Apioideae species.

MATERIAL AND METHODS

Plant material, DNA extraction, and sequencing

A total of 31 Apioideae species were collected for DNA extraction with the support of three institutions (Medicinal herb garden; College of Pharmacy Seoul National University, Seoul National University Forests, Hantaek Botanical Garden). Seven *Angelica* species, *Anthriscus sylvestris*, three *Bupleurum* species, *Centella asiatica*, *Conioselinum tenuissimum*, three *Cnidium* species, *Coriandrum sativum*, *Cryptotaenia japonica*, *Dystaenia takesimana*, *Heracleum moellendorffii*, *Levisticum officinale*, *Oenanthe javanica, Ostericum sieboldii*, *Pastinaca sativa*, *Pleurospermum camtschaticum*, and *Sium suave* samples were collected from the herb garden in Goyang, Kyunggi-do operated by Seoul National University, College of Pharmacy. *Peucedanum hakuunense* sample was collected from Baekwoon mountain in the Gwangyang region supplied by Seoul National University Forest. Four *Peucedanum japonicum* samples were collected from Geumodo Island (Pj-1), Hantaek Botanical garden (Pj-2), Jeju island (Pj-3), and Wando region (Pj-4) in South Korea; Sample from Geumodo and Hantaek is the cultivated sample while a sample from Jeju and Wando is a wild plant.

The genomic DNA for each sample was extracted using the modified cetyltrimethylammonium bromide (CTAB) method [24]. The quantity and quality of genomic DNAs were examined using both Nanodrop 1000 spectrometer (Thermo Fisher Scientific, USA) and Ethidium Bromide (EtBr) stained agarose gel electrophoresis. For Next-generation sequencing (NGS), genomic libraries with 300 bp

insert size were prepared according to the paired-end (PE) standard protocol (Illumina, USA) and sequenced using Illumina Miseq / HiseqX genome analyzer (Illumina, USA) at the LabGenomics (www.labgenomics.co.kr). The library for each plant sample was separately tagged with a different Illumina index and pooled for sequencing in a single lane. After sequencing 101 cycles, PE reads for each plant sample were collected according to the index.

Table 1-1. Summary of 31 Apioideae samples used in this study

No.	NGS ID	Scientific Name	Korean Name	Clade / Tribe	Source
1	IM180813-54	*Centella asiatica*	병풀	Mackinlayoideae	SNU Medical Herb Garden
2	IM160307-5	*Bupleurum falcatum*	시호	Burpleureae	SNU Medical Herb Garden
3	IM160307-6	*Bupleurum latissimum*	섬시호	Burpleureae	SNU Medical Herb Garden
4	IM180813-47	*Bupleurum longeradiatum*	개시호	Burpleureae	SNU Medical Herb Garden
5	IM180813-51	*Pleurospermum camtschaticum*	왜우산풀	Pleurospermeae	SNU Medical Herb Garden
6	IM180813-48	*Cryptotaenia japonica*	파드득나물	Oenantheae	SNU Medical Herb Garden
7	IM180813-50	*Oenanthe javanica*	미나리	Oenantheae	SNU Medical Herb Garden
8	IM180813-52	*Sium suave*	개발나물	Oenantheae	SNU Medical Herb Garden
9	IM151201-8	*Anthriscus sylvestris*	전호	Scandiceae	SNU Medical Herb Garden
10	IM180813-40	*Ostericum sieboldii*	묏미나리	Acronema Clade	SNU Medical Herb Garden
11	IM180813-44	*Coriandrum sativum*	고수	Coriandreae	SNU Medical Herb Garden
12	IM180813-27	*Cnidium officinale*	천궁	Sinodielsia Clade	SNU Medical Herb Garden
13	IM180813-29	*Conioselinum tenuissimum*	고본	Sinodielsia Clade	SNU Medical Herb Garden
14	IM180813-41	*Levisticum officinale*	유럽당귀	Sinodielsia Clade	SNU Medical Herb Garden
15	IM180813-45	*Heracleum moellendorffii*	어수리	Tordylieae	Lab. of Functional Plants
16	IM180813-46	*Pastinaca sativa*	설탕당근	Tordylieae	SNU Medical Herb Garden
17	IM180813-30	*Angelica amurensis*	지리강활	Selineae	SNU Medical Herb Garden
18	IM151201-3	*Angelica dahurica*	구릿대	Selineae	SNU Medical Herb Garden
19	IM180813-31	*Angelica decursiva*	바디나물	Selineae	SNU Medical Herb Garden
20	IM151201-7	*Angelica gigas*	참당귀	Selineae	SNU Medical Herb Garden
21	IM180813-42	*Angelica jaluana*	삼수구릿대	Selineae	SNU Medical Herb Garden
22	IM151201-2	*Angelica japonica A.Gray*	갯강활	Selineae	SNU Medical Herb Garden
23	IM180813-28	*Angelica reflexa*	강활	Selineae	SNU Medical Herb Garden
24	IM180813-43	*Cnidium japonicum*	갯사상자	Selineae	SNU Medical Herb Garden
25	IM180813-53	*Cnidium monnieri*	사상자	Selineae	SNU Medical Herb Garden

26	IM180813-49	*Dystaenia takesimana*	섬바디	Selineae	Lab. of Functional Plants
27	IM180813-21	*Peucedanum hakuunense*	백운기름나물	Selineae	SNU Forest
28	IM150601-18	*Peucedanum japonicum* (Pj-1)	갯기름나물	Selineae	Lab. of Functional Plants
29	IM200128-1	*Peucedanum japonicum* (Pj-2)	갯기름나물	Selineae	Hantaek Botanical
30	IM150921-6	*Peucedanum japonicum* (Pj-3)	갯기름나물	Selineae	Lab. of Functional Plants
31	IM150601-19	*Peucedanum japonicum* (Pj-4)	갯기름나물	Selineae	Lab. of Functional Plants

Plastome and nrDNA, assembly and annotation

Complete plastome and 45S nrDNA sequences were generated by de novo assembly of the low coverage whole genome sequence (dnaLCW) [25~26]. In brief, trimmed high-quality reads with Phred scores of 20 or more were obtained from the total PE raw reads using the CLC-quality trim tool and then were assembled by a CLC genome assembler (ver. 4.06 beta, CLC Inc, Rarhus, Denmark). The untapped initial contigs representing the plastid genome were selected from the total contigs using MUMmer [27] by selecting close topological related plastid genome sequences registered in the NCBI Genbank database as a reference sequence. The representative contigs were arranged based on reference plastome sequences and merged into a single draft sequence by connecting overlapping terminal sequences. The genes of the assembled plastome were annotated using the GeSeq (https://chlorobox.mpimp-golm.mpg.de/geseq.html) [28] and manually curated using Artemis through homologous genes comparison with other closely related species under Apioideae [29]. Circular maps of each Apioideae plastome were drawn using OGDRAW (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) [30].

For 45S nrDNA assembly, the longest initial contigs, including 45S cistron unit, were selected from the total contigs by comparison with reported 45S nrDNA sequences of *Panax ginseng* (KM036295), *Ledebourella seseloides* (KX757774, KX757775), *Glehnia littoralis* (KX757778, KX757779) and *P. japonicum* (KX757776, KX757777) from the previous study [25, 31]. Assembly errors and gaps within the draft sequences were manually corrected by mapping raw PE reads. The structures of 45S nrDNA sequences were predicted by both RNAmmer and BLASTN searches.

Structure analyses of plastome and nrDNA

The completed plastome and 45S nrDNA sequences on selected Apioideae species were aligned to seek variation on both structural variations along with informative polymorphic data using the web-based program MAFFT (http://mafft.cbrc.jp/alignment/ server/) [31]. Plastome structure variations were visualized through mVISTA (http://genome.lbl.gov/vista/mvista/submit.html) [33] under Shuffle LAGAN, and BlastZ tool (http://nature.snu.ac.kr/tools/ blastz_v3.php). Few misaligned regions were manually curated using the BioEdit program [34]. Afterward, polymorphic regions; Insertion/Deletion (InDels), and single nucleotide polymorphisms (SNPs) were investigated at an inter-specific level. To assess the variable position on the plastid gene, a nucleotide diversity (π) was calculated by sliding window analysis using DnaSP v6 [35].

To uncover dynamic plastome structure changes on Apioideae species, especially the junction changes related to the IR region, the LSC-IRs junction location on each Apioideae species was analyzed from the plastome annotation obtained from GeSeq analysis. Furthermore, to ensure the junction changed dramatically within Apioideae species, plastome information on 30 representative species of the Araliaceae family (*Aralia*, *Brassaiopsis*, *Cheirodendron*, *Dendropanax*, *Diplopanax*, *Eleutherococcus*, *Fatsia*, *Hydrocotyle*, *Kalopanax*, *Metapanax*, *Panax*, *Raukaua*, and *Schefflera*) was selected from the NCBI Genbank database for comparative analysis (Table S1-1).

Phylogenetic analyses

Phylogenetic analysis was conducted to understand the relationship between Apioideae species, including representative species of tribes and clades of Apioideae from NCBI GenBank (https://www.ncbi.nlm.nih.gov/nucleotide). As a result, the following 29 Apioideae species selected for analysis are listed in Table S1-2. To increase accuracy on phylogenetic analysis, only the protein-coding sequence (CDS) regions of the plastome was selected. The genes from the IRa region and controversial genes such as *ycf15* from IRb was excluded from the selected genes. 78 genes that are common in all species were selected for further analysis. Each CDS sequence was gathered into a single file and aligned using the MAFFT web-based program [32]. Before phylogenetic analysis, the nucleotide substitution model was selected by jModelTest v2.1.10 [36]. The following model was selected under the Akaike information criterion (AIC) set as GTR + G + I model. Finally, the maximum likelihood (ML) tree was constructed using RAxML Program [37] under with selected model and 1000 bootstrap replicate.

Divergence time estimation

To understand the evolutionary history of Apioideae species, the divergence time of the Apioideae species was estimated. MCMCTree method in the PAML 4.9h package [38] was used to estimate the divergence time. First, the BASEML in PAML was used to estimate the neutral substitution rate among Apioideae species, then MCMCTree was used to calculate the gradient vector (G) and Hessian matrix (H) with the time unit of 100Mya. Based on the substitution rate (0.004918) generated from the BASEML

analysis, prior for the parameter rgene and sigma2 in MCMCTree were set as G (1,250) and G (1, 4.5), respectively. Next, MCMC chain were analyzed, with every 10,000,000 generations with the sampling frequency of 50 and a burn-in phase of 1,000,000 generations. The stability of the run was verified using TRACER v. 1.72 [39] to ensure the effective sample size (ESS) greater of all parameters is over 200. The final tree was viewed in FigTree v1.4 [40]. As for the calibration point, two fossil data of tribe Bupleurere and Pleurospermeae from Banasiak et al. [41] and the strategies were adopted from Wen et al. [23]. Each calibration point was constrained to a lognormal distribution; the first calibration point was pointed at the stem node of Bupleureae (44.80 Mya) and the second was pointed at the stem node Pleurospermeae (43.52 Mya).

RESULTS

Genome organization and gene content of the plastomes of Apioideae species

The plastome sequences of 31 Apioideae samples were assembled with low-coverage whole-genome sequencing, otherwise known as the dnaLCW method [25, 26]. Information regarding plastome components, length of the quadripartite structure, GC contents, IR junctions are listed under Table 1-3. Among 31 species, *C. sativum* represented the smallest genome size (143,618 bp), whereas *P. japonicum* (Pj-2) represented the largest genome size (164,656 bp). The GC contents vary between 37.38 to 37.97%, where *S. suave* presented the smallest and *P. camtschaticum* presented the largest GC contents. Quadripartite region within Apioideae species ranges are varies as LSC: 75,853~99,223 bp, SSC: 16,945~18,108 bp, IRs: 13,474~35,760 bp in length. Comparative analysis between quadripartite regions, *C. sativum*; smallest plastome size represented largest LSC region (99,223 bp) and the smallest IR region (13,474 bp), whereas *P. japonicum* presented the smallest LSC region (75,583 bp) and largest IR region (35,759~35,760 bp), respectively. The number of genes in the plastome varied from 122 to 141. The differences in gene numbers are strongly affected by the IR region. Such changes regarding the length of the IR region are determined from the boundary between LSC and IR region refer as LSC/IR junction (otherwise known as JLB and JLA depending on the position of the IR region). 31 Apioideae species, the LSC/IR junction are grouped with two sets nearby the plastid gene of *rps19* gene or *ycf2* region.

Other junctions are located near the *petB* gene, *rpl22* gene, *rpl23* gene, *trn*V-GAC, and *trn*L-CAA region. *P. japonicum* demonstrated the largest IR region (35,760 bp) and the largest IR junction (*petB* intron) in the table. *P. japonicum* illustrated that two different plastome sizes exist between *P. japonicum* accessions with 147 kbp and 164 kbp.

Same dnaLCW method was used to assembled a 45S nrDNA sequence of the 31 Apioideae species. 45S nrDNA sequences were assembled into a single contig, including a 45S citron unit. Due to the gaps at the GC-rich regions along IGSs regions, it was impossible to assemble. The length of the 45S nrDNA units of each Apioideae species is ranged from the smallest 5,803 bp (*Angelica jaluana*) to 5,827 bp (*Centella asiatica*). The citron units are comprised of five parts, ranged as follows: 18S (1,809~1,816 bp), 5.8S (156~164 bp), and 26S(3,359~3,400 bp) ribosomal RNA unit and two internal transcribed spaces ITS1(209~253 bp) and ITS2 (224~242 bp) (Table 1-2). Between each compartment of the 45S citron unit, 18S and 5.8S presented minor variation in the unit length between the species, whereas ITS1, ITS2, and 26S represented variation changes caused by the InDel polymorphisms. Most of the SNP polymorphisms discovered under 45S nrDNA are located within ITS1 and ITS2 regions.

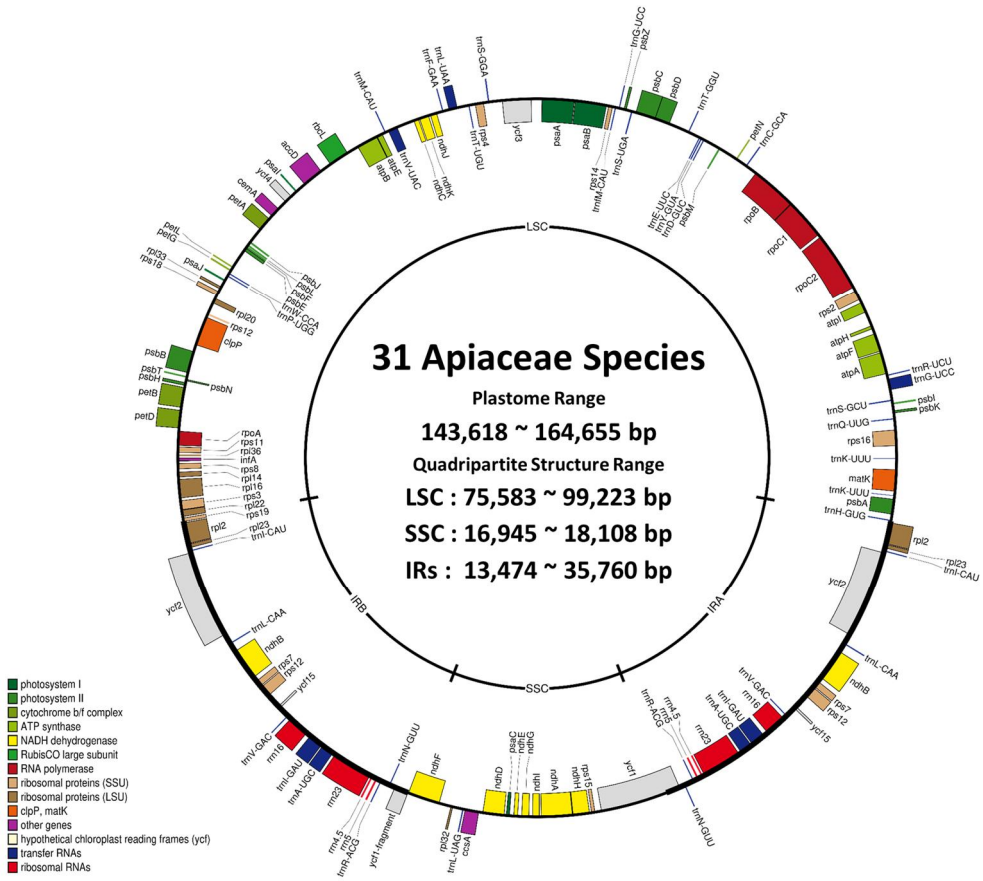


Figure 1-1. Plastid genome map of Apioideae species. The following figure demonstrates an assembled plastome map of 31 Apioideae species with plastome sizes ranging from 143 kbp to 164 kbp. Illustrated plastome map is based on *C. asiatica*, represented a close topological relationship to the Araliaceae family (Figure 1-3). Plastome is composed of a quadripartite structure of two SC regions (Large and Small) and IR regions. Information regarding the length of each LSC, SSC, and IR region is presented in Table 1-3. Each plastome map was generated using OGDRAW. Genes transcribed clockwise and counterclockwise are indicated on the outside and inside of the large circle, respectively.

Table 1-2. Information of 45S nrDNA sequence of assembled Apioideae species.

No.	NGS ID	Scientific Name	Length (bp)	18s	ITS1	5.8s	ITS2	26s
1	IM180813-54	*Centella asiatica*	5,827	1,810	253	158	242	3,364
2	IM160307-5	*Bupleurum falcatum*	5,813	1,809	213	164	227	3,400
3	IM160307-6	*Bupleurum latissimum*	5,815	1,809	215	164	237	3,390
4	IM180813-47	*Bupleurum longeradiatum*	5,814	1,809	242	160	239	3,364
5	IM180813-51	*Pleurospermum camtschaticum*	5,807	1,809	243	162	230	3,363
6	IM180813-48	*Cryptotaenia japonica*	5,809	1,809	209	164	238	3,389
7	IM180813-50	*Oenanthe javanica*	5,808	1,809	210	164	236	3,389
8	IM180813-52	*Sium suave*	5,809	1,809	211	164	236	3,389
9	IM151201-8	*Anthriscus sylvestris*	5,812	1,809	216	164	224	3,399
10	IM180813-40	*Ostericum sieboldii*	5,812	1,809	245	159	236	3,363
11	IM180813-44	*Coriandrum sativum*	5,805	1,809	242	158	234	3,362
12	IM180813-27	*Angelica genuflexa*	5,806	1,809	214	163	231	3,389
13	IM180813-29	*Conioselinum tenuissimum*	5,806	1,809	217	163	233	3,384
14	IM180813-41	*Levisticum officinale*	5,806	1,809	245	156	235	3,361
15	IM180813-45	*Heracleum moellendorffii*	5,816	1,816	244	159	235	3,362
16	IM180813-46	*Pastinaca sativa*	5,812	1,809	247	159	235	3,362
17	IM180813-30	*Angelica amurensis*	5,806	1,809	215	163	231	3,388
18	IM151201-3	*Angelica dahurica*	5,806	1,809	215	163	231	3,388
19	IM180813-31	*Angelica decursiva*	5,806	1,810	214	163	231	3,388
20	IM151201-7	*Angelica gigas*	5,806	1,809	215	163	231	3,388
21	IM180813-42	*Angelica jaluana*	5,803	1,809	241	161	233	3,359
22	IM151201-2	*Angelica japonica A.Gray*	5,807	1,809	215	163	232	3,388
23	IM180813-28	*Angelica reflexa*	5,805	1,809	215	162	231	3,388
24	IM180813-43	*Cnidium japonicum*	5,807	1,809	242	160	234	3,362
25	IM180813-53	*Cnidium monnieri*	5,807	1,809	242	160	234	3,362
26	IM180813-49	*Dystaenia takesimana*	5,804	1,809	241	160	233	3,361
27	IM180813-21	*Peucedanum hakuunense*	5,807	1,809	215	163	232	3,388
28	IM150601-18	*Peucedanum japonicum* (Pj-1)	5,809	1,809	216	163	233	3,388
29	IM200128-1	*Peucedanum japonicum* (Pj-2)	5,809	1,809	216	163	233	3,388
30	IM150921-6	*Peucedanum japonicum* (Pj-3)	5,809	1,809	216	163	233	3,388
31	IM150601-19	*Peucedanum japonicum* (Pj-4)	5,809	1,809	216	163	233	3,388

Plastome diversity of Apioideae species

Nucleotide diversity (pi) of the plastome of 60 Apioideae species, including outgroup, was calculated to observe which gene was affected during divergence (Table 1-3). The pi value range ranges from 0 to 0.086, the *trn*E-UUC-*trn*Y-GUA-*trn*D-GUC gene revealed the highest nucleotide diversity. However, this result may have occurred due to the tRNA genes inverted rotated within three species. *Angelica gigas* and *Peucedanum japonicum* from the Selineae tribe and *Carum carvi* from the Careae tribe demonstrated unique tRNA gene inversion, whereas other species have demonstrated the tRNA order of *trn*D-GUC, *trn*Y-GUA, and *trn*E-UUC. More information will be discussed in the later section. Excluding *trn*E-*trn*Y-*trn*D genes pi value, *psbI-trnS*, *trnT-trnL*, *rpl16-rps19* within LSC region and *ycf1a-ndhF*, *ndhE-ndhG*, *ndhA*, and *ycf1* gene within SSC region represented with high nucleotide divergence. The divergence hotspot coverage from a quadripartite area is mostly detected within LSC and SSC. IR region involves less inter or intraspecific divergence. It is interesting to note that pi value near IR junction *petB-petD*, *rpl16-rps19* and *ycf1a-ndhF* genes reveal relatively high pi value than other closely related gene regions.

Table 1-3. Information of plastome sequence of 60 Apioideae species.

No	Clade / Tribe	Scientific Name	IR junction Position	Length (bp)	LSC (bp)	SSC (bp)	IRs (bp)	G+C %	Gene	CDS	tRNA	rRNA	NGS ID
1	Araliaceae	*Panax ginseng cv Chunpoong*	*rps19*	156,248	86,128	18,084	26,018	38.07	130	85	37	8	KM088019
2	Araliaceae	*Aralia elata*	*rps19*	156,220	86,263	18,111	25,923	38.08	132	87	37	8	KT153023
3	Araliaceae	*Hydrocotyle verticillata*	*rps19*	153,207	84,352	18,739	25,058	37.59	132	87	37	8	NC_015818
4	Mackinlayoideae	*Centella asiatica*	*rps19*	154,760	86,166	18,108	25,243	37.68	130	85	37	8	IM180813-54
5	Chamaesium clade	*Chamaesium delavayi*	*rps19*	154,684	85,029	17,415	26,120	38.34	130	85	37	8	MN119367
6	Burpleureae	*Bupleurum latissimum*	*rps19*	155,568	85,417	17,547	26,302	37.65	130	85	37	8	IM160307-6
7	Burpleureae	*Bupleurum falcatum*	*rps19*	155,817	85,736	17,499	26,291	37.66	130	85	37	8	IM160307-5
8	Burpleureae	*Bupleurum longeradiatum*	*rps19*	156,036	85,864	17,588	26,292	37.66	130	85	37	8	IM180813-47
9	Pleurospermeae	*Pleurospermum camtschaticum*	*rps19*	155,413	85,942	17,493	25,989	37.97	130	85	37	8	IM180813-51
10	Physospermopsis clade	*Hansenia weberbaueriana*	*rps19*	158,625	88,260	18,237	26,064	37.68	130	85	37	8	NC_035053
11	Komarovia clade	*Chuanminshen violaceum*	*rps19*	154,530	84172	17,800	26,279	37.79	130	85	37	8	KU921430.2
12	Oenantheae	*Tiedemannia filiformis*	*rps19*	154,737	84,585	17,140	26,506	37.32	130	85	37	8	HM596071.1
13	Oenantheae	*Sium suave*	*rps19*	155,018	84,887	17,707	26,212	37.38	130	85	37	8	IM180813-52
14	Oenantheae	*Cryptotaenia japonica*	*rps19*	153,398	83,613	16,945	26,420	37.51	130	85	37	8	IM180813-48
15	Oenantheae	*Oenanthe javanica*	*rps19*	152,113	83,921	17,280	25,456	37.64	130	85	37	8	IM180813-50
16	Oenantheae	*Cicuta virosa*	*rps19*	154,569	84,177	17,578	26,407	37.52	130	85	37	8	NC_037711
17	Arcuatopterus Clade	*Sillaphyton podagraria*	*rps19*	156,912	84,492	17,417	27,508	37.68	130	85	37	8	NC_033344
18	Scandiceae	*Anthriscus sylvestris*	*rps19*	154,951	84,818	17,313	26,410	37.5	130	85	37	8	IM151201-8
19	Acronema Clade	*Ligusticum delavayi*	*rps19*	155,623	85,066	16,741	26,908	37.55	130	85	37	8	NC_049052
20	Acronema Clade	*Ostericum sieboldii*	*rps19*	158,186	88,256	17,344	26,293	37.63	130	85	37	8	IM180813-40
21	Acronema Clade	*Pterygopleurum neurophyllum*	*rps19*	154,369	84,411	21,210	24,374	37.61	130	85	37	8	NC_033345
22	Torilidinae	*Caucalis platycarpos*	*rps19*	171,083	85,042	17,553	34,244	37.84	130	85	37	8	KX832334.1
23	Daucinae	*Cuminum cyminum*	*rps19*	157,839	83,927	17,598	28,157	37.79	130	85	37	8	NC_046879
24	Daucinae	*Daucus carota*	*rps19*	155,911	84,242	17,571	27,049	37.66	130	85	37	8	NC_008325
25	Pyramidoptereae	*Crithmum maritimum*	*rpl16*	158,355	85,230	17,139	27,993	37.55	133	88	37	8	NC_015804
26	Careae	*Carum carvi*	*rps3*	155,449	83,672	17,549	27,114	37.77	125	80	37	8	NC_029889
27	Apieae	*Petroselinum crispum*	*rpl2*	152,890	86,116	17,506	24,634	37.78	129	84	37	8	NC_015821
28	Apieae	*Anethum graveolens*	*rpl2*	153,357	86,506	17,517	24,667	37.65	129	84	37	8	NC_029470

29	Apieae	*Foeniculum vulgare*	*rpl2*	153,628	86,659	17,470	24,750	37.65	129	84	37	8	NC_029469
30	Cachrys clade	*Prangos trifida*	*rpl2*	153,510	86,482	17,446	24,791	37.75	129	84	37	8	NC_037852
31	Coriandreae	*Coriandrum sativum*	*trn*V-GAC	143,618	99,223	17,447	13,474	37.44	122	79	35	8	IM180813-44
32	Sinodielsia Clade	*Levisticum officinale*	*ycf2*	148,547	93,908	17,703	18,468	37.58	126	82	36	8	IM180813-41
33	Sinodielsia Clade	*Cnidium officinale*	*ycf2*	148,520	93,977	17,607	18,468	37.6	126	82	36	8	IM180813-27
34	Sinodielsia Clade	*Conioselinum tenuissimum*	*rpl22*	158,510	84,874	17,662	27,987	37.61	131	86	37	8	IM180813-29
35	Tordylieae	*Heracleum moellendorffii*	*trn*L-CAA	149,340	92,249	17,507	19,792	37.5	127	82	37	8	IM180813-45
36	Tordylieae	*Pastinaca sativa*	*trn*L-CAA	149,619	92,385	17,620	19,807	37.42	127	82	37	8	IM180813-46
37	Selineae	*Seseli montanum*	*ycf2*	147,823	92,621	17,480	18,861	37.57	126	82	36	8	NC_027451
38	Selineae	*Ledebouriella seseloides*	*ycf2*	147,830	93,202	17,324	18,652	37.53	126	82	36	8	NC_034643
39	Selineae	*Dystaenia takesimana*	*ycf2*	147,725	93,030	17,557	18,569	37.55	126	82	36	8	IM180813-49
40	Selineae	*Peucedanum praeruptorum*	*ycf2*	147,197	92,161	17,610	18,713	37.58	125	82	35	8	MN016968
41	Selineae	*Peucedanum japonicum* (Pj-3)	*ycf2*	147,525	92,563	17,628	18,667	37.54	126	82	36	8	IM150921-6
42	Selineae	*Peucedanum japonicum* (Pj-4)	*ycf2*	147,550	92,632	17,606	18,656	37.54	126	82	36	8	IM150601-19
43	Selineae	*Peucedanum japonicum* (Pj-1)	*petB*	164,652	75,583	17,551	35,759	37.45	141	96	37	8	IM150601-18
44	Selineae	*Peucedanum japonicum* (Pj-2)	*petB*	164,656	75,585	17,551	35,760	37.46	141	96	37	8	IM200128-1
45	Selineae	*Arracacia xanthorrhiza*	*ndhB*	143,989	94,820	17,439	15,865	37.48	124	81	35	8	NC_032364
46	Selineae	*Cnidium monnieri*	*ycf2*	147,227	94,222	17,545	17,730	37.47	126	82	36	8	IM180813-53
47	Selineae	*Cnidium japonicum*	*ycf2*	147,038	94,017	17,571	17,725	37.43	126	82	36	8	IM180813-43
48	Selineae	*Angelica acutiloba*	*ycf2*	147,074	93,367	17,573	18,067	37.53	126	82	36	8	NC_029391
49	Selineae	*Angelica japonica A.Gray*	*rpl23*	154,302	86,603	17,547	25,076	37.55	127	83	36	8	IM151201-2
50	Selineae	*Peucedanum terebinthaceum*	*ycf2*	147,925	93,368	17,571	18,493	37.48	126	82	36	8	NC_053641
51	Selineae	*Peucedanum hakuunense*	*ycf2*	147,426	91,915	17,425	19,043	37.52	126	82	36	8	IM180813-21
52	Selineae	*Glehnia littoralis*	*ycf2*	147,477	93,496	17,555	18,213	37.51	126	82	36	8	NC_034645
53	Selineae	*Angelica dahurica*	*ycf2*	146,811	93,546	17,629	17,818	37.52	126	82	36	8	IM151201-3
54	Selineae	*Angelica nitida*	*ycf2*	146,512	93,298	17,950	17,632	37.48	126	82	36	8	MF594405.1
55	Selineae	*Angelica jaluana*	*ycf2*	145,959	92,545	17,230	18,092	37.57	126	82	36	8	IM180813-42
56	Selineae	*Angelica decursiva*	*rpl22*	154,736	84,138	17,594	26,502	37.61	131	86	37	8	IM180813-31
57	Selineae	*Angelica decursiva*	*ycf2*	146,719	93,256	17,497	17,983	37.56	127	82	37	8	KT781591.1
58	Selineae	*Angelica gigas*	*ycf2*	146,933	93,152	17,605	18,088	37.55	126	82	36	8	IM151201-7
59	Selineae	*Angelica amurensis*	*ycf2*	146,983	93,186	17,579	18,109	37.52	126	82	36	8	IM180813-30
60	Selineae	*Angelica reflexa*	*ycf2*	147,281	93,187	17,662	18,216	37.54	126	82	36	8	IM180813-28

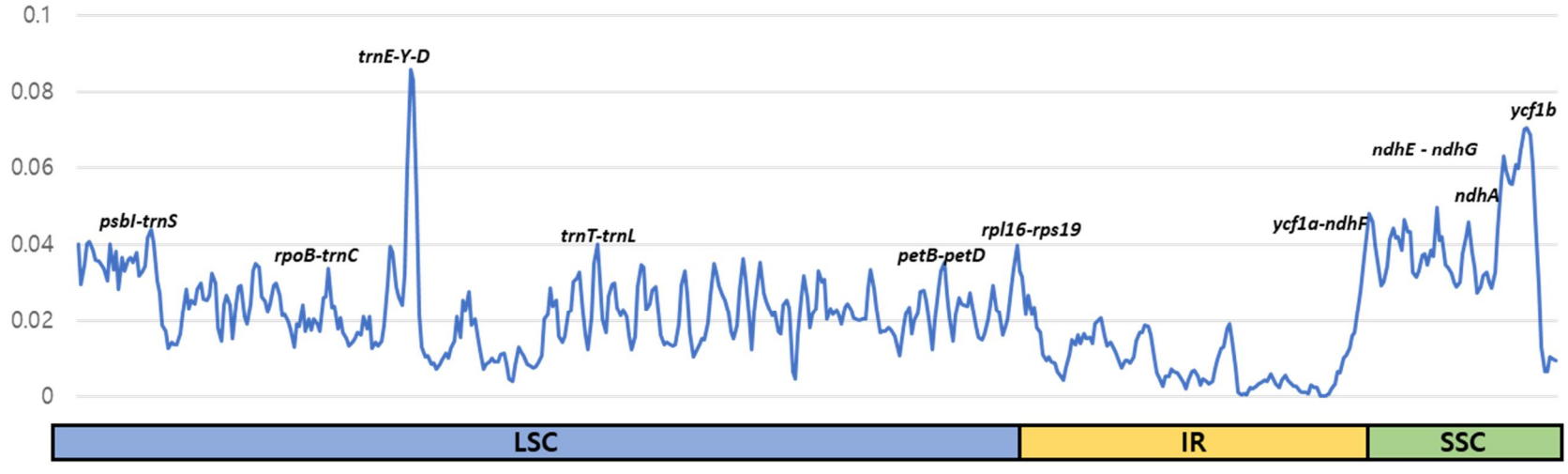


Figure 1-2. Nucleotide diversity (pi) of plastome of 60 Apioideae species.

Plastome-based phylogenetic analysis of the Apioideae

A total of 60 Apioideae species plastome sequences were used for phylogenetic analysis (Figure 1-3). *Panax ginseng, Aralia elata, Hydrocotyle verticillate* in the Araliaceae family were included as an outgroup. To ensure the accuracy of the phylogenetic relationship on subfamily Apioideae, 78 plastid CDS regions that coincide with all species were selected. The phylogenetic relationship of subfamily Apioideae in this study were almost consistent with the phylogenetic relationships in other previous studies [2, 23] except the position of *Prango trifida* in the *Cachrys* clade, with slightly low branch supporting bootstrap value (BS = 60) (Figure 1-3A). The phylogenetic analysis indicated that Burpleureae, Oenantheae, *Acronema* clade, Apieae, *Sinodielsia* clade, Tordylieae, and Selineae were monophyletic relationship. The branches of most clade and tribes in the Apioideae illustrated a strong bootstrap supporting value (BS = 100). Few branches along superclade illustrated depressed bootstrap supporting values: in Burpleureae and Pleurospermeaea-Selineae clade (BS = 82), *Cachrys* clade and Coriandreae-Selineae (BS = 60), Tordylieae and Selineae (BS = 74). The Cachrys clade showed the lowest supporting value, because the position of the Cachrys clade may illustrated sister clade to the Apieae as presented in other previous studies [2, 23]. Most species resolved relationships with high internal support values (BS = 100).

In this study, two endemic and rare Apioideae species were included; *Dystaenia takeshimana* and *Peucedanum hakuunense* belong to the Selineae tribe, illustrating a sister clade with *Peucedanum praeruptorum* and *Peucedanum terebinthaceum,* respectively. The Selineae tribe was separated into two groups. The first group comprises *Seseli montanum*, *Ledebourella seseloides*, *Dystaenia*

takeshimana, *Peucedanum praeruptorum*, and *Peucedanum japonicum*. The second group comprises mostly species under the *Angelica* genus with *Arracacia xanthorrhiza*, *Cnidium monnieri*, *Cnidium japonicum*, *Ghlenia littoralis*, *Peucedanum hakuunense*, and *Peucedanum terebinthaceum*. The phylogenetic result has illustrated polyphyletic relationship under three genera, *Angelica*, *Cnidium,* and *Peucedanum*. Between *Angelica* species, *Angelica dahurica* and *Angelica nitida* formed a sister clade with *Glehnia littoralis,* whereas *Angelica japonica* formed a sister clade with two *Peucedanum* species, *Peucedanum hakunense* and *Peucedanum terebinthaceum*. The remaining five *Angelica* species such as *A. jaluana*, *A decursiva*, *A. gigas*, *A. amurensis*, and *A. reflexa*, demonstrated monophyletic relationship within the Selineae tribe. Between *Cnidium* species, *Cnidium officinale* was sister to *Conioselinum tenuissimum* within the Sinodielsia clade, while other two *Cnidium* species (*C. monnieri* and *C. japonicum*) were sister to the clade of *Angelica-Peucedanum* within the Selineae tribe. While two *Peucedanum* species (*P. hakuunense* and *P. terebinthaceum*) were nested within the *Angelica* clade, *P. japonicum* and *P. praeruptorum* formed a clade, and then the clade was sister to *D. takeshimana*.

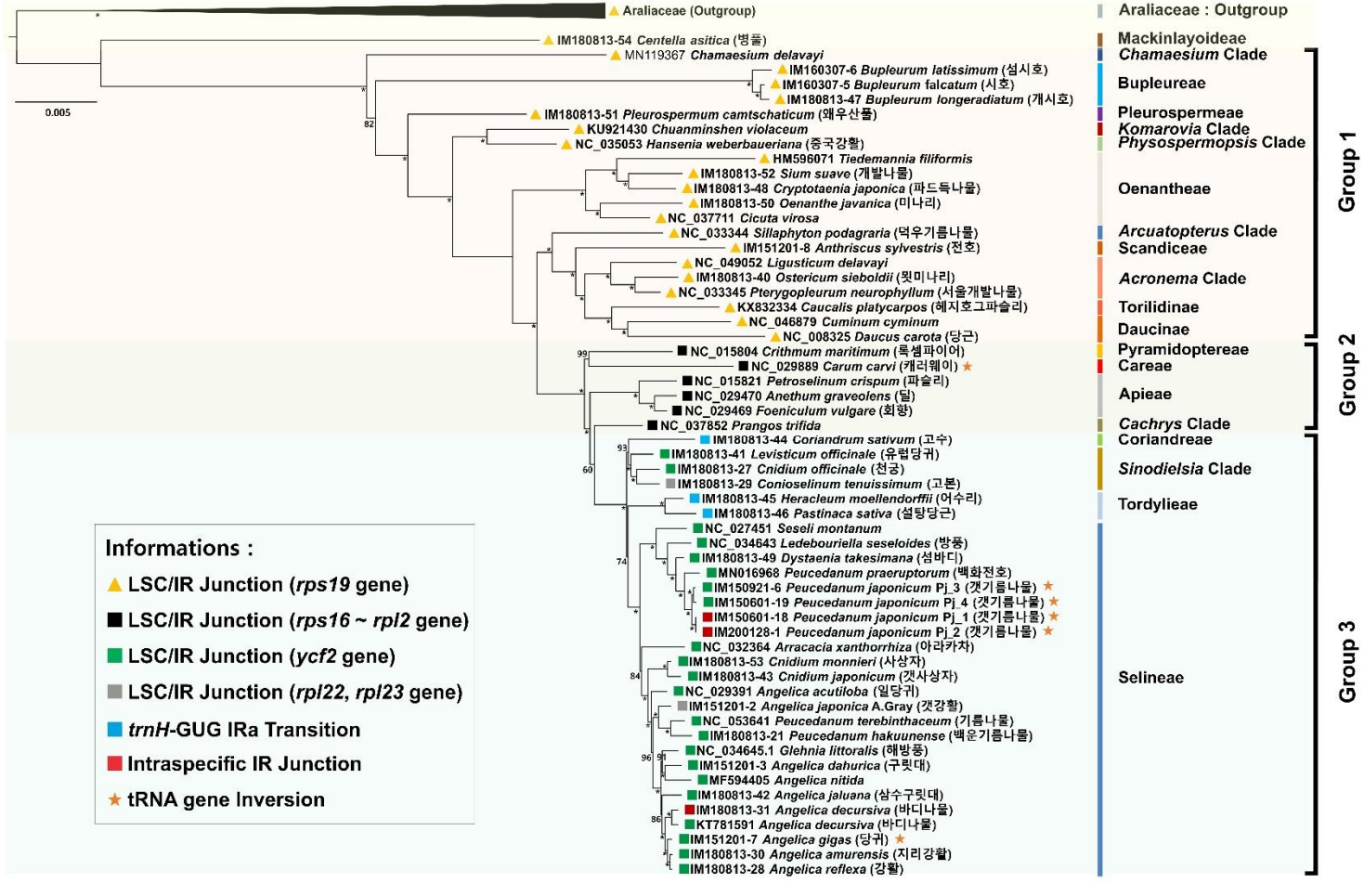

Araliaceae (Outgroup)
IM180813-54 Centella asitica (병풀)
MN119367 Chamaesium delavayi
IM160307-6 Bupleurum latissimum (섬시호)
IM160307-5 Bupleurum falcatum (시호)
IM180813-47 Bupleurum longeradiatum (개시호)
IM180813-51 Pleurospermum camtschaticum (왜우산풀)
KU921430 Chuanminshen violaceum
NC_035053 Hansenia weberbaueriana (중국강활)
HM596071 Tiedemannia filiformis
IM180813-52 Sium suave (개발나물)
IM180813-48 Cryptotaenia japonica (파드득나물)
IM180813-50 Oenanthe javanica (미나리)
NC_037711 Cicuta virosa
NC_033344 Sillaphyton podagraria (덕우기름나물)
IM151201-8 Anthriscus sylvestris (전호)
NC_049052 Ligusticum delavayi
IM180813-40 Ostericum sieboldii (뭣미나리)
NC_033345 Pterygopleurum neurophyllum (서울개발나물)
KX832334 Caucalis platycarpos (헤지호그파슬리)
NC_046879 Cuminum cyminum
NC_008325 Daucus carota (당근)
NC_015804 Crithmum maritimum (록셈파이어)
NC_029889 Carum carvi (캐러웨이)
NC_015821 Petroselinum crispum (파슬리)
NC_029470 Anethum graveolens (딜)
NC_029469 Foeniculum vulgare (회향)
NC_037852 Prangos trifida
IM180813-44 Coriandrum sativum (고수)
IM180813-41 Levisticum officinale (유럽당귀)
IM180813-27 Cnidium officinale (천궁)
IM180813-29 Conioselinum tenuissimum (고본)
IM180813-45 Heracleum moellendorffii (어수리)
IM180813-46 Pastinaca sativa (설탕당근)
NC_027451 Seseli montanum
NC_034643 Ledebouriella seseloides (방풍)
IM180813-49 Dystaenia takesimana (섬바디)
MN016968 Peucedanum praeruptorum (백화전호)
IM150921-6 Peucedanum japonicum Pj_3 (갯기름나물)
IM150601-19 Peucedanum japonicum Pj_4 (갯기름나물)
IM150601-18 Peucedanum japonicum Pj_1 (갯기름나물)
IM200128-1 Peucedanum japonicum Pj_2 (갯기름나물)
NC_032364 Arracacia xanthorrhiza (아라카차)
IM180813-53 Cnidium monnieri (사상자)
IM180813-43 Cnidium japonicum (갯사상자)
NC_029391 Angelica acutiloba (일당귀)
IM151201-2 Angelica japonica A.Gray (갯강활)
NC_053641 Peucedanum terebinthaceum (기름나물)
IM180813-21 Peucedanum hakuunense (백운기름나물)
NC_034645.1 Glehnia littoralis (해방풍)
IM151201-3 Angelica dahurica (구릿대)
MF594405 Angelica nitida
IM180813-42 Angelica jaluana (삼수구릿대)
IM180813-31 Angelica decursiva (바디나물)
KT781591 Angelica decursiva (바디나물)
IM151201-7 Angelica gigas (당귀)
IM180813-30 Angelica amurensis (지리강활)
IM180813-28 Angelica reflexa (강활)
0.005
82
99
60
93
74
84
96
91
86
Informations :
LSC/IR Junction (rps19 gene)
LSC/IR Junction (rps16 ~ rpl2 gene)
LSC/IR Junction (ycf2 gene)
LSC/IR Junction (rpl22, rpl23 gene)
trnH-GUG IRa Transition
Intraspecific IR Junction
tRNA gene Inversion
Araliaceae : Outgroup
Mackinlayoideae
Chamaesium Clade
Bupleureae
Pleurospermeae
Komarovia Clade
Physospermopsis Clade
Oenantheae
Arcuatopterus Clade
Scandiceae
Acronema Clade
Torilidinae
Daucinae
Pyramidoptereae
Careae
Apieae
Cachrys Clade
Coriandreae
Sinodielsia Clade
Tordylieae
Selineae
Group 1
Group 2
Group 3

Figure 1-3. Plastome-based phylogenetic analysis on 60 representatives of Apioideae species, including Araliaceae family. Phylogenetic analysis on Apioideae species, including three Araliaceae species (*A. elata*, *P. ginseng*, *H. verticillata*) based on the common protein-coding sequence. The phylogenetic tree was constructed using maximum likelihood (ML), species under major clade, and the tribe of Apioideae species is labeled with the LSC/IR junction position labeled with colored triangle and square shape. Support values marked in the figure represent bootstrap support. * represent the best support (100%).

Table 1-4. The number of each LSC/IR junction types discovered within Apioideae plastomes in Figure 1-3.

Sequence Source	Group 1 *rps19* gene junction	Group 2 *rps*16 ~ *rpl*2 junction	Group 3 *ycf2* junction	Group 3 *rpl*22, *rpl*23 junction	Group 3 *trn*H-GUG IRa transition	Group 3 Intraspecific IR junction	Total No. Species
This study	10	0	13	2	3	3	31
GenBank	14	6	9	0	0	0	29
Total	24	6	22	2	3	3	60

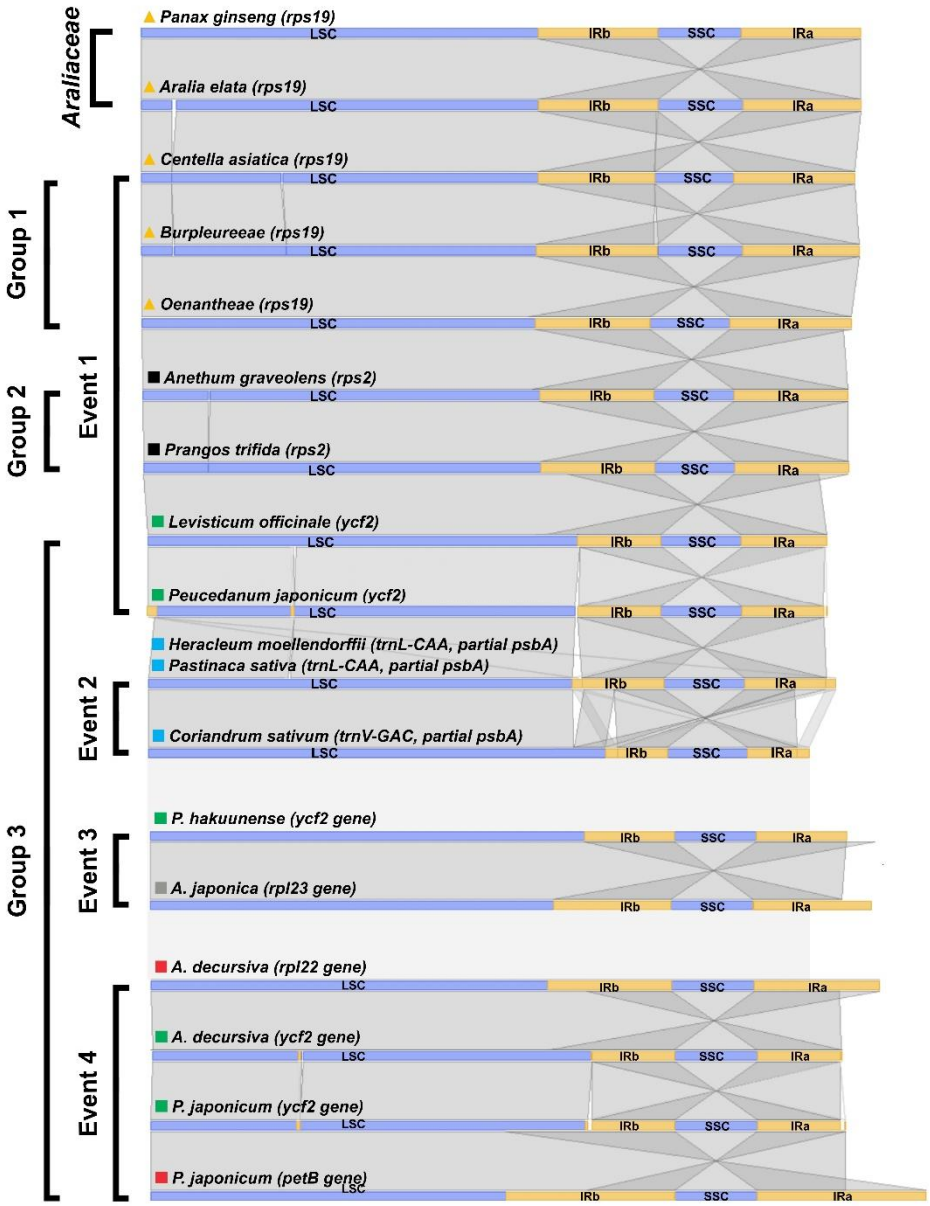


Figure 1-4. Illustration of the blastZ analysis on Apioideae species representative of groups related to the changes in LSC/IR junction from Figure 1-3, illustrating the changes in plastome structure of Apioideae. The species are also groups as an event of IR junction changes referenced from the Figure 1-5.

Major events regarding reposition of IR junction in Apioideae

According to the phylogenetic relationship, the tribes and clades of the Apioideae were categorized into three groups based on the positions of the IR junctions. Species in Group 1 composed of both Araliaceae and basal Apioideae species represented the IR junction at the *rps19* region (Figure 1-3, Table 1-3). Species in Group 2 represented the IR junction is at the *rpl2* region. Species in Group 3 represented the IR junction at the *ycf2* region (Figure 1-3 Table 1-3). In addition, the plastome comparison of Araliaceae and the three groups illustrated a gradual contraction of the IR length (Figure 1-4, Table 1-3). On the other hand, the species that represented with the *trn*H-GUG transition and intraspecific IR junction under Group 3 have illustrated the expansion of the IR length on Apioideae species (Figure 1-4). Plastome comparison revealed that a portion of the IR region around 1.2kbp size on *Coriandrum sativum, Heracleum moellendorffii* and *Pastinaca sativa* overlaps the *trn*H-GUG gene on LSC region of *Peucedanum japonicum* (Figure 1-4; *trn*H-GUG transition). The results represent the *trn*H-GUG, and the partial *psbA* gene has been inserted into the IR region. For *C. sativum,* the inserted sequence is found near 3' end of the *trn*V-GAC region, whereas *for Heracleum moellendorffii* and *Pastinaca sativa*, it is found near 3' end of the *trn*L-CAA region. Because Figure 1-3 showed the IR junction in most Group 3 species is in the *ycf2* region, the position of the inserted gene region suggests that the ancestor of both species experienced IR contraction before *trn*H-GUG and partial *psbA* sequence inserted into the IR region. BlastZ analysis on four species revealed an extension of the IR region in *Conioselinum tenuissimum*, *Angelica japonica*, *Angelica decursiva*, and *Peucedanum japonicum*. The four mentioned species formed a sister clade with the species that had

an IR junction in the *ycf2* region (Figure 1-3). *Conioselinum tenuissimum* and *Angelica japonica* illustrated IR expansion limited to those two species, but *Angelica decursiva* and *Peucedanum japonicum* demonstrated IR expansion occurred within the species (Figure 1-4; Intraspecific IR). Observing IR junction changes on both phylogenetic analysis and BlastZ analysis, a scenario of a significant event regarding the reposition of IR junction under Apioideae species was hypothesized.

The first event is the repositioning of the LSC-IR junction, making the two dependent IR reduction at the right border of LSC when Apioideae species diverged from Araliaceae (Figure 5A). Araliaceae is a sister clade of Apiaceae, the plastome sequence of representative species under the Araliaceae family (*Aralia*, *Brassaiopsis*, *Cheirodendron*, *Dendropanax*, *Diplopanax*, *Eleutherococcus*, *Fatsia*, *Hydrocotyle*, *Kalopanax*, *Metapanax*, *Panax*, *Raukaua,* and *Schefflera*) demonstrated IR region differences around 1 kbp in length (25,058~26,136 bp) with IR junction located near 5' end of the *rps19* gene (Table S1-1). Compared to the plastome sequence from the Araliaceae family, Apioideae species under Group 1 also represented IR junction located near 5' end of the *rps19 gene* with a variation within IR length ranging between 25,456 bp(*Oenanthe javanica*) ~ 28,157 bp(*Cuminum cyminum*) (Table 1-3). Two species under Group 1, revealed a change in IR junction or IR length. *Pterygopleurum neurophyllum* represented IR junction near *rpl2* due to pseudogenization on *rps19* gene but maintained the IR length around 25 kbp. Whereas, *Caucalis platycarpos* represented IR junction near *rps19* gene with the largest IR length (34,244 bp) among the Group 1 Apioideae species due to the 10 kbp insertion within IR region identified as mitochondria-to-plastid transfer and nuclear-to-plastid transfer [22]. Though two

species (*P. neurophyllum, C. platycarpos*) presented a variation in the IR length, the IR junction under most Group 1 species seems to maintain a stable state at the *rps19* gene IR region around 25 ~ 28 kbp in length (Figure 1-5A, Table 1-3). As the Apioideae species continue to diverge, species under Group 2 (*C. carvi*, *C. maritimum*, *F. vulgare*, *A. graveolens*, *P. crispum*, and *P. trifida*), seems to lose its stability of the IR junction and begins to reposition the IR junction near the *rps19* gene (*rpl16-rpl2*) (Figure 1-5A). IR junction changes dramatically as Group 3 starts to separate from Group 2. Most species' within Sinodielsia clade and Selineae tribe, the IR junction was repositioned near the 3' end of the *ycf2* gene (Figure 1-5A).

The second event is the reposition of IRa-LSC junction to the left LSC border, exhibited within *C. sativum*, *H. moellendorffii,* and *P. sativa*, which involves the transition of *trn*H-GUG and *psbA* gene into the IR region (Figure 1-5B, Table 1-3). Location of the IR junction related with the transition of *trn*H-GUG and partial *psbA* is differed by the species, that, either positioned near 3' end of the *trn*L-CAA (*H. moellendorffii*, *P. sativa*) or near 3' end of the *trn*V-GAC gene (*C. sativum*) (Figure 1-5B). Gene order of both species has illustrated junction within IRa include *trn*H-GUG partial *psbA* gene. The plastome sequence beginning with partial *psbA* followed by *matK* as a start position on the LSC region, whereas the *trn*H-GUG gene and remainder of partial *psbA* transition into the IRa region. Such transition is also illustrated in the blastZ analysis (Figure 1-4; *trn*H-gug transition). *C. sativum,* unlike the other Group 3 species, experienced more significant IR contraction up to the *trn*V-GAC gene in the IRb region. Then simultaneously experienced the IR expansion on the IRa region where *trnH-GUG* and partial *psbA* transition into IRa and the IR junction has again

repositioned near the partial *psbA* gene. As the IRa region experienced expansion, IRb also expanded, including *trn*H-GUG and partial *psbA* insertion due to copy-correct mechanism during genome rearrangement process. A similar process occurred on both *H. moellendorffii* and *P. sativa* that IR contracted up to *trn*L-CAA region before *trn*H-GUG and partial *psbA* transition into IRa.

The third event related to the reposition of LSC-IR junction related to IR expansion occured specific to *Angelica japonica* and *Conioselinum tenuissimum* species under Group 3 (Figure 1-5C, Table 1-3). As mentioned, the IR junction on most Group 3 species is located near the 3' end of *ycf2* region. Two species; *A. japonica* and *C. tenuissimum* species are associated under Group 3, illustrates a monophyletic relationship with three species; *C. officinale*, *P. hakuunense,* and *P. terebinthaceum*, respectively (Figure 1-3, Figure 1-5C). However, both *A. japonica* and *C. tenuissimum* have revealed IR junction located near 3' end of *rpl22* and *rpl23* region, which represents an IR expansion specifically occurred under *A. japonica* and *C. tenuissimum*.

The fourth event is the appearance of an intraspecific variation regarding the IR region that demonstrated two different plastome sizes with different LSC/IR junctions exhibited under *A. decursiva* and *P. japonicum* species (red square, Figure 1-3, Figure 1-5D, Figure 1-6). Phylogenetic analysis revealed that each *A. decursiva* and *P. japonicum* species exhibited a monophyletic relationship. (Figure 1-3). *A. decursiva* demonstrated unique intraspecific IR variation by introducing a new plastome size of 154,736 bp with LSC/IR junction near the *rpl22* gene region. On the other hand, the plastome of *A. decursiva* from the NCBI Genbank database exhibited a size of 146,719

bp with LSC/IR junction located in between *ycf2*/*trn*L-CAA (Figure 1-5D, Figure 1-6). Regarding *P. japonicum*, two *P. japonicum* plastomes (NC_034644, KU866531) with a genomic size of 164,653 bp without intraspecific variation from our previous study has been uploaded in the NCBI Genbank database [31]. Newly assembled *P. japonicum* samples for this study (Pj-1, Pj-2, Pj-3, Pj-4) have exhibited the genomic size of 164,652 bp, 164,656 bp, 147,525 bp, and 147,549 bp, respectively (Table 1-3). The assembled plastome of *P. japonicum* accessions (Pj-1, Pj-2) has demonstrated similarities with the previous study; 164 kbp with LSC/IR junction located in *petB* intron. However, *P. japonicum* accessions (Pj-3, Pj-4) were also introduced to another *P. japonicum* plastome of 147 kbp in size with LSC/IR junction located in between *ycf2*/*trn*L-CAA (Figure 1-5D, Figure 1-6). The changes in the IR region revealed that *A. decursiva* exhibited 8 kbp differences as *P. japonicum* exhibited 17 kbp differences. For convenience, species with intraspecific IR variation with a long IR region will be referred to as an L-type, and intraspecific IR variation with a short IR region will be referred to as an S-type.

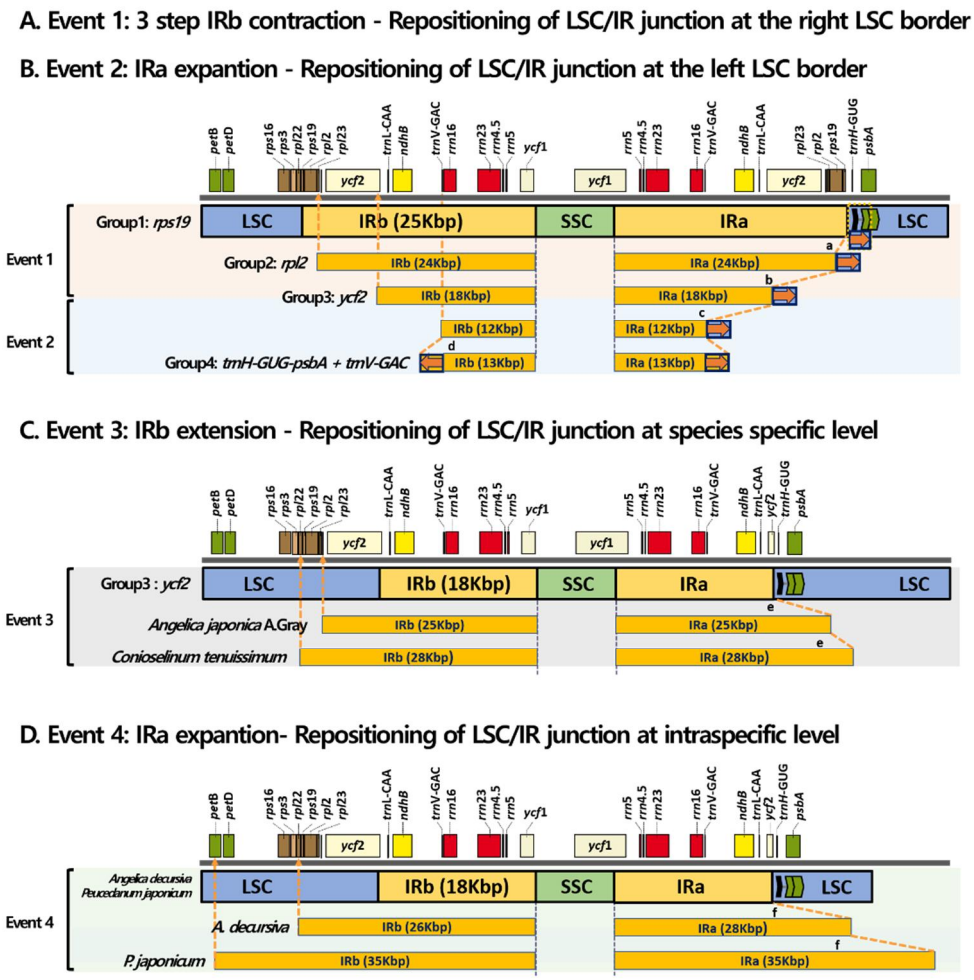
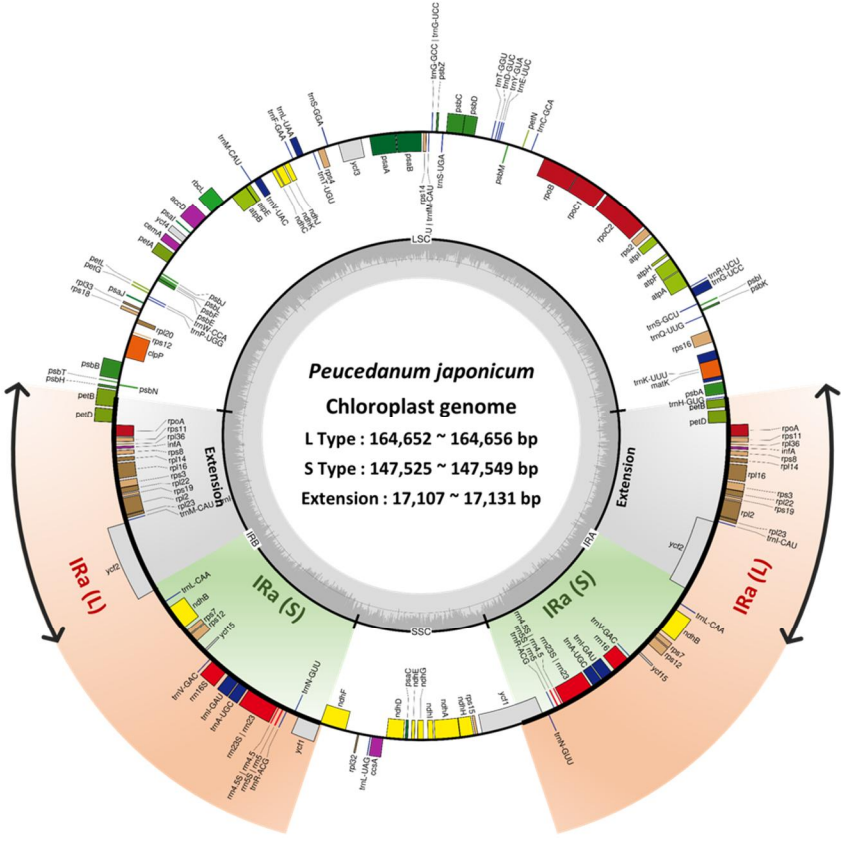


Figure 1-5. Schematic representation of the event of IR junction changes within Apioideae plastome. (A) Schematic linear plastome figure motif represents reposition of the right border of LSC/IR junction. The illustration represents the changes on IR junction from *rps19*; *rps16-rpl2*; *ycf2* gene. (B) Schematic linear plastome figure motif represents reposition of the left border of LSC/IR junction. The illustration represents the IR junction changes related to *trn*H-GUG and partial *psbA* gene insert into the IR region. (C) Schematic linear plastome figure motif represents species-specific IR

expansion. The illustration represents the LSC/IR junction changes limited to species-specific IR expansion on *A. japonica* and *C. tenuissimum*. (D) Schematic linear plastome figure motif represents intraspecific IR expansion. The illustration represents the LSC/IR junction changes on an intraspecific level resulting from IR expansion on *A. decursiva* and *P. japonicum*.

Validation of intraspecific IR variation through molecular marker analysis

The plastome of *Angelica decursiva* and *Peucedanum japonicum* indicated intraspecific variations of IR region, creating differences in plastome size around 7 kbp and 17 kbp, respectively (Figure 1-5D, Figure 1-6). Other than the LSC/IR junction boundary changes that affected duplicated gene copies in IR regions, two species have exhibited no changes in genes contents and gene order. The gene content between the two different plastome types in *A. decursiva* and *P. japonicum* is the same when excluding the IRa region; 114 genes, including 80 protein-coding genes, 30 transfer RNA (tRNA) genes, and four ribosomal RNA (rRNA) genes were commonly annotated. Plastome comparison between two species illustrated low intraspecific polymorphisms. Both InDel and SNP found in *A. decursiva* were 35 and 112, respectively (Table S1-3). The polymorphism regarding *P. japonicum* will be further discussed in Chapter 2. The InDel based molecular marker from *P. japonicum* accessions has indicated a wide range of intraspecific diversity within *P. japonicum* population (Figure 1-7). Four InDel markers were designed from tandem repeat and applied to the 20 *P. japonicum* accessions. Of 20 *P. japonicum* germplasms, number 1~10 represents *P. japonicum* germplasms with L-type, and number 11~20 represents *P. japonicum* germplasms with S-type. Applying marker to 20 *P. japoncum* germplasm in the gel figure indicated various intraspecific diversity within *P. japonicum*. The L-type population exhibited very narrow diversity; most of the population resulted with a similar genotype, while the S-type exhibits broad intraspecific diversity; depending on a marker, the genotype is separated into 2~4 types (Figure 1-7).


Peucedanum japonicum
Chloroplast genome
L Type : 164,652 ~ 164,656 bp
S Type : 147,525 ~ 147,549 bp
Extension : 17,107 ~ 17,131 bp
LSC
SSC
IRa (L)
IRa (S)
IRa (S)
IRa (L)
Extension
Extension
IRA
IRB

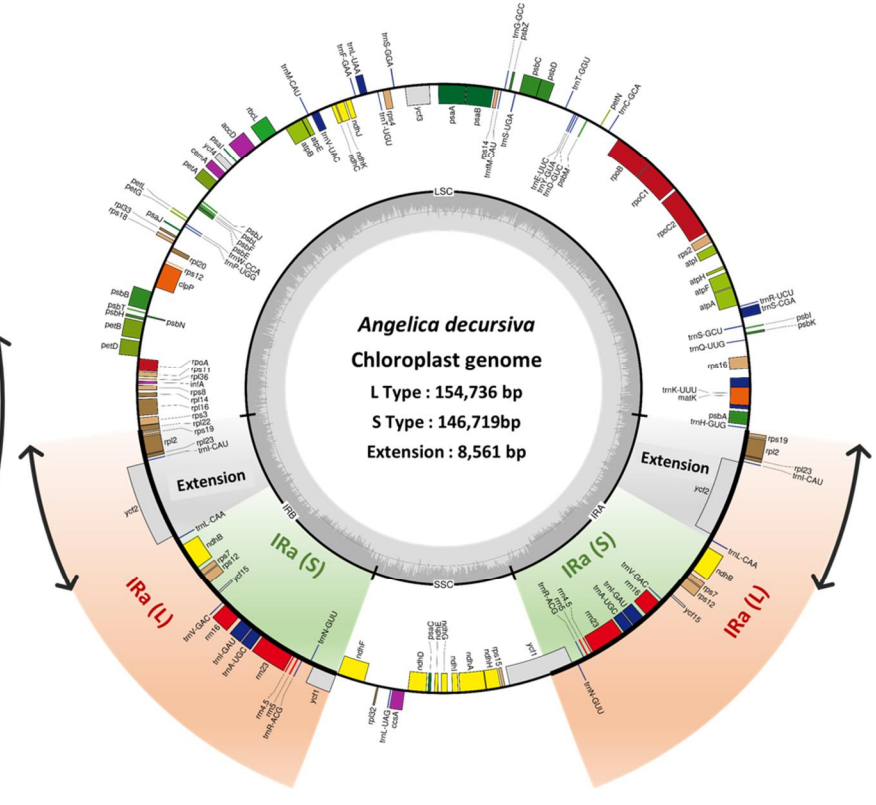

Angelica decursiva
Chloroplast genome
L Type : 154,736 bp
S Type : 146,719bp
Extension : 8,561 bp
LSC
SSC
IRa (L)
IRa (S)
IRa (S)
IRa (L)
Extension
Extension
IRA
IRB

Figure 1-6. Plastome map of *A. decursiva and P. japonicum* representing intraspecific IR variatrion. The following figure represents intraspecific IR variation discovered between the two plastome maps; one assembled for this study and the other from the NCBI Genbank database. The following plastome map indicates the size variation on the inverted repeat regions. The red-colored region indicates the long IR region. The green-colored region indicates the short IR region. The gray-colored region and the black-colored arrow indicate the difference between the long and short IR regions. The plastome map was generated using OGDRAW. Genes transcribed clockwise and counterclockwise are indicated on the outside and inside of the large circle, respectively.

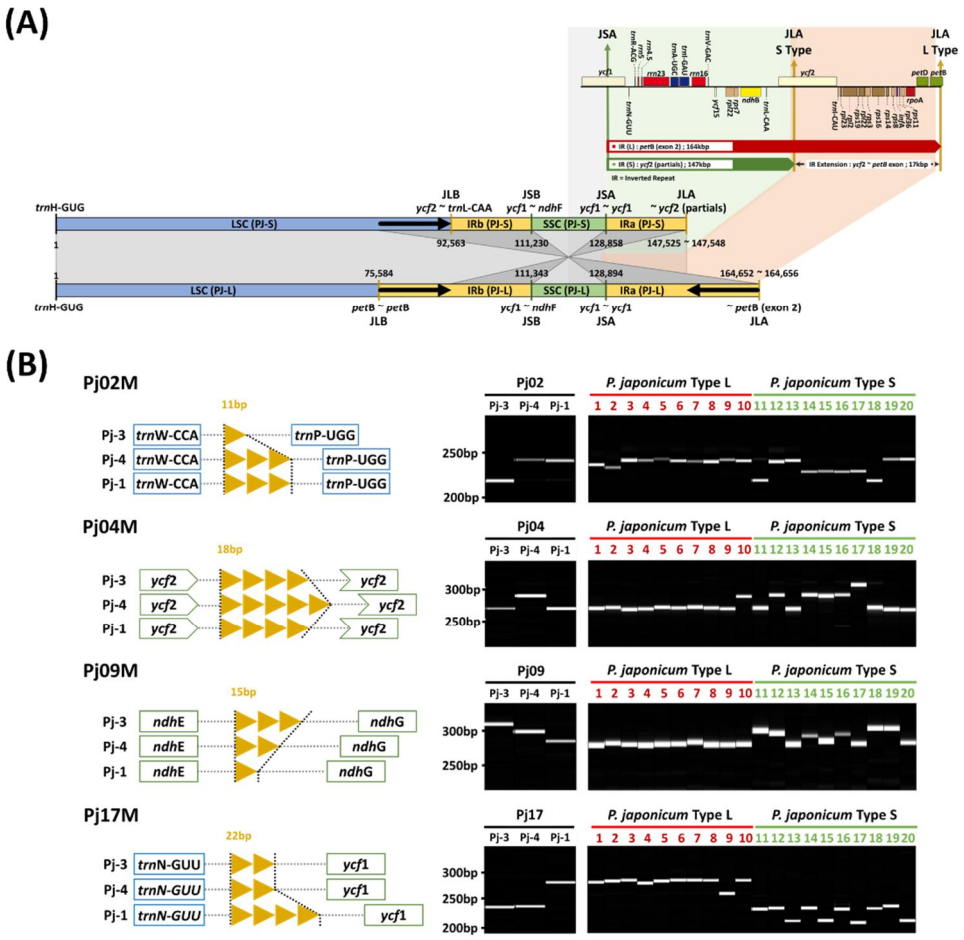


Figure 1-7. Intraspecific plastome diversity of *Peucedanum japonicum* between long (L) and short (S) type inverted repeat. (A) Schematic representation of the plastome comparison between *P. japonicum* L and S-type. The linear bar represents a quadripartite structure of the plastome. Black arrow on the near IR junction region represent the changes of IR junction between L and S-type. The top right figure represents IR's detailed plastome gene map; red color illustrates the IR region under the L-type, green color illustrates the IR region under the S-type. The colored pentagon shape box represents the difference between L and S-types of plastid gene contents

under the IR region. (B) Tandem repeat structure of four InDel regions with intraspecific diversity and application of molecular markers to 20 *P. japonicum* germplasms. The left figure represents the InDel motif detected between three *P. japonicum* plastome sequences (Pj-1, Pj-3, Pj-4) on the following molecular markers (Pj02M, Pj04M, Pj09M, Pj17M). The yellow color represents the tandem repeat motif. The size of each repeat motif is written with the same color. The right figure represents the gel result of applying the molecular marker for intraspecific diversity on 20 *P. japonicum* accessions (no. 1~10: L-type, no. 11~20: S-type). The left gel figure indicates the control band on each marker tested with three *P. japonicum* accessions (Pj-1, Pj-3, Pj-4). The genotype result is represented in Table 1-5.

Table 1-5. Summary regarding genotype evalution on four intraspecific InDel on 20 *Peucedanum japonicum* germplasms in Figure 1-7.

Primer	Product size	*P. japonicum* Type L germplasm										*P. japonicum* Type S germplasm									
	Pj-6 / Pj-7 / Pj-3 (bp)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Pj02	211 / 233 / 233	D	C	D	D	D	D	D	D	D	D	A	D	C	C	C	C	C	A	D	D
Pj04	261 / 279 / 261	A	A	A	A	A	A	A	A	A	B	A	B	A	B	B	B	C	A	A	A
Pj09	299 / 289 / 274	A	A	A	A	A	A	A	A	A	A	C	B	A	B	A	B	A	C	C	A
Pj17	233 / 233 / 277	D	D	D	D	D	D	D	D	C	D	B	B	A	B	A	B	A	B	B	A

Plastome structure rearrangement regarding tRNA gene.

Three Apioideae species, *Angelica gigas*, *Carum carvi*, and *P. japonicum*, have shown an inversion of tRNA gene from *trn*D-GUC, *trn*Y-GUA, and *trn*E-UUC into *trn*E-UUC, *trn*Y-GUA, and *trn*D-GUC. In addition, a 37 bp sequence linking the 3' end of the *trn*E-UUC gene and the 5' end of the *trn*D-GUC gene in the form of an inverted repeat after aligning the whole plastome sequence of 60 species was uncovered. Under the normal tRNA gene order *trn*D-GUC, *trn*Y-GUA, and *trn*E-UUC, the first 37 bp repeat sequence was at the 5' end of the *trn*D-GUC gene, whereas the second 37 bp sequenced was inverted and positioned at 69~75 bp away from the 3' end of the *trn*E-UUC gene. Compared to other Apioideae plastomes, the 37 bp repeat sequence in *A. gigas* and *C. carvi*, *P. japonicum* plastome was oppositely positioned. The first 37 bp repeat sequence is placed 68~70 bp away from the 3' end of the *trn*E-UUC gene, and the second 37 bp repeat sequence was placed near the start codon of 5' end of the *trn*D-GUC gene. The 37 bp sequence in the left zone (normal: *trn*D-GUC, inverted: *trn*E-UUC) was highly conserved with minimal polymorphisms. However, there was a variation in the 37 bp in the right zone (normal: *trn*E-UUC, inverted: *trn*D-GUC). A comparison of aligned plastomes from 60 Apioideae plastomes revealed that plastomes under Group 1, which includes the Araliaceae family, had 37 bp sequence in the right zone with high polymorphism compared to the 37 bp sequence in the left zone (Maximum polymorphism > 20 SNPs; Table 1-6). As a result, it is difficult to comprehend both sequences as repetitive sequences under Group 1. In both left and right zones, the species in Group 2 and Group 3 had a 37 bp repeat with a minimal polymorphism (Maximum polymorphism = 5 SNPs; Table 1-7). The modification of the IR junction

influenced changes in the plastome sequence, causing a 37 bp repeated sequence near the right zone to be linked to a 37 bp sequence in the left zone in the form of an inverted repeat. Random events associated with homologous recombination appear to have caused the inversion in *A. gigas* and *C. carvi*, *P. japonicum* species.

Table 1-6. Summary of 37 bp repeat near *trn*E-UUC, *trn*D-GUC gene on Group 1 IR junction species.

		Same	Diff	Missing	Total			Same	Diff	Missing	Total
1	*Panax ginseng cultivar Chunpoong (KM088019).* L: AAAAAGGAGAGAAGATTTATGTACTTATTGGATCCGT R: AAAACGGAAAGAT-CCCCCGGATCGAATCATATCATT C: ****.***.***.-.....*..**...***..*	19	18	0	37	13	*Sium suave (IM180813-52).* L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT R: AAAACCGAA-G--GATT---------- C: ****..***-.--...----------	7	26	4	37
2	*Aralia elata (KT153023).* L: AAAAAGGAGAGAGGATTTTGATGTACTTATTGAATGCGT R: AAAACAGAAAGATCTCTCGGATTGAATCATATCATT C: ****.**.***....*..***.*.*.**...**....---	19	17	3	39	14	*Cryptotaenia japonica (IM180813-48).* L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT R: AAAACCGAA-G--GATT---------- C: ****..***-.--****----------	7	26	4	37
3	*Hydrocotyle verticillata (NC_015818).* L: AAAAAGGAGAGAAGATTTATGTACTTATTGGATCCGT R: AAAACGGAAAGAT-CCCCCGGATCGAATCATATCATT C: ****.***.***.-.....*..**...***..*	19	18	0	37	15	*Oenanthe javanica (IM180813-50).* L: AAAAGGAAAAGATGATTGATGTACTTATTGAATCTGT R: AAAACCGAAGGATTCCTTGGATC-TAATCATATCGGT C: ****...**.***...*....*.-*.**...***.**	21	16	0	37
4	*Centella asiatica (IM180813-54).* L: AAAAGGGCGAGATAATTGATATACTTATTGGATCCAT R: AAATTTGTATATT--------CTATATACAATTTATT C: ***...*.....*-------...*.....*...*	18	19	18	55	16	*Cicuta virosa (NC_037711).* L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT R: AAAACCGAAGGATTCCTTGGATC-TAATCATATCGGT C: ****..***.***...*....*.-*.**...***.**	22	15	0	37
5	*Chamaesium delavayi (MN119367).* L: AAAAAGGAAAGATGATTGAGATTGAGGTACTTATTGAATCTGT R: AAAAAGTAAAGATTCCTT-------GGATCTAATAATATCGGT C: ******.******...*.-------**..**.**...***.**	24	12	7	43	17	*Sillaphyton podagraria (NC_033344).* L: AAAAGGGAAAGATGATTGATGTACTTATTTAATCTGT R: AAAACGTAAATATTCCTTGGAT-CTAATCATATCGGT C: ****.*.***.**...*....*-**.**...***.**	16	20	1	37
6	*Bupleurum latissimum (IM160307-6)* L: AAAAAGGAAAGATGATTGATCTACTTATTGAATCTGT R: ----------ATT------CAAATCATATCGGT C: ----------***------*..**...***..*	10	7	20	37	18	*Anthriscus sylvestris (IM151201-8).* L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT R: AAAATGGAAAAATTATTTGGATA-TAATTATATCGGT C: ****.*****.**.***....**-*.***..***.**	24	12	1	37
7	*Bupleurum falcatum (IM160307-5)* L: AAAAAGGAAAGATGATTGATCTACTTATTGAATCTGT R: ----------ATT------CAAATCATATCGGT C: ----------***------*..**...***..*	10	7	20	37	19	*Ligusticum delavayi (NC_049052).* L: AAAAGGGAAAGATGAGTGATGCACTTATTTAATCTGT R: AAAACGGAAAGATTCCTTGGATA-TAATCATATCGGT C: ****.********...*.....*-*.**...***.**	22	14	1	37
8	*Bupleurum longeradiatum (IM180813-47).* L: AAAAAGGAAAGATGATTGATCTACTTATTGAATCTGT R: ----------ATT------CAAATCATATCGGT C: ----------***------*..**...***..*	10	7	20	37	20	*Ostericum sieboldii (IM180813-40).* L: AAAAGGGAAAGATGATTGATGCACTTATTTAATCTGT R: AAAACGGAAAGATTCCTTGGATA-TAATCATATCGGT C: ****.********...*.....*-*.**...***.**	22	15	0	37
9	*Pleurospermum camtschaticum (IM180813-51).* L: AAAAAGGAAAGATGATTGATGTA----CTTATTGAATCTGT R: AAAACGGAAGGGAAGGATTCCTTGGATCTAATCACATCGGT C: ****.****.*.....----**.**...***.**	23	8	16	47	21	*Pterygopleurum neurophyllum (NC_033345).* L: AAAAGGGAAAGATGATTGATGCACTTATTTAATCTGT R: AAAACGGAAAGATTCCTTGGATA-TAATCATATCGGT C: ****.********...*.....*-*.**...***.**	21	15	1	37
10	*Hansenia weberbaueriana (NC_035053).* L: AAAAAGGAAAGATGATTGATGTACTTATTGAATCTGT R: AAAACGGAAAGATTCCTTGGAT-CTAATCATATTGGT C: ****.********...*....*.**.**...**..**	22	14	1	37	22	*Caucalis platycarpos (KX832334.1).* L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT R: AAAACGGAAAGATTCCTTGGATA-TAATCATATCGGT C: ****.********...*....**.*.**...***.**	24	13	0	37
11	*Chuanminshen violaceum (KU921430.2).* L: AAAAAGGAAAGATGATTGATGTACTTATTGAATCTGT R: AAAACGGAAAGATTCCTTGGATCTAATCATATTG-GT C: ****.********...*....*.....*.*.-**	22	15	0	37	23	*Cuminum cyminum (NC_046879).* L: AAAAGGGAAAGATTATTGATGTACTTATTGAATCTGT R: ---------- C: ----------	0	0	37	37
12	*Tiedemannia filiformis (HM596071.1).* L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT R: AAAACCGAA-G--GATT---------- C: ****..***-*--****----------	7	4	26	37	24	*Daucus carota (NC_008325).* L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT R: AAAACGGAAGGAT--TCCGTGTATATAATCATAT--- C: ****.****.***--*...****..**.*.*...---	23	14	0	37

* L: refer to the 37 bp sequence in the left zone, and R: refer to the 37 bp sequence in the right zone. C refer to the clustered sequence of 37 bp from both left and right zone.

Table 1-7. Summary of 37 bp repeat near *trn*E-UUC, *trn*D-GUC gene on Group 2 and Group 3 IR junction species.

25	*Crithmum maritimum (NC_015804.1).*	Same	Diff	Missing	Total	43	*Peucedanum japonicum (IM150601-18)*	Same	Diff	Missing	Total
	L: AAAAGGGAAAGATGATGGATGTACTTATTGAATCTGT						L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT				
	R: AAAAGGGAAAGATGATGGATGTACTTATTGAATCTGT						R: AAAACGGAAAGATGATTGATGTACTTATTGAATCTGT				
	C: **********	37	0	0	37		C: ****.**********	36	1	0	37
26	*Carum carvi (NC_029889.1).*	Same	Diff	Missing	Total	44	*Peucedanum japonicum (IM200128-1)*	Same	Diff	Missing	Total
	L:AAAACGGAAAGATGATGGATGTACTTATTGAATCTGT						L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT				
	R:AAA-CGGAAAGATGATGGATGTACTTATTGAATCTGT						R: AAAACGGAAAGATGATTGATGTACTTATTGAATCTGT				
	C: ***.**********	36	0	1	37		C: ****.**********	36	1	0	37
27	*Petroselinum crispum (NC_015821.1).*	Same	Diff	Missing	Total	45	*Arracacia xanthorrhiza (NC_032364.1).*	Same	Diff	Missing	Total
	L: AAAGGGAAAAGATGATTGATGTACTTATTGAATCTGT						L: AAAAGGGAAAGATGATTGGTGTACTTATTGAATCTGT				
	R: AAAACGGAAAAATGATGGATGTACTTATTGAATCTGT						R: AAAACGGAAAGATGATGGATGTACTTATTGAATCTGT				
	C: ***.*.***.*****.**********	32	5	0	37		C: ****.**********.*.**********	34	3	0	37
28	*Anethum graveolens (NC_029470.1).*	Same	Diff	Missing	Total	46	*Cnidium monnieri (IM180813-53).*	Same	Diff	Missing	Total
	L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT						L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT				
	R: AAAACGTAAAAATTATGGATGTACTTATTGAATCTGT						R: AAAACGGAAAGATGATGGATGTACTTATTGAATCTGT				
	C: ****.*.***.**.**.**********	32	5	0	37		C: ****.**********.**********	35	2	0	37
29	*Foeniculum vulgare (NC_029469.1).*	Same	Diff	Missing	Total	47	*Cnidium japonicum (IM180813-43).*	Same	Diff	Missing	Total
	L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT						L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT				
	R: AAAACGGAAAAATTCTGGATGTACTTATTGAATCTGT						R: AAAACGGAAAGATGATGGATGTACTTATTGAATCTGT				
	C: ****.*****.**..*.**********	32	5	0	37		C: ****.**********.**********	35	2	0	37
30	*Prangos trifida (NC_037852.1).*	Same	Diff	Missing	Total	48	*Angelica acutiloba (NC_029391.1).*	Same	Diff	Missing	Total
	L: AAAAGGGAAAGATGATGGATGTACTTATTGAATCTGT						L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT				
	R: AAAAGGGAAAGATGATGGATGTACTTATTGAATCTGT						R: AAAACGGAAAGATGATGGATGTACTTATTGAATCTGT				
	C: **********	37	0	0	37		C: ****.**********.**********	35	2	0	37
31	*Coriandrum sativum (IM180813-44).*	Same	Diff	Missing	Total	49	*Angelica japonica A. gray (IM151201-2).*	Same	Diff	Missing	Total
	L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT						L: AAAGGGGAAAGATGATTGATGTACTTATTGAATCTGT				
	R: AAAACGGAAAGATGATGGATGTACTTATTGAATCTGT						R: AAAACGGAAAGATGATGGATGTACTTATTGAATCTGT				
	C: ****.**********.**********	35	2	0	37		C: ***..**********.**********	35	2	0	37
32	*Levisticum officinale (IM180813-41).*	Same	Diff	Missing	Total	50	*Peucedanum terebinthaceum (NC_053641).*	Same	Diff	Missing	Total
	L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT						L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT				
	R: AAAACGGAAAGATGATGGATGTACTTATTGAATCTGT						R: AAAACGGAAAGATGATGGATGTACTTATTGAATCTGT				
	C: ****.**********.**********	35	2	0	37		C: ****.**********.**********	35	2	0	37
33	*Cnidium officinale (IM180813-27).*	Same	Diff	Missing	Total	51	*Peucedanum hakuunense (IM180813-21)*	Same	Diff	Missing	Total
	L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT						L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT				
	R: AAAACGTAAAGATGATGGATGTACTTATTGAATCTGT						R: AAAACGGAAAGATGATGGATGTACTTATTGAATCTGT				
	C: ****.*.*********.**********	34	3	0	37		C: ****.**********.**********	35	2	0	37
34	*Conioselinum tenuissimum (IM180813-29).*	Same	Diff	Missing	Total	52	*Glehnia littoralis (NC_034645.1).*	Same	Diff	Missing	Total
	L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT						L: AAAACCGAAAGATGATTGATGTACTTATTGAATCTGT				
	R: AAAACGTAAAGATGATGGATGTACTTATTGAATCTGT						R: AAAACCGAAAGATGATGGATGTACTTATTGAATCTGT				
	C: ****.*.*********.**********	34	3	0	37		C: ********** **********	36	1	0	37
35	*Heracleum moellendorffii (IM180813-45).*	Same	Diff	Missing	Total	53	*Angelica dahurica (IM151201-3).*	Same	Diff	Missing	Total
	L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT						L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT				
	R: AAAACGGAAAGATGATGGATGTACTTATTGAATCTGT						R: AAAACCGAAAGATGATGGATGTACTTATTGAATCTGT				
	C: ****.**********.**********	35	2	0	37		C: ****..**********.**********	34	3	0	37

36	*Pastinaca sativa (IM180813-46).*	Same	Diff	Missing	Total	54	*Angelica nitida (MF594405.1).*	Same	Diff	Missing	Total
	L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT						L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT				
	R: AAAACGGAAAGATGATGGATGTACTTATTGAATCTGT						R: AAAACCGAAAGATGATGGATGTACTTATTGAATCTGT				
	C: ****.**********.**********	35	2	0	37		C: ****..**********.**********	34	3	0	37
37	*Seseli montanum (NC_027451.1).*	Same	Diff	Missing	Total	55	*Angelica jaluana (IM180813-42).*	Same	Diff	Missing	Total
	L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCT-T						L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT				
	R: AAAACGGAAAGATGATTGATGTACTTATTGAATCTGT						R: AAAACCGAAAGATGATGGATGTACTTATTGAATCTGT				
	C: ****.**********-*	36	1	0	37		C: ****..**********.**********	34	3	0	37
38	*Ledebouriella seseloides (NC_034643.1).*	Same	Diff	Missing	Total	56	*Angelica decursiva (IM180813-31).*	Same	Diff	Missing	Total
	L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT						L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT				
	R: AAAACGGAAAGATGATTGATGTACTTATTGAATCTGT						R: AAAACCGAAAGATGATGGATGTACTTATTGAATCTGT				
	C: ****.**********	36	1	0	37		C: ****..**********.**********	34	3	0	37
39	*Dystaenia takesimana (IM180813-49).*	Same	Diff	Missing	Total	57	*Angelica decursiva (KT781591.1).*	Same	Diff	Missing	Total
	L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT						L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT				
	R: AAAACGGAAAGATGATTGATGTACTTATTGAATCTGT						R: AAAACCGAAAGATGATGGATGTACTTATTGAATCTGT				
	C: ****.**********	36	1	0	37		C: ****..**********.**********	34	3	0	37
40	*Peucedanum praeruptorum (MN016968).*	Same	Diff	Missing	Total	58	*Angelica gigas (IM151201-7).*	Same	Diff	Missing	Total
	L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT						L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCCGT				
	R: AAAACGGAAAGATGATTGATGTACTTATTGAATCTGT						R: AAAACCGAAAGATGATGGATGTACTTATTGAATCTGT				
	C: ****.**********	36	1	0	37		C: ****..**********.**********.**	35	2	0	37
41	*Peucedanum japonicum (IM150921-6)*	Same	Diff	Missing	Total	59	*Angelica amurensis (IM180813-30).*	Same	Diff	Missing	Total
	L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT						L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT				
	R: AAAACGGAAAGATGATTGATGTACTTATTGAATCTGT						R: AAAACCGAAAGATGATGGATGTACTTATTGAATCTGT				
	C: ****.**********	36	1	0	37		C: ****..**********.**********	34	3	0	37
42	*Peucedanum japonicum (IM150601-19)*	Same	Diff	Missing	Total	60	*Angelica reflexa (IM180813-28).*	Same	Diff	Missing	Total
	L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT						L: AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGT				
	R: AAAACGGAAAGATGATTGATGTACTTATTGAATCTGT						R: AAAACCGAAAGATGATGGATGTACTTATTGAATCTGT				
	C: ****.**********	36	1	0	37		C: ****..**********.**********	34	3	0	37

* L: refer to the 37 bp sequence in the left zone, and R: refer to the 37 bp sequence in the right zone. C refer to the clustered sequence of 37 bp from both left and right zone.

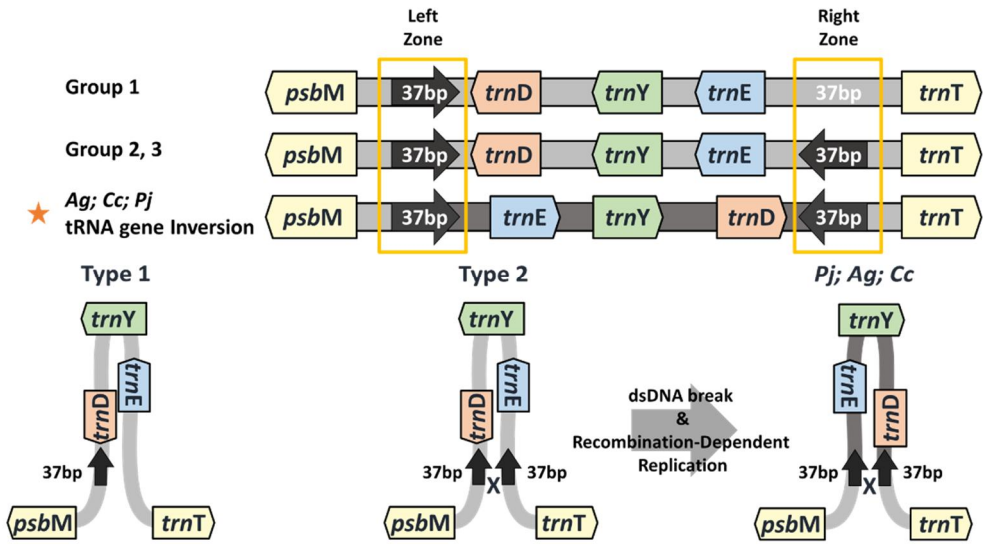


Figure 1-8. Schematic figure representing the *trn*E-UUC, *trn*Y-GUA, and *trn*D-GUC inversion under *Angelica giga* (*Ag*), *Carum carvi* (*Cc*), *and Peucedanum japonicum* (*Pj*). The black arrow represents a 37 bp sequence located near the junction of the tRNA gene linked in a repeat sequence. The dark gray area under *Ag*, *Cc*, and *Pj* tRNA gene order represents the inverted region compared to other Apioideae species.

Divergence time estimation among Apioideae species

Based on plastid CDS sequences, divergence time upon major clade of Apioideae was estimated with median height along with 95% HPDs (Figure 1-9). The divergence time between Araliaceae and Apiaceae was estimated as 65.82 Mya with a 95% HPD interval of 56.44~70.68 Mya in the Late Cretaceous period. The divergence time of Apiaceae subfamily Mackinlayoideae from subfamily Apioideae was estimated at 62.69 Mya with a 95% HPD interval of 53.15~69.39 Mya in the Late Paleocene period. Within Apioideae, estimated divergence time on each of the following tribes and clade under Group 1 were shown as follows: Chamaesieae at 45.03 Mya (39.28~52.19 Mya in 95% HPD), Burpleureae at 41.28 Mya (36.29~47.38 Mya in 95% HPD) and Pleurospermeae at 34.56 Mya (31.51~40.32 Mya in 95% HPD) in Eocene period. *Komarovia* Clade and *Physospermopsis* clade were separated around 31.26 Mya (27.46~36.81 Mya in 95% HPD), Oenantheae tribe has diverged at 27.65 Mya (23.57~33.06 Mya in 95% HPD). The following tribe and clade under *Arcuatopterus* clade, Scandiceae, *Acronema* clade, Torilidinae, Daucinae were separated at 23.28 Mya (21.16~30.52 Mya in 95% HPD) in the late Oligocene period. Split between Group 1 and Group 2 (ancestor to Careae, Pyramidoptereae, *Cachrys* clade, Apieae) occurred around 25.28 Mya (21.16~30.52 Mya in 95% HPD). Under the tribe and clade of Group 2 species, Careae and Pyramidoptereae formed a sister clade and diverged around 17.36 Mya (12.74~22.74 Mya in 95% HPD). *Cachrys* Clade and Apieae were separated from Group 3 around 17.15 Mya (12.89~22.40 Mya in 95% HPD). *Cachrys* clade was split from tribe Apieae around 15.33 Mya (11.02~20.49 Mya in 95% HPD). Tribe and clade under Group 3, Coriandreae and *Sinodielsia* clade, were separated from Tordylieae and Selineae tribe

at 11.99 Mya (8.65~16.59 Mya in 95% HPD). Later the common ancestor of Tordylieae separated from the Selineae tribe around 11.72 Mya (8.45~16.22 Mya in 95% HPD). Selineae tribe demonstrated a separation into two groups with polyphyletic relationship under the *Peucedanum* genus. The group that contains *S. montanum*, *L. seseloides*, *D. takeshimana*, *P. praeruptorum,* and *P. japonicum*; diverged from other species of Selineae around 9.45 Mya (6.72~13.30 in 95% HPD). The second group contains *Angelica*, *Cnidium*, and *Peucedanum* genus (*P. terebinthaceum* and *P. hakuunense*) were separated around 8.96 Mya (6.38~12.65 Mya in 95% HPD) in the Miocene period. The following species under the Selineae tribe, *A. decursiva* and *P. japonicum* with intraspecific IR variation have demonstrated a separation within the species 1.17 Mya (0.51~2.24 Mya in 95% HPD) and 0.88 Mya (0.40~1.67 Mya in 95% HPD) in the middle of the Quaternary period, respectively (Figure 1-9).

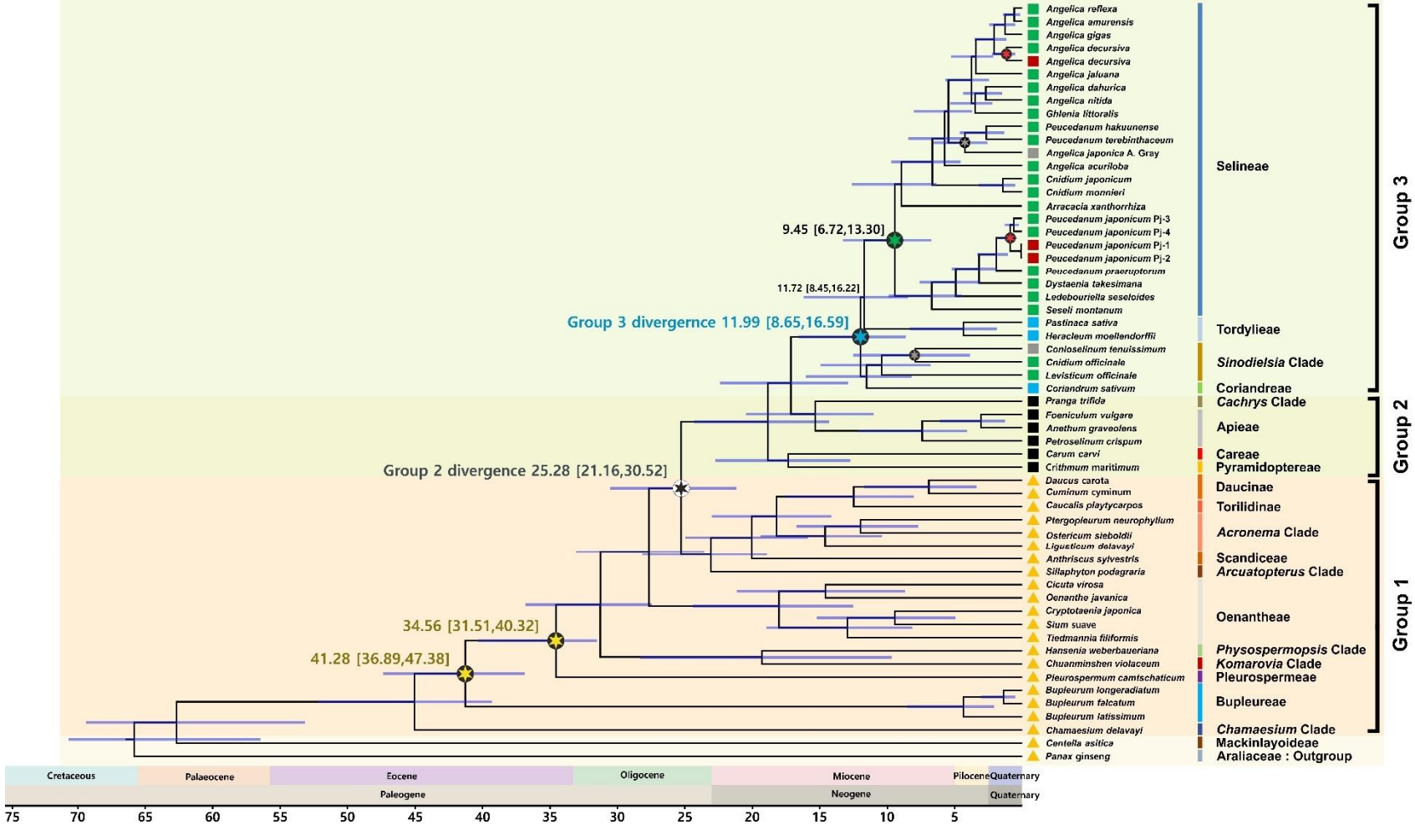

Angelica reflexa
Angelica amurensis
Angelica gigas
Angelica decursiva
Angelica decursiva
Angelica jaluana
Angelica dahurica
Angelica nitida
Ghlenia littoralis
Peucedanum hakuunense
Peucedanum terebinthaceum
Angelica japonica A. Gray
Angelica acuriloba
Cnidium japonicum
Cnidium monnieri
Arracacia xanthorrhiza
Peucedanum japonicum Pj-3
Peucedanum japonicum Pj-4
Peucedanum japonicum Pj-1
Peucedanum japonicum Pj-2
Peucedanum praeruptorum
Dystaenia takesimana
Ledebouriella seseloides
Seseli montanum
Pastinaca sativa
Heracleum moellendorffii
Conioselinum tenuissimum
Cnidium officinale
Levisticum officinale
Coriandrum sativum
Pranga trifida
Foeniculum vulgare
Anethum graveolens
Petroselinum crispum
Carum carvi
Crithmum maritimum
Daucus carota
Cuminum cyminum
Caucalis playtycarpos
Ptergopleurum neurophyllum
Ostericum sieboldii
Ligusticum delavayi
Anthriscus sylvestris
Sillaphyton podagraria
Cicuta virosa
Oenanthe javanica
Cryptotaenia japonica
Sium suave
Tiedmannia filiformis
Hansenia weberbaueriana
Chuanminshen violaceum
Pleurospermum camtschaticum
Bupleurum longeradiatum
Bupleurum falcatum
Bupleurum latissimum
Chamaesium delavayi
Centella asitica
Panax ginseng
Selineae
Tordylieae
Sinodielsia Clade
Coriandreae
Cachrys Clade
Apieae
Careae
Pyramidoptereae
Daucinae
Torilidinae
Acronema Clade
Scandiceae
Arcuatopterus Clade
Oenantheae
Physospermopsis Clade
Komarovia Clade
Pleurospermeae
Bupleureae
Chamaesium Clade
Mackinlayoideae
Araliaceae : Outgroup
Group 3
Group 2
Group 1
9.45 [6.72,13.30]
11.72 [8.45,16.22]
Group 3 divergernce 11.99 [8.65,16.59]
Group 2 divergence 25.28 [21.16,30.52]
34.56 [31.51,40.32]
41.28 [36.89,47.38]
Cretaceous
Palaeocene
Eocene
Oligocene
Miocene
Pilocene
Quaternary
Paleogene
Neogene
Quaternary
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5

Figure 1-9. Chronogram illustration on divergence time estimation between Apioideae species estimated through MCMC tree analysis using 78 chloroplast coding sequences (CDS). The tree was calibrated with tertiary fossil pollen data of tribes Burpleureae and Pleurospermeae from the study by Banasiak et al., and the second calibration point from the study by Wen et al. [23, 41]. Each calibration point is labeled with a yellow star. The purple bar represents the 95% highest posterior density (95% HPD) for each node. Scale axis is based on million years ago (Mya). For example, the following point labeled with black, blue, and green stars indicates where groups 1, 2 and 3 were separated. Gray and red stars indicates *trn*H-GUG transition and intraspecific IR junction were separated, respectively.

DISCUSSION

Dynamic IR junction shift with a phylogenetic relationship in Apioideae

The phylogenetic result of Apioideae species in this study was similar to other previous studies [2, 23, 42]. However, several differences when comparing the phylogenetic result from a recent study by Wen et al. was detected [23]. For example, the *Prango trifida* of the *Cachrys* clade is found between the Apieae and Coriandreae with low branch supporting value (BS=60), whereas the previous study was found between Pimpinelleae and Apieae (Figure 1-3A). Furthermore, the phylogenetic trees illustrated in the previous study have mentioned polyphyletic relationships on the tribe level on Pimpinelleae and Tordylieae that could not discover due to the limited sampling in this study. One of the goals in this study was to analyze complete plastomes of the Apioideae species distributed in Korea; therefore, 31 species under ten genera of Apioideae species, including rare and endemic species such as *D. takeshimana* and *P. hakuunense*, was collected. Compared to the number of distributed Apioideae species in Korea, the number of collected species in this study is small. However, given the large number of studies on Apioideae that focus on intra-genus level analysis, by including a representative species of tribe and clade under Apioideae species from NCBI, was able to investigate the relationship of Apioideae species on a larger scale.

The plastome of *Peucedanum japonicum* and *Peucedanum hakuunense*, which are the rare and endemic species in Korea were assembled. By including other

plastomes of *Peucedanum* species (*P. terebinthaceum* and *P. praeruptorum*), the *Peucedanum* species are demonstrated polyphyletic relationship under groups of the Selineae tribe (Figure 1-3). The polyphyletic relationships on *Peucedanum* species have often been mentioned in other previous studies [2, 12]. However, in a recent study on the Apioideae [23], the polyphyly of *Peucedanum* genus has not been mentioned. The genus *Peucedanum* is one of the worldwidely distributed genus, as more and more plastome sequences are accessible, such polyphyletic problem would become a significant issue. Unfortunately, the issue related to the polyphyly of the *Peucedanum* genus could not be resolved in this study. As for the moment, a more plastome sequence or large amounts of nuclear genes is required for resolving the phylogenetic relationship of *Peucedanum*. Furthermore, *Peucedanum insolens* which is endemic to Korea, has recently been changed its nomenclature into *Sillaphyton podagraria*, with the support of both morphological and genetic analysis [43]. These imply that a comprehensive study on both plastome and morphological study is necessary to solve the polyphyly of *Peucedanum* and origin of *Sillaphyton podagraria*.

Throughout investigating the IR junction within Apioideae species, an IR junction scenario that affected the plastome diversity of Apioideae was hypothesized. The phylogenetic analysis has illustrated that the tribe and clade on Apioideae species experienced difference structural changes within plastome size by the changes on the quadripartite junction between LSC and IR region (Figure 1-4, Table 1-3). Plastid genome size is greatly affected by the expansion and contraction of the IR region, but there are some exceptions. Within selected Apioideae species, *C. platycarpos* exhibited the largest plastome size (Table 1-3). Thus, assumed that its IR junction must be greater

than the other species under Group 1. However, the sequence analysis through blast analysis indicated that *C. platycarpos* exhibited an inserted sequence with IR region, where the inserted sequence represented high sequence homology from the *C. sativum* mitochondria sequence in blastn analysis [23]. The inserted sequence has demonstrated alteration in the length of the IR region, but IR junction within *C. platycarpos* still remains near the *rps19* gene (Table 1-3). The results imply that while the plastome size is strongly affected by the IR region, the junction between LSC and IR region is not. It seems more likely that the changes in the IR junction are more closely related to the phylogenetic relationship among tribe and clade of the associated species under Apioideae. Coordinating the event regarding IR junction on Apioideae species has visualized the IR contraction has occurred compared to the basal group and Araliaceae (Figure 1-4, Table 1-3). Investigating 30 representative species under the Araliaceae family illustrated that the LSC/IR junction boundary was located near the *rps19* gene with IR region length around 25 kbp (Table S1-1). As the Apioideae species diverge, the LSC/IR junction boundary has repositioned its junctions and eventually reached junction boundary near 3' end of the *ycf2* gene region on Selineae tribe along with IR region length decreased to 18 kbp (Figure 1-5, Table 1-3).

From such changes observed under Apioideae species, the scenario on plastome variation related to IR expansion or contraction resulting from the repositioning of IR junction near the LSC border was hypothesized (Figure 1-4, Figure 1-5). The first two events are related to the reposition of the IR junction regarding IR contraction. As Apioideae species diverged from Araliaceae, IR junction started to change its position from near 3' end of *rps19* to 3' end of *ycf2* (Figure 1-5A, B).

However, during that reposition, species under Coriandreae and Tordylieae have experienced repositioning of the IR junction near the LSC's left border related to *trn*H-GUG transition into the IR region. From this point, the IR expansion begins to appear within the Apoioideae species. Third and fourth events have illustrated IR expansion specific to certain species: *A. japonica* and *C. tenuissimum* (Figure 1-5C) or IR expansion occur within intraspecific level: *A. decursiva* and *P. japonicum*.

Discovery of intraspecific IR variation within Apioideae species.

Among the Apioideae species, the unique discovery was the two kinds of the plastome that differs in the genome size around 7 kbp and 17 kbp were found in both *Angelica decursiva* and *Peucedanum japonicum,* respectively (Figure 1-6). The discovery is noteworthy, as this is the first to report on the phenomenon of a plant species representing intraspecific differences on plastome size provided by the IR region greater than 2 kbp. Plastome variation related to the IR region regarding IR expansion or contraction has been well studied under plant species even before the super barcoding method was developed [45]. However, studies were primarily based on interspecific comparison since the conserved genomic feature on plastome exhibits high interspecific diversity than intraspecific diversity. At the interspecific level, variation in plastome size is greater or equal to 2 kbp difference is often noticed between species under the same genus [23, 45, 46, 47]. However, such a variation on intraspecific level, plastome variation up to 7 kbp or 17 kbp has not yet been reported.

Both *A. decursiva* and *P. japonicum* experienced intraspecific IR expansion.

The IR junction between two species demonstrated a shift from *ycf2* to *rpl22* and *petB* intron region, without modification of gene contents and gene order (Figure 1-5D, Figure 1-6). To ensure that intraspecific IR expansion is not generated by the sequencing or assembly error, molecular markers were tested onto the *P. japonicum* population and verified the intraspecific plastome variation on the IR region (Figure 1-7). Due to a limited number of accessions on the *A. decursiva* sample, the validation of the molecular markers was not obtainable. However, as the same phenomenon appeared within *P. japonicum*, it is fair to assume the intraspecific variation exists in the plastomes of *A. decursiva* as well. Population based study on validating Intraspecfic IR on *A. decusiva* would be necessary for future study. Outline of phylogenetic relationship from and the four events of IR junction illustrated that the IR junction boundary of Apioideae species experienced changes as the species diverges (Figure 1-3, 1-5). Though the plastome seems to reach the stable state around the Selineae tribe, the exhibition of intraspecific IR expansion on both *A. decursiva* and *P. japonicum* implies that the IR junction changes might still be active (Figure 1-3, Figure 1-6).

Another event related with IR expansion; a specifc IR expansion on *Angelica japonica* demonstrated its plastome size 154,302 bp with IR junction (the *rpl22* gene) which differ from its sister species *Peucedanum hakuuenese* and *P. terebinthaceum* (147 kbp, the *ycf2* gene). Considering the intraspecific IR expansion event, the hypothesis of the existence of another *A. japonica* species with the plastid genome size around 147 kbp with IR junction near the *ycf2* gene region was speculated. The confirmation of this hypothesis would be an assignment for future study.

tRNA gene inversion caused by repeat variation

Based on the discovery, the inversion of the three tRNA genes: *trn*E-UUC, *trn*Y-GUA, *trn*D-GUC, may have started around the common ancestor of Group 2 (Figure 1-3). The 37 bp repeat sequences were found nearby the junction of the three tRNA genes on both left and right zones (Figure 1-8). Between the 37 bp sequences, the number of polymorphisms appeared significantly on right zone 37 bp under Group 1 species compared to Group 2 and 3. As Group 2 species separated from Group 1, the accumulated evolutionary force pushed the 37 bp inverted repeat region near the tRNA gene in the right zone then tRNA inversion occurred through a random recombination-dependent replication process. Thus, the inversion process of the *trn*E-UUC, *trn*Y-GUA, *trn*D-GUC region was completed (Figure 1-8).

Divergence time estimation within Apioideae.

The timeline of the events was estimated regarding the changes upon IR junction through divergence time estimation. The IR junction changes detected from the node near Pyramidoptereae and Daucinae have estimated around 25.25 Mya (21.16~30.52 Mya in 95% HPD) during the Late Oligocene (Figure 1-9). The IR junction near *rpl2* region observed in Group 2 has occurred around 17.15 Mya (12.89~22.40 Mya in 95% HPD). The IR junction near *trn*H-GUG transition (*C. sativum*, *H. moellendorffii,* and *P. sativa*) observed in Group 3 has occurred around 11.99 Mya (8.65~16.59 Mya in 95% HPD). IR junction near *ycf2* gene region observed in Group 3 has occurred around 9.45 Mya (6.72~13.30 Mya in 95% HPD) (Figure 1-9).

The separation between Group 1 and Group 2 occurred around 25 Mya, near the late Oligocene warming event (Figure 1-9). Such changes in the environment may have accelerated the diversification of the species. The uplift history of the Qinghai-Tibetan Plateau (QTP) may have strongly influenced the high speciation rates on the species under Group 2 [44]. The uplift of QTP resulted from the collision of the Indian plate with Eurasia around 55~50 Mya [44]. From so on, uplift progressed during Oligocene and Miocene in the direction of north, west, south as the times passed. Favre et al. stated that during Miocene, the uplift of QTL had influenced the Asian monsoon, leading to a warm and humid climate and subtropical vegetations. Most species under Group 2 are widely distributed in Asia. The following species under Group 2 and 3, LSC/IR junction, experienced dynamic changes mostly around the Miocene epoch. Through the uplift of QTL, the environmental changes on both geological and climatic may have strongly influenced the adaptive evolution that has affected the plastome regarding Group 2 and 3 of Apioideae species.

Overall, the plastome of 60 Apioideae species, including newly assembled 31 Apioideae species in Korea, has revealed a significant event related to the changes in IR junction compared to its sister family Araliaceae. Phylogenetic analysis subdivided Apioideae species into three groups related to the changes in IR junction leading to IR contraction from the *rps19* gene to the ycf2 gene. A few IR expansions from *trn*H-GUG transition into IRa region, species-specific IR expansion, and unique intraspecific IR expansion. This study provides valuable information for further studies related to the Apioideae species.

REFERENCES

1. Plunkett, GM., Chandler, GT., Lowry II, PP., Pinney, SM., Sprenkle, TS., Van Wyk, BE., and Tilney, PM. (2004). Recent advances in understanding Apiales and a revised classification. *South African Journal of Botany* 70(3), 371-381.

2. Downie, SR., Spalik, K., Katz-Downie, DS., and Reduron, JP. (2010). Major clades within Apiaceae subfamily Apioideae as inferred by phylogenetic analysis of nrDNA ITS sequences. *Plant Diversity and Evolution* 128(1), 111.

3. Koch, WDJ. (1825). *Generum tribuumque plantarum umbelliferarum nova dispositio: Cum tab. tribus lithograptis*.

4. Bentham, G., and Hooker, JD. (1867). Umbelliferae. *Genera plantarum* 1(3), 859-931.

5. Drude, O. (1897). Umbelliferae. *Die naturlichen Pflanzenfamilien* 3, 63-150.

6. Koso-Poljansky, BM. (1916). Sciadophytorum systematis lineamenta. *Bull. Soc. Imp. Naturalistes Moscou* 29, 93-221.

7. Cerceau-Larrival, M. (1962). *Plantules et polens d'Ombellifères: leur intérêt systématique et phylogénique*. Éditions du Muséum.

8. Pimenov, MG., and Leonov, MVE. (1993). *The genera of the Umbelliferae: a nomenclator*. Royal Botanic Gardens, Kew.

9. Plunkett, GM., Soltis, DE., and Soltis, PS. (1997). Clarification of the relationship between Apiaceae and Araliaceae based on *matK* and *rbcL* sequence data. *American Journal of Botany* 84(4), 565-580.

10. Downie, SR., Plunkett GM., Watson MF., Spalik K., Katz-Downie DS.,

Valiejo-Roman CM., Terentieva EI., Troitsky AV., Lee BY., and Lahham J. (2001). Tribes and clades within Apiaceae subfamily Apioideae: the contribution of molecular data. *Edinburgh Journal of Botany* 58 (2), 301-330.

11. Downie, SR., Ramanath, S., Katz-Downie, DS., and Llanas, E. (1998). Molecular systematics of Apiaceae subfamily Apioideae: phylogenetic analyses of nuclear ribosomal DNA internal transcribed spacer and plastid *rpo*C1 intron sequences. *American Journal of Botany* 85(4), 563-591.

12. Spalik, K., Reduron, JP., and Downie, SR. (2004). The phylogenetic position of *Peucedanum* sensu lato and allied genera and their placement in tribe Selineae (Apiaceae, subfamily Apioideae). *Plant Systematics and Evolution* 243(3), 189-210.

13. Spalik, K., Downie, SR., and Watson, MF. (2009). Generic delimitations within the Sium alliance (Apiaceae tribe Oenantheae) inferred from cpDNA *rps16*-5′ *trn*K (UUU) and nrDNA ITS sequences. *Taxon* 58(3), 735-748.

14. Spalik, K., Banasiak, Ł., Feist, MAE., and Downie, SR. (2014). Recurrent short-distance dispersal explains wide distributions of hydrophytic umbellifers (Apiaceae tribe Oenantheae). *Journal of Biogeography* 41(8), 1559-1571.

15. Valiejo-Roman, CM., Terentieva, EI., Samigullin, TH., Pimenov, MG., Ghahremani-Nejad, F., and Mozaffarian, V. (2006). Molecular data (nrITS-sequencing) reveal relationships among Iranian endemic taxa of the Umbelliferae. *Feddes Repertorium: Zeitschrift für botanische Taxomonie und Geobotanik* 117(5-6), 367-388.

16. Sun, FJ., and Downie, SR. (2010). Phylogenetic relationships among the perennial, endemic Apiaceae subfamily Apioideae of western North America: additional data from the cpDNA *trn*F-*trn*L-*trn*T region continue to support a highly

polyphyletic Cymopterus. *Plant Diversity and Evolution* 128(1), 151.

17. Downie, SR., and Katz-Downie, DS. (1996). A molecular phylogeny of Apiaceae subfamily Apioideae: evidence from nuclear ribosomal DNA internal transcribed spacer sequences. *American Journal of Botany* 83(2), 234-251.

18. Katz-Downie, DS., Valiejo-Roman, CM., Terentieva, EI., Troitsky, AV., Pimenov, MG., Lee, B., and Downie, SR. (1999). Towards a molecular phylogeny of Apiaceae subfamily Apioideae: additional information from nuclear ribosomal DNA ITS sequences. *Plant Systematics and Evolution* 216(3), 167-195.

19. Downie, SR., Katz-Downie, DS., and Watson, MF. (2000). A phylogeny of the flowering plant family Apiaceae based on chloroplast DNA *rpl16* and *rpo*C1 intron sequences: towards a suprageneric classification of subfamily Apioideae. *American Journal of Botany*, 87(2), 273-292.

20. Downie, SR., Katz-Downie, DS., and Cho, K. J. (1996). Phylogenetic analysis of Apiaceae subfamily Apioideae using nucleotide sequences from the chloroplast *rpo*C1 intron. *Molecular Phylogenetics and Evolution* 6(1), 1-18.

21. Downie, SR., and Jansen, RK. (2015). A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. *Systematic Botany* 40(1), 336-351.

22. Spooner, DM., Ruess, H., Iorizzo, M., Senalik, D., and Simon, P. (2017). Entire plastid phylogeny of the carrot genus (Daucus, Apiaceae): Concordance with nuclear data and mitochondrial and nuclear DNA insertions to the plastid. *American Journal of Botany* 104(2), 296-312.

23. Wen, J., Xie, DF., Price, M., Ren, T., Deng, YQ., Gui, LJ., ... and He, XJ. (2021). Backbone phylogeny and evolution of Apioideae (Apiaceae): New insights from phylogenomic analyses of plastome data. *Molecular Phylogenetics and Evolution* 161, 107183.

24. Allen, GC., Flores-Vergara, MA., Krasynanski, S., Kumar, S., and Thompson, WF. (2006). A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nature Protocols* 1(5), 2320-2325.

25. Kim, K., Lee, SC., Lee, J., Lee, HO., Joh, HJ., Kim, NH., ... and Yang, TJ. (2015). Comprehensive survey of genetic diversity in chloroplast genomes and 45S nrDNAs within Panax ginseng species. *PloS one* 10(6), e0117159.

26. Kim, K., Lee, SC., Lee, J., Yu, Y., Yang, K., Choi, BS., ... and Yang, TJ. (2015). Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of Oryza AA genome species. *Scientific Reports* 5(1), 1-13.

27. Kurtz, S., Phillippy, A., Delcher, AL., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, SL. (2004). Versatile and open software for comparing large genomes. *Genome Biology* 5(2), 1-9.

28. Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, ES., Fischer, A., Bock, R., and Greiner, S. (2017). GeSeq–versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* 45(W1), W6-W11.

29. Carver, T., Simon, RH., Matthew, B., Julian, P., and Jacqueline AM. (2012). Artemis: An Integrated Platform for Visualization and Analysis of High-Throughput Sequence-Based Experimental Data. *Bioinformatics* 28,(4), 464-69.

30. Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW

(OGDRAW) version 1.3. 1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* 47(W1), W59-W64.

31. Lee, HO., Joh, HJ., Kim, K., Lee, SC., Kim, NH., Park, JY., ... and Yang, TJ. (2019). Dynamic chloroplast genome rearrangement and DNA barcoding for three apiaceae species known as the medicinal herb "Bang-Poong". *International Journal of Molecular Sciences* 20(9), 2196.

32. Katoh, K., and Standley, DM. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30(4), 772-780.

33. Frazer, KA., Pachter, L., Poliakov, A., Rubin, EM., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Research*, 32(suppl_2), W273-W279.

34. Hall, T., Biosciences, I., and Carlsbad, C. (2011). BioEdit: an important software for molecular biology. *GERF Bull Biosci* 2(1), 60-61.

35. Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, JC., Guirao-Rico, S., Librado, P., Ramos-Onsins, SE., and Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution* 34(12), 3299-3302.

36. Darriba, D., Taboada, GL., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9(8), 772-772.

37. Silvestro, D., and Michalak, I. (2012). raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity & Evolution* 12(4), 335-337.

38. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood.

Molecular Biology and Evolution 24(8), 1586-1591.

39. Rambaut, A., Drummond, AJ., Xie, D., Baele, G., and Suchard, MA. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* 67(5), 901.

40. Rambaut, A. (2012). FigTree v1. 4.

41. Banasiak, Ł., Piwczyński, M., Uliński, T., Downie, SR., Watson, MF., Shakya, B., and Spalik, K. (2013). Dispersal patterns in space and time: a case study of Apiaceae subfamily Apioideae. *Journal of Biogeography* 40(7), 1324-1335.

42. Wen, J., Yu, Y., Xie, DF., Peng, C., Liu, Q., Zhou, SD., and He, XJ. (2020). A transcriptome-based study on the phylogeny and evolution of the taxonomically controversial subfamily Apioideae (Apiaceae). *Annals of Botany* 125(6), 937-953.

43. Pimenov, MG., Tatiana, AO., Galina VD,, and Tahir, HS. (2016). Sillaphyton, a new genus of the Umbelliferae, endemic to the Korean Peninsula. *Botanica Pacifica: a Journal of Plant Science and Conservation* 5(2), 31-41.

44. Favre, A., Päckert, M., Pauls, SU., Jähnig, SC., Uhl, D., Michalak, I., and Muellner-Riehl, AN. (2015). The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biological Reviews* 90(1), 236-253.

45. Plunkett, GM., and Downie, SR. (2000). Expansion and contraction of the chloroplast inverted repeat in Apiaceae subfamily Apioideae. *Systematic Botany* 25(4), 648-667.

46. Goulding, SE., Wolfe, KH., Olmstead, RG., and Morden, CW. (1996). Ebb and flow of the chloroplast inverted repeat. *Molecular and General Genetics MGG* 252(1), 195-206.

47. Zhu, A., Guo, W., Gupta, S., Fan, W., and Mower, JP. (2016). Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytologist* 209(4), 1747-1756.

CHAPTER II

Dynamic plastome evolution among *Peucedanum japonicum* accessions distributed in South Korea

ABSTRACT

Peucedanum japonicum is a plant species distributed in the East Asian region, used for oriental medicine primary as antifebrile. Our previous study demonstrated an identical plastome size on two *Peucedanum japonicum* accessions (164,653 bp) without any intraspecific polymorphism. However, comparative plastome analysis in Chapter 1 has revealed two kinds of plastomes with different inverted repeat length, the long (L) and short (S) types of IR region. In this chapter, the diversity of intraspecific polymorphism of *P. japonicum*, at population level was explored. Total 38 germplasm distributed in different region around South Korea were collected and low-coverage genome sequencing was conducted to on nine accessions, five of the L-type and four of the S-type. Plastome based phylogenetic analysis showed clear clusterization of each accession onto L and S-types, while those of 45S nrDNA sequence formed monophyletic group regardless of plastome types. Furthermore, the separation on the S-type *P. japonicum* accession; Pj-8 before L and S-types diverged, suggests the L-type divergence from the S-type. The genetic diversities among *P. japonicum* are either depended on the type of *P. japonicum*. Upon three molecular markers (The L and S-type differentiation, 17 InDels, and 8 KASPs) revealed the L-type possesses low intraspecific polymorphism while the S-type has higher diversity. SNP-based KASP marker analysis distinguished both L and S type into sub-groups, illustrated broader genetic diversity on *P. japonicum* germplasms. The intraspecific diversity on *P. japonicum* would benefit the selection and management of the best traits within *P.*

japonicum during breeding experiments.

Keywords: Genetic diversity, Intra-specific polymorphism, *Peucedanum japonicum*

INTRODUCTION

Peucedanum japonicum Thunberg is a member of the *Peucedanum* genus under the Selineae tribe of the Apioideae. Its habitat is near the coastal region around the cliff area, distributed in the East Asian region of Korea, Japan, China, Taiwan, and the Philippines [1]. Although like most other *Peucedanum* species, *P. japonicum* are rich in coumarin components, its root has been used as traditional oriental medicine to treat coughs, colds, headaches, antifebrile, anti-obesity, anti-inflammatory, and anti-cancer [2~7]. Lately, as the interest in the healthy product has become popular, the aerial part of the *P. japonicum* such as leaves and flowers, is used to manufacture herbal tea, diet product, and vegetable food [8].

Plastid (or chloroplast) is a key organelle of photosynthesis and other biomechanical pathways in the plants converting light into an essential energy source for a plant to survive. Plastid genome (plastome) located within the cytoplasm is highly conserved from uni-parental inheritance feature and a circular structure composed of a quadripartite structure of the two single-copy regions and two inverted repeat regions. The plastome size is smaller than other organellar genomes such as mitochondria, with sizes ranging between 120~217 kbp [9~10] in most Angiosperm. The conserved structure of the plastome provides a moderate substitution rate and low intra-species polymorphism compared to inter-species polymorphism and less heteroplasmy and recombination than other genomes. The conserve feature of plastome provides a source for comparative analysis in plant species [11].

The study by Lee and Joh et al., [12] have reported a complete assembly on two *P. japonicum* plastome with a genome size of 164,653 bp without any intraspecific polymorphism. The study commented that an identical plastome between two *P. japonicum* accession resulted from the same germplasm of *P. japonicum*. Considering the low genetic diversity from the cultivated *P. japonicum* population from the previous study, the *P. japonicum* from the wild distributed plant would provide higher genetic diversity was hypothesized. Plastome assembly on the collections of *P. japonicum* germplasms from different distributed regions in Korea in Chapter 1 revealed a genetic variation within *P. japonicum* plastome with two different plastomes (164 kbp and 147 kbp) by the changes in IR junction. From the 17 kbp difference within inverted repeat regions, have renounced *P. japonicum* as Long (L) and Short (S) type depending on the length of IR regions.

To further understand the broadness of the genetic diversity within *P. japonicum*, nine *P. japonicum* plastomes, including three newly assembled *P. japonicum* accessions, were analyzed. Furthermore, the molecular marker developed in this study was applied to the population of 38 collected *P. japonicum* from various regions around Korea. The finding from this study will be helpful to answer the question regarding the broad genetic diversity of *P. japonicum* and provide a valuable marker for future experiments.

METERIALS AND METHODS

Taxon sampling and DNA extraction

A total of 38 *P. japonicum,* both wild and cultivated landrace plants, were collected in Korea, and the location of the samples are as follows: Geumodo Island, Wando region, Hamyang region, Goheung region, Ulleung island, Keuneong region; Oedolgae region; Seopjikoji beach in Jeju island, Marado island, Haenam region, Oedo island, Bakripo beach in the west coast region, Hantaek Botanical garden in Yong-In Kyunggi-do region and a Pharmaceutical Herb garden in Goyang Kyunggi-do region operated by Seoul National University (SNU); College of Pharmacy. The genomic DNA of each sample was extracted using the modified cetyltrimethylammonium bromide (CTAB) method [13] and Exgene Plant SV Midi Kit (Geneall Biotechnology, Seoul) following the manufacturer's protocol. In addition, the quantity and quality of genomic DNAs were examined using both Nanodrop 1000 spectrometer (Thermo Fisher Scientific, USA) and Ethidium Bromide (EtBr) stained agarose gel electrophoresis.

Plastome and 45S rDNA nucleotide sequencing, assembly, and annotation

Total nine *P. japonicum* accession were used in this study. Among nine accessions, two accessions: Pj-1 (NC_034644) and Pj-2 (KU866531) were downloaded from NCBI Genbank, and four accessions: Pj-3 (Geumodo Island), Pj-4 (Hantaek Botanical

Garden), Pj-6 (Jeju Island), Pj-7 (Wando Region) were used in Chapter 1. For this study, three *P. japonicum* accessions from the Keuneong from Jeju island (Pj-5), SNU pharmaceutical herb garden (Pj-8), and Haenam region (Pj-9) were selected for plastome assembly. Complete plastome and 45S nrDNA sequences were generated by de novo assembly of the low coverage whole genome sequence (dnaLCW) from Phyzen Genome Institution (http://phyzen.com); described more detail in the study by Kim et al. [14. 15]. In brief, trimmed high-quality reads with Phred scores of 20 or more were obtained from the total paired-end (PE) raw reads using the CLC-quality trim tool and then were assembled by a CLC genome assembler (ver. 4.06 beta, CLC Inc, Rarhus, Denmark). The untapped initial contigs representing the plastome were selected from the total contigs using MUMmer [16], by comparison with the *P. japonicum* plastid genome sequences (NC_034644, KU866531.1) from the previous study by Lee and Joh et al., as a reference sequence [12]. The representative contigs were arranged based on the reference sequences and merged into a single draft sequence by connecting overlapping terminal sequences. Detected assembly errors and gaps in the draft sequence were manually corrected by mapping the raw PE reads. Genes of the plastome were annotated using the GeSeq program (https://chlorobox.mpimp-golm.mpg.de/geseq.html) [17] and manually curated based on the reference *P. japonicum* sequence from the NCBI Genbank database [12]. Circular maps of assembled plastomes were drawn using OGDRAW (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) [18].

For 45S nrDNA assembly, the longest initial contigs, including the 45S cistron unit, were selected from the total contigs by comparison with reported 45S nrDNA

sequences of *P. japonicum* (KX757776.1) from the NCBI Genbank database [12]. In addition, the structures of 45S nrDNA sequences were predicted by both RNAmmer [19] and BLASTN (http://www.ncbi.nlm.nih.gov/BLAST/) [20] searches.

Table 2-1. Summary of 38 *P. japonicum* germplasms used in this study

Germplasm No.	Collected Location	Habitats	Assembled plastome	Germplasm No.	Collected Location	Habitats	Assembled plastome
1	Geumodo region	Cultivated	Pj-3	20	Ulleung island	Wild	
2	SNU Pharmaceutical herb garden	Cultivated		21	SNU Pharmaceutical herb garden	Cultivated	
3	Hantaek botanical garden	Cultivated		22	Hantaek botanical garden	Cultivated	Pj-4
4	Hamyang region	Cultivated		23	Hantaek botanical garden	Cultivated	
5	Goheung region	Cultivated		24	Jeju island	Wild	Pj-6
6	Hantaek botanical garden	Cultivated		25	Haenam region	Wild	Pj-9
7	Keuneong region in Jeju island	Wild	Pj-5	26	SNU Pharmaceutical herb garden	Cultivated	
8	Oedolgae region in Jeju island	Wild		27	Wando region	Wild	Pj-7
9	Ulleung island	Wild		28	Hantaek botanical garden	Cultivated	
10	Ulleung island	Wild		29	SNU Pharmaceutical herb garden	Cultivated	
11	Ulleung island	Wild		30	Bakripo beach	Wild	
12	Ulleung island	Wild		31	Oedo island	Wild	
13	Ulleung island	Wild		32	SNU Pharmaceutical herb garden	Cultivated	
14	Ulleung island	Wild		33	Seopjikoji beach in Jeju island	Wild	
15	Ulleung island	Wild		34	Marado island	Wild	
16	Ulleung island	Wild		35	SNU Pharmaceutical herb garden	Cultivated	Pj-8
17	Ulleung island	Wild		36	SNU Pharmaceutical herb garden	Cultivated	
18	Ulleung island	Wild		37	Hantaek botanical garden	Cultivated	
19	Ulleung island	Wild		38	Bakripo beach	Wild	

Table 2-2. Summary of quality and quantity of the generated sequences data of *Peucedanum japonicum*

Location	Type	IDs	Raw data	Chloroplast genome			45S nrDNA		
			amounts (Gbp)	Length (bp)	Coverage (x) [d]	Sequence ID	Length (bp)[e]	Coverage (x) [d]	Sequence ID
Rural Development Administration	Cultivated	Pj-1	2.08	164,653	369.41	NC_034644	5,815	943.43	KX757776
Korea Food & Drug Administration	Cultivated	Pj-2	5.18	164,653	945.25	KU866531	5,815	2,418.35	KX757777
Geumodo Island	Cultivated	Pj-3	1.18	164,652	113.7	IM150601-18	5,815	306.23	IM150601-18nr
Hantaek Botanical Garden	Cultivated	Pj-4	40.96	164,656	527.4	IM200128-1	5,815	1,195.07	IM200128-1nr
Keuneong (Jeju)	Wild	Pj-5	1.36	164,660	76.69	IM201104-4	5,815	553.32	IM201104-4nr
Jeju Island	Wild	Pj-6	1.35	147,525	236.81	IM150921-6	5,815	432.06	IM150921-6nr
Wando Region	Wild	Pj-7	1.06	147,549	75.9	IM150601-19	5,815	124.50	IM150601-19nr
SNU Pharmaceutical herb garden	Cultivated	Pj-8	1.59	147,592	144.98	IM201104-5	5,815	186.49	IM201104-5nr
Haenam region	Wild	Pj-9	1.59	147,578	144.24	IM201104-6	5,815	619.46	IM201104-6nr

Plastome and 45S rDNA nucleotide variations

Each completed plastid genome and 45S nrDNA sequences were aligned to find intraspecific variation using the web-based program MAFFT (http://mafft.cbrc.jp/alignment/server) [21]. Sequence variations were visualized through mVISTA (http://genome.lbl.gov/vista/mvista/submit.shtml) [22] under Shuffle LAGAN, and BlastZ tool (http://phyzen.iptime.org/tools/cv.php). Few misaligned regions were manually curated using the BioEdit program [23]. Afterward, polymorphic regions such as single nucleotide polymorphisms (SNPs) and insertion/deletions (InDels) polymorphisms were investigated at the intra-species level.

Seeking intraspecific diversity among *P. japonicum* accessions through comparative analysis

The complete plastome assembly among nine *P. japonicum* accessions demonstrated differences with the plastome size related to the IRs region of plastome structure. For this study, an experiment using the molecular marker IR01 provided from the previous study [12] was designed to investigate the diversity of the *P. japonicum* population related to the IR regions of the plastome structure. The molecular marker IR01 comprises three primers; one pair of control primers and one LSC primer amplifying the LSC/IR junction on *P. japonicum* L-type [12]. Since the marker is developed based on L-type *P. japonicum*, and the plastome size of the S-type *P. japonicum* is similar to *L. seseloides* and *G. littoralis*, the IR01 marker was capable of distinguishing *P.*

japonicum L and S-type. The PCR conditions are programmed as follows: 94°C for 5 min; followed by 35 cycles of 94°C for 1 min 30 sec, 50°C for 1 min 30 sec, 72°C for 1 min 30 sec; and final extension at 72°C for 7 min [12]. For the accuracy of each marker, each markers were tested using three *P. japonicum* accessions (Pj-3, Pj-6, and Pj-7).

Table 2-3. Marker information from previous study regarding plastome structure variation within Inverted Repeat region

Marker ID	Primer sequence (5'- 3')	Product size Pj-1 / Pj-2 / Pj-3 (bp)	Location
IR01	LSC P: CCTAGCTGCTGTTGAAGCTC Control F: GACGACTGAGCCAACTTGAT Control R: TCGAGACGTTCTTCAAACCA	576 / na / na 262 / 262/ 262	psbA *psbH* - *petB*

*na refers to as none available band type.

Molecular marker analysis

To validate intra-species variation, 17 molecular markers designed from InDel polymorphic region detected among nine accessions using the Primer 3 program (http://bioinfo.ut.ee/primer3-0.4.0/) were tested on 38 collected germplasms [24]. PCR was implemented using genomic DNAs of three accessions (Pj-3, Pj-6, Pj-7) from Geumodo, Jeju and Wando regions as a control marker, respectively (Table 2-2). PCR conditions for molecular markers were as follows: Initial denaturation at 94°C for 5 min, followed by 35 cycles of 94°C for 1 min 30 sec, 50 ~ 58°C for 1 min 30 sec, 72°C for 1 min 30 sec; and final extension at 72°C for 7 min. The amplified PCR fragments were analyzed through agarose gel electrophoresis and capillary electrophoresis using a Fragment analyzer (Advanced Analytical Technologies Inc., USA) following the manufacturer's instructions. To understand clustered patterns of the genetic relationship among *P. japonicum* population, cluster analysis of UPGMA (Unweighted Pair Group Method with Arithmetic Mean) method from NTSYS 2.10 program was used to view the relationship among populations [25]. In addition, the dendrogram was generated using the Sequential agglomerative hierarchical non-overlapping (SAHN) clustering program provided within NTSYS 2.10.

KASP marker analysis

Eight KASP (Kompetitive allele-specific PCR) markers were developed for a faster and more accurate intraspecific genotype evaluation on *P. japonicum* accessions. To design the KASP marker for genotyping, approximately 50 bp of sequences upstream and

downstream of the SNP locus from *rpoA*, *ycf4*, *psaA*, *psbK*, *rpl2*, *rpo*C2, *rps15*, *ycf1*, *ycf4* gene were analyzed. For the experiment with KASP markers, 5 ul of DNA was dispensed to the 96 wells of the microtiter plate, and the other two wells were filled with sterilized distilled water (SDW) as a negative control (NTC) [26]. First, the KASP Master mix and KASP Assay mix were dissolved thoroughly and vortexed. Then the mixture of 5 ul of Master mix and 0.14 ul of Assay mix was prepared and dispensed to each well. In other words, 96-well plates have a total reaction volume of 10 ul. Finally, Genotyping was performed by reading fluorescence resonance energy transfer (FRET) using qPCR instruments, Roche LC480 (Roche Diagnostics, Penzberg, Germany). The thermal cycling conditions are as follows: 15 min at 94°C, 10 cycles of 20 sec at 94°C and 60 sec at 66~55°C (drop 0.6 per cycle), and 26 cycles of 20 sec at 94°C and 60 sec at 55°C. The results were then plotted with measured FAM and HEX values to check if the clusters were well constructed. To see better results, the Recycling protocol recommended by LGC genomics was employed (LGC Genomics, Hoddesdon, UK) with the condition of 3 cycles of 20 sec at 94°C and 60 sec at 57°C and 1 cycle of 60 sec at 30°C. After the cycling, each fluorescence read was taken for KASP genotyping. To understand clustered patterns of the genetic relationship among *P. japonicum* population, cluster analysis of UPGMA (Unweighted Pair Group Method with Arithmetic Mean) method from NTSYS 2.10 program was used to view the relationship among populations [25]. The dendrogram was generated using Sequential agglomerative hierarchical non-overlapping (SAHN) clustering program provided within NTSYS 2.10.

Phylogenetic analyses

To better understand the uniqueness of the plastid genome on *P. japonicum*, phylogenetic analysis and the species under Apioideae with a close relationship to *P. japonicum* species from Chapter I was executed. Nine *P. japonicum* accessions (NCBI, Chapter 1, three newly assembled sequences) were used, and the *S. montanum* was selected as an outgroup. Besides the newly sequenced Apiodeae plastome *D. takeshimana* ID: IM180813-49, sequence information regarding *S. montanum;* ID: NC_027451.1, *L. seseloides*; ID: NC_034643, KT153021, and *P. praeruptorum*; ID: MN016968 was downloaded from NCBI Genbank database (https://www.ncbi.nlm.nih.gov/nucleotide). To understand the intraspecific genetic variation under *P. japonicum* accessions, a phylogenetic analysis on the 45S nrDNA sequence was implemented. Complete 45S citron sequence information on *L. seseloides* (KX757774, KX757775) was obtained from the previous study [12], and *D. takeshimana* was newly assembled in Chapter 1 were used. Each CDS and nrDNA sequence were gathered into a single sequence file and aligned using MAFFT web-based program [21]. Gblock web-based program was used on aligned sequences to increase the quality and accuracy on a default setting [27, 28]. Maximum likelihood phylogenetic analysis was implemented using raxmlGUI version 2.0 program [29]. Before running the analysis, the sequence was first checked with jModeler 2.1.10 to select the optimal model selection. Then, the GTR+G+I nucleotide substitution model was applied with ML estimation of the base frequencies with 1000 replicated ML bootstrap analysis was conducted to test the support of the branches. The phylogenetic tree was visualized with Figtree 1.4 [30].

RESULTS

Plastid, 45S nrDNA genomes sequencing, assembly, and annotation

Nine *P. japonicum* accessions demonstrated two types of plastome sizes of 164 kbp and 147 kbp with 17 kbp differences. *P. japonicum* accession Pj-1, Pj-2, Pj-3, Pj-4, Pj-5 exhibited practically constant plastome size of 164.4 kbp, with minor 8 bp difference (164,652 bp to 164,660 bp) (Figure 2-1, Table 2-2), whereas *P. japonicum* accession Pj-6, Pj-7, Pj-8, Pj-9 exhibits plastome size of 147.5 kbp with 67 bp differences (147,525 bp to 147,592 bp) (Figure 2-1, Table 2-2). Each *P. japonicum* accession presented a typical quadruple genomic structure, including two single-copy and two IR regions. However, the range of quadruple structure of nine *P. japonicum* accessions varies in each region. Large single-copy region presented with the range of 75,584~92,804 bp; small single-copy region presented with the range 17,551~17,628 bp; inverted repeat region presented with the range 18,760~35,760 bp (Figure 2-1, Table 2-4). For convenience, *P. japonicum* species with the plastome size of 164 kbp was renounced as the L-type and 147 kbp as the S-type. Shift within LSC/IR junction (Junction of LSC-IRb: JLB and Junction of LSC-IRa: JLA) between *P. japonicum* accessions affected the length and gene contents within LSC and IR region; however, the boundary of IRb/SSC, SSC/IRa, otherwise known as JSB/JSA junctions remained unchanged for all *P. japonicum* samples (JSB: *ycf1-ndhF* / JSA: *rps15-ycf1*).

The 45S nrDNA sequences on nine *P. japonicum* accessions were assembled into a single contig consisting of a complete transcribed unit (45S nrDNA unit) and partial IGS. This study could not assemble the complete length of IGS due to the nucleotide gaps at GC-rich regions as described in the previous study [12]. The component of 45S nrDNA units' was identical among nine accessions (Table 2-5). Each unit were 5,815 bp in length, consist of 18S rRNA (1,809 bp), 5.8S rRNA (163 bp), 26S rRNA (3,394 bp), ITS1 (216 bp), ITS2 (233 bp) (Table 2-5).

Table 2-4. Summary of full plastome assemble information on nine *P. japonicum* accessions

ID	Chloroplast genome									Location	GenBank acc. no. (Accession No.)
	Length (bp)	LSC (bp)	SSC (bp)	IRs (bp)	Gene	Protein	tRNA	rRNA	G+C %		
Pj-1	164,653	75,584	17,551	35,759	141	96	37	8	37.46	Rural Development Administration	NC_034644
Pj-2	164,653	75,584	17,551	35,759	141	96	37	8	37.46	Korea Food & Drug Administration	KU866531
Pj-3	164,652	75,583	17,551	35,759	141	96	37	8	37.45	Geumodo Island	IM150601-18
Pj-4	164,656	75,585	17,551	35,760	141	96	37	8	37.45	Hantaek Botanical Garden	IM200128-1
Pj-5	164,660	75,588	17,552	35,760	141	96	37	8	37.45	Keuneong (Jeju)	IM201104-4
Pj-6	147,525	92,563	17,628	18,667	126	82	36	8	37.54	Jeju Island	IM150921-6
Pj-7	147,549	92,631	17,606	18,656	126	82	36	8	37.54	Wando Region	IM150601-19
Pj-8	147,592	92,804	17,576	18,606	126	82	36	8	37.55	SNU Pharmaceutical herb garden	IM201104-5
Pj-9	147,578	92,655	17,605	18,659	126	82	36	8	37.54	Haenam region	IM201104-6

Table 2-5. 45s nrDNA sequence information on nine *P. japonicum* accessions

Species	IDs	Length (bp)	18s	ITS1	5.8s	ITS2	26s	GenBank acc. no. (Accession No.)
Peucedanum japonicum	Pj-1nr	5,815	1809	216	163	233	3394	KX757776
	Pj-2nr	5,815	1809	216	163	233	3394	KX757777
	Pj-3nr	5,815	1809	216	163	233	3394	IM150601-18nr
	Pj-4nr	5,815	1809	216	163	233	3394	IM200128-1nr
	Pj-5nr	5,815	1809	216	163	233	3394	IM201104-4nr
	Pj-6nr	5,815	1809	216	163	233	3394	IM150921-6nr
	Pj-7nr	5,815	1809	216	163	233	3394	IM150601-19nr
	Pj-8nr	5,815	1809	216	163	233	3394	IM201104-5nr
	Pj-9nr	5,815	1809	216	163	233	3394	IM201104-6nr

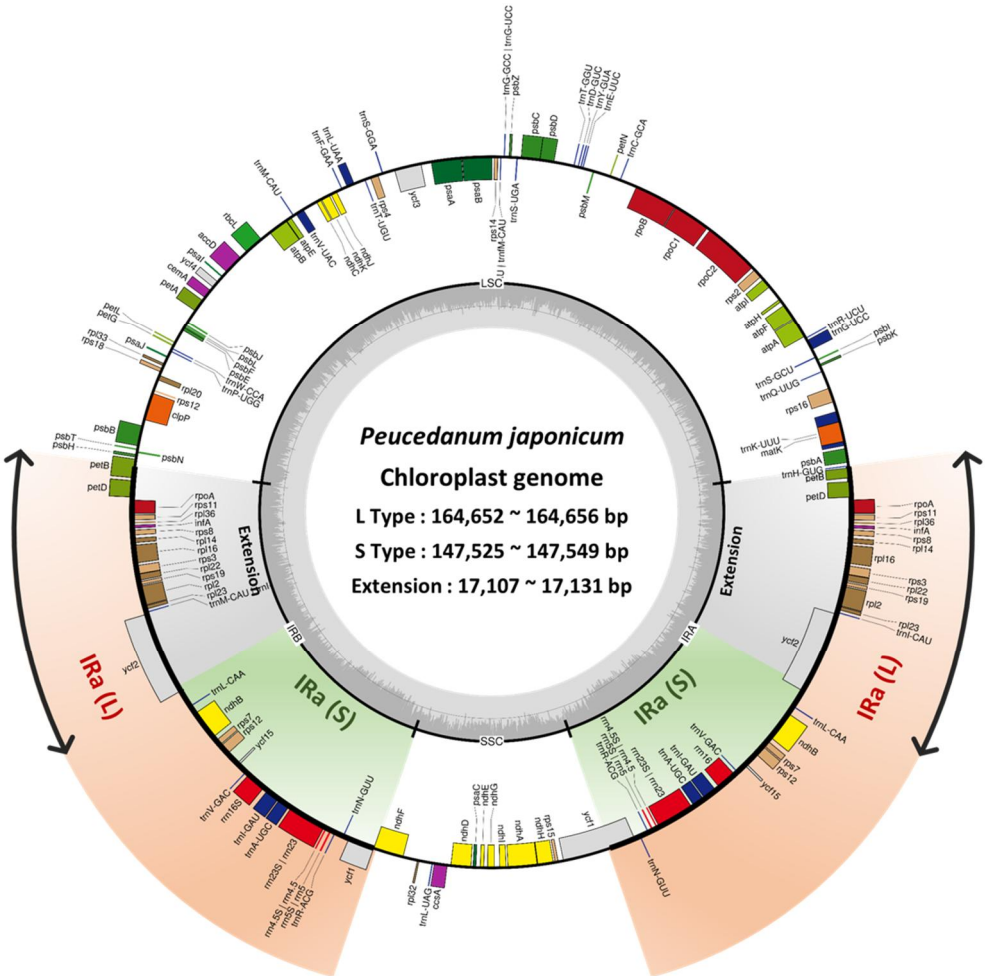


Figure 2-1. The plastome map of *P. japonicum* Long and Short types. The following plastome map illustrated the difference between L and S-types through the colored region on the plastome indicate the inverted repeat regions. The red shade indicates the IR region of the L-type. Green shade indicates the IR region of the S-type. Gray shade indicates the difference in the IR region between L and S-types. The plastome map was generated using OGDRAW. Genes transcribed clockwise and counterclockwise are indicated on the outside and inside of the large circle, respectively.

Distribution of variants along with *P. japonicum* plastomes sequences

The comparison of plastome sequence on *P. japonicum* presented a total of 388 intraspecies polymorphisms comprised of 73 InDels with polymorphic differences and 315 SNPs among nine *P. japonicum* accessions (Table 2-6, Table S2-1). The InDels and SNP located within the IRa region are excluded for an accurate evaluation.

The majority of the 73 intraspecific InDels are found within intergenic regions. Among them, 18 InDels comprise simple sequence insertion, and deletion ranged from 3 bp to 36 bp, mainly in the intergenic regions (Table S2-1). The largest of the InDel is the 36 bp insertion in between the JLA/*trn*H-GUG intergenic region, which emerges in the Pj-8 accession. The other InDels were composed of repeat sequences, comprised of 51 tandem repeats and four simple repeats. InDel size ranges from 5 bp to 66 bp, with repeat motifs ranging from 5 bp to 22 bp. Three tandem repeats were located within the *ycf2*, 23S ribosomal, and *ycf1* genic regions (Table S2-1b). Amongst the three, the largest InDel is located in the *ycf2* gene with a tandem repeat motif of 18 bp. The following list of accessions represents the same number and order of tandem repeats; Pj-3, Pj-4, Pj-5, Pj-7 accessions, and Pj-6, Pj-9 accessions presented similar tandem repeat orders. Pj-8 accession, on the other hand, presented a very different tandem repeat order than all other accessions. Three tandem repeats have been discovered only within Pj-8 accessions. The largest InDel is located in the *trn*N-GUU-*ycf1* intergenic region with a 22 bp tandem repeat motif, specific to the L-type of *P. japonicum*. Apart from the InDel size and motif ranges, Pj-6 demonstrated a 10 bp sequence insert between tandem repeat motifs on *ndhE-ndhG* intergenic region. The 10 bp inserted

sequence is part of the tandem repeat (TR) motif, suggesting a slip and slide mechanism triggered the TR difference between the other S-type accessions. The four small repeated sequences comprised of 'AT' or 'TA' combinations were discovered within the *trn*R-UCU - *atpA*, *atpF* - *atpH*, *rps4* - *trn*T-UGU, and *ndhC* - *trn*V-UAC intergenic regions. Each repeat specifies S-type accessions. Of the total 73 intraspecific InDels that discovered, a large portion is specific to the S-type *P. japonicum*. These variations represent a difference in the genomic size between S-type *P. japonicum* accessions (Pj-6: 147,525 bp, Pj-7: 147,549 bp, Pj-8: 147,592 bp, Pj-9: 147,578 bp). The genome size difference between each L-type *P. japonicum* demonstrates 1~8 bp, caused by 2 bp length of simple repeat and single nucleotide insertion from the mono-polymer repeat region found within LSC and IR regions. On the other hand, S-type represents higher genetic diversity than the L-type.

Total 315 SNPs were discovered under the plastome comparison between *P. japonicum* accessions. Most SNPs are located from the LSC region, followed by the SSC region then the IR region (Figure 2-2, Table S2-2). Comparison between *P. japonicum* accessions, Pj-8 accession represented a far greater SNP than other accessions. Considering the number of SNPs under each accession, about 202 SNPs were discovered within Pj-8 accessions alone from the total number of SNPs. Among discovered SNPs, about 156 SNPs are located within the genic region while the rest of 159 SNPs are located within the intergenic region (Figure 2-2, Table 2-6, Table S2-2). The 156 SNPs are discovered within 48 plastid genes (Figure 2-2, Table S2-2). Within SNPs from the genic region, about 56 SNPs have been validated as synonymous

substitution, while 99 SNPs have been validated as non-synonymous substitution (Figure 2-2, Table S2-2). The following SNPs discovered from certain plastid gene is specific to the *P. japonicum* accessions. The SNPs discovered from *rpl2* exon 2 and *rps19* plastid gene is specific to Pj-6. The SNPs discovered from *ndhG* and *rps4* plastid gene is specific to Pj-7. The SNPs discovered from *atpA*, *rpoC1*, *psaB*, *cemA*, *psbE*, *rps18*, *psbT*, *infA*, rps8, *rpl16* exon 2, *rps3*, *rpl23*, *ndhB* exon 2, *ccsA*, *ndhI*, *ndhA* exon 1, and *ndhB* exon 2 plastid gene is specific to Pj-8. The SNP discovered from the *rps7* plastid gene is specific to Pj-6, Pj-7, and Pj-9 accessions. The SNPs discovered from the *ndhJ* plastid genes are specific to separating both L and S-type accessions. Among the SNPs, the unique SNP in the *rpoA* gene presented a non-synonymous substitution with nonsense mutation type. From the closer inspection, the SNP on the *rpoA* gene, when it translates, Pj-1, Pj-2, Pj-3, Pj-4, Pj-5, and Pj-8 *P. japonicum* accessions translates α-amino acid "Serine" into a stop codon. The stop codon on the type of such accessions is located 51 bp closer to the 5' region than the stop codon of the *rpoA* gene on Pj-6, Pj-7, and Pj-9 accessions (Figure 2-3).

Table 2-6. Information on intra-specific plastome variation between *P. japonicum* accessions

No	SNPs	Pj-1 (L)	Pj-2 (L)	Pj-3 (L)	Pj-4 (L)	Pj-5 (L)	Pj-6 (S)	Pj-7 (S)	Pj-8 (S)	Pj-9 (S)
1	Pj-1 (L)	/	0	0	0	9	31	32	119	30
2	Pj-2 (L)	0	/	0	0	9	31	32	119	30
3	Pj-3 (L)	0 (1)	0 (1)	/	0	9	31	32	119	30
4	Pj-4 (L)	0 (3)	0 (3)	0 (4)	/	9	31	32	119	30
5	Pj-5 (L)	2 (7)	2 (7)	2 (8)	2 (4)	/	34	35	119	30
6	Pj-6 (S)	20 (24)	20 (24)	20 (24)	20 (24)	20 (24)	/	4	127	20
7	Pj-7 (S)	17 (23)	17 (23)	17 (23)	17 (23)	17 (23)	12 (30)	/	126	11
8	Pj-8 (S)	52 (83)	52 (83)	52 (83)	52 (83)	52 (83)	55 (89)	55 (87)	/	125
9	Pj-9 (S)	18 (27)	18 (27)	18 (27)	18 (27)	18 (27)	13 (30)	3 (16)	56 (87)	/
	InDels	Pj-1 (L)	Pj-2 (L)	Pj-3 (L)	Pj-4 (L)	Pj-5 (L)	Pj-6 (S)	Pj-7 (S)	Pj-8 (S)	Pj-9 (S)

* Numbers within parenthesis ( ) refers to the number of monopolymer repeat

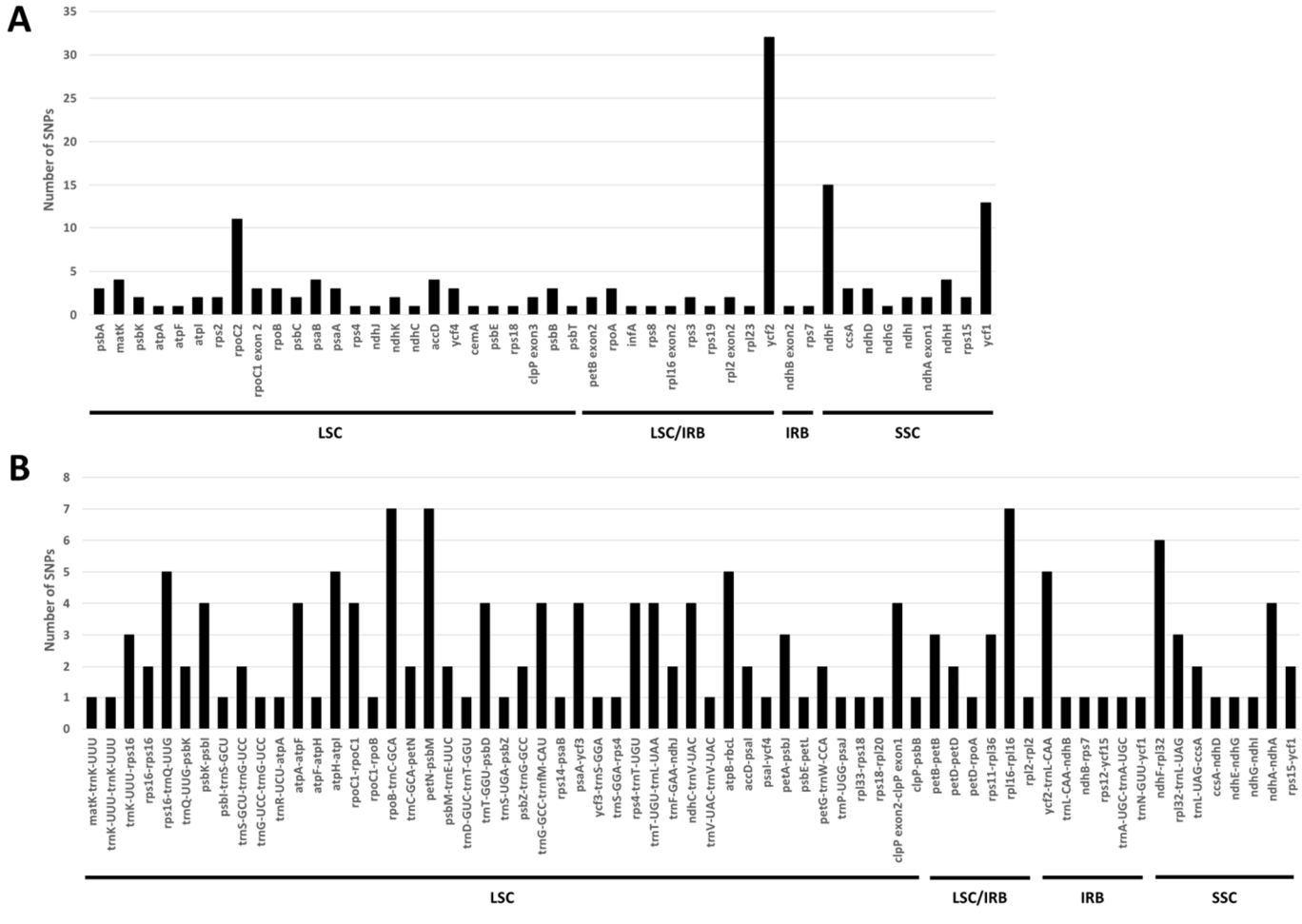

A
Number of SNPs
35
30
25
20
15
10
5
0
psbA
matK
psbK
atpA
atpF
atpI
rps2
rpoC2
rpoC1 exon 2
rpoB
psbC
psaB
psaA
rps4
ndhJ
ndhK
ndhC
accD
ycf4
cemA
psbE
rps18
clpP exon3
psbB
psbT
petB exon2
rpoA
infA
rps8
rpl16 exon2
rps3
rps19
rpl2 exon2
rpl23
ycf2
ndhB exon2
rps7
ndhF
ccsA
ndhD
ndhG
ndhI
ndhA exon1
ndhH
rps15
ycf1
LSC
LSC/IRB
IRB
SSC
B
Number of SNPs
8
7
6
5
4
3
2
1
0
matK-trnK-UUU
trnK-UUU-trnK-UUU
trnK-UUU-rps16
rps16-rps16
rps16-trnQ-UUG
trnQ-UUG-psbK
psbK-psbI
psbI-trnS-GCU
trnS-GCU-trnG-UCC
trnG-UCC-trnG-UCC
trnR-UCU-atpA
atpA-atpF
atpF-atpH
atpH-atpI
rpoC1-rpoC1
rpoC1-rpoB
rpoB-trnC-GCA
trnC-GCA-petN
petN-psbM
psbM-trnE-UUC
trnD-GUC-trnT-GGU
trnT-GGU-psbD
trnS-UGA-psbZ
psbZ-trnG-GCC
trnG-GCC-trnfM-CAU
rps14-psaB
psaA-ycf3
ycf3-trnS-GGA
trnS-GGA-rps4
rps4-trnT-UGU
trnT-UGU-trnL-UAA
trnF-GAA-ndhJ
ndhC-trnV-UAC
trnV-UAC-trnV-UAC
atpB-rbcL
accD-psaI
psaI-ycf4
petA-psbJ
psbE-petL
petG-trnW-CCA
trnP-UGG-psaJ
rpl33-rps18
rps18-rpl20
clpP exon2-clpP exon1
clpP-psbB
petB-petB
petD-petD
petD-rpoA
rps11-rpl36
rpl16-rpl16
rpl2-rpl2
ycf2-trnL-CAA
trnL-CAA-ndhB
ndhB-rps7
rps12-ycf15
trnA-UGC-trnA-UGC
trnN-GUU-ycf1
ndhF-rpl32
rpl32-trnL-UAG
trnL-UAG-ccsA
ccsA-ndhD
ndhE-ndhG
ndhG-ndhI
ndhA-ndhA
rps15-ycf1
LSC
LSC/IRB
IRB
SSC

Figure 2-2. Distribution of intraspecific SNPs across nine *P. japonicum* accessions. A. SNPs located in the CDS region of the plastome B. SNPs located across the intergenic region of the plastome. LSC, SSC, IR refers to the both Large and Small single copy and Inverted repeat region of the plastome. LSC/IRB refers to the region which belong to the LSC region in the S-type while remains in IRB region in the L-type.

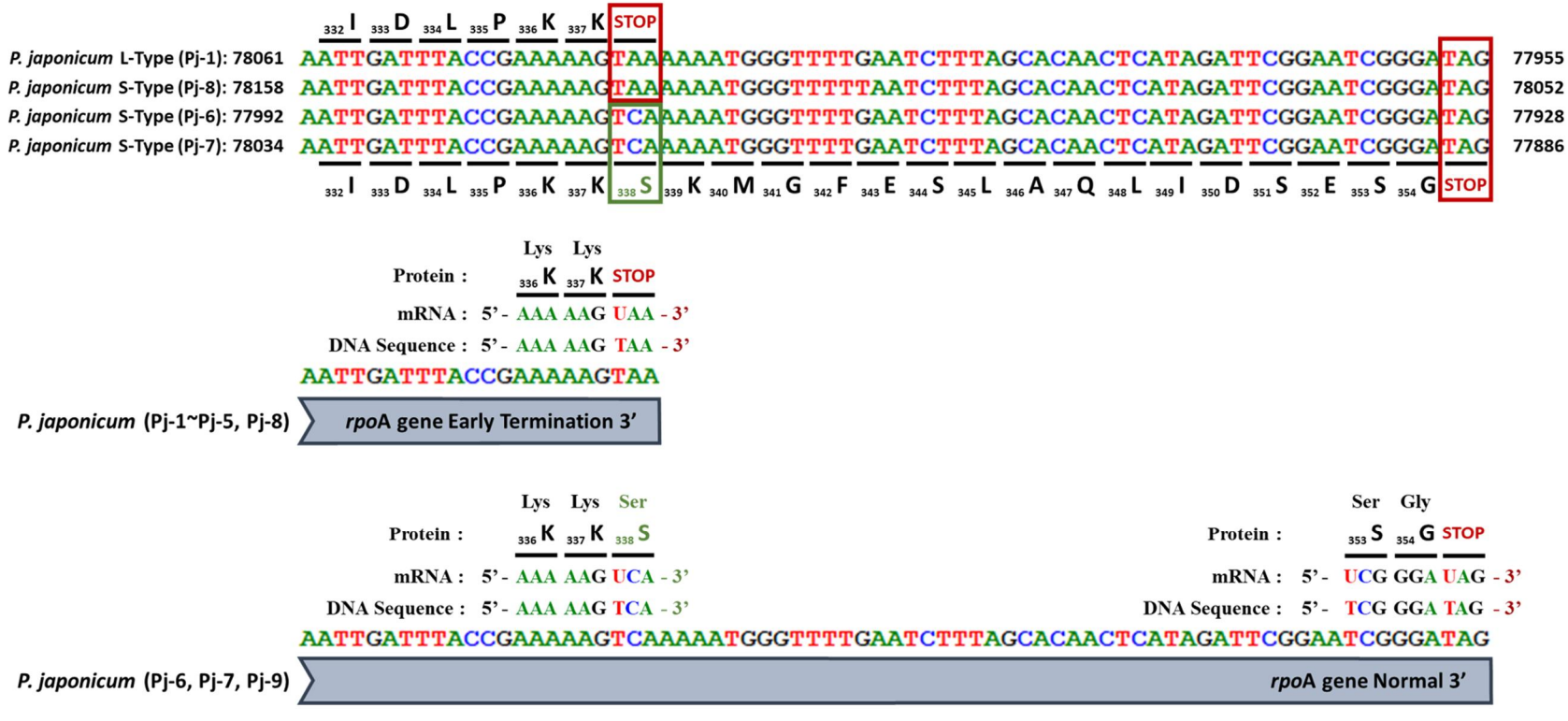


Figure 2-3. Illustration of single SNP causing early termination of *rpoA* gene within *P. japonicum* accessions. The SNP within *rpoA* gene is a non-synonymous substitution with nonsense mutation type caused early termination of a gene that revealed gene size differences between the accessions.

Distribution of variants along with *P. japonicum* 45s nrDNA sequences

The 45S nrDNA sequence was assembled into a single contig with a completed 45S citron unit with the size of 5,815 bp from each *P. japonicum* species. Comparison on each *P. japonicum* nrDNA citron unit presented four SNP polymorphisms discovered at 5.8S, ITS2, and 26S regions (Figure 2-4, Table 2-7). Unlike the plastome, the polymorphisms on the nrDNA sequence presented the intraspecific polymorphism within the L-type *P. japonicum*s that separate from the previous study (Pj-1, Pj-2) and the newly assembled *P. japonicum* accessions (Pj-3, Pj-4, Pj-5) [12]. In addition, the SNPs located in 5.8S and ITS2 are specific to Pj-1, Pj-2, and Pj-8, whereas SNP presented at 26s region presented polymorphism that is specific to the Pj-1, Pj-2, Pj-7 and Pj-8 accessions in 3,802 position and specific to the Pj-3 and Pj-4 accessions in 4,047 position (Figure 2-4, Table 2-7).

Table 2-7. Information of intra-specific 45s nrDNA variation on *P. japonicum* accessions

45S nrDNA Position	Location of SNP	Pj-1 (L)	Pj-2 (L)	Pj-3 (L)	Pj-4 (L)	Pj-5 (L)	Pj-6 (S)	Pj-7 (S)	Pj-8 (S)	Pj-9 (S)
[a] 5.8S	2160	C	C	T	T	T	T	T	C	T
[b] ITS2	2213	A	A	T	T	T	T	T	A	T
[c] 26S	3802	T	T	C	C	C	C	T	T	C
[d] 26S	4047	C	C	T	T	C	C	C	C	C

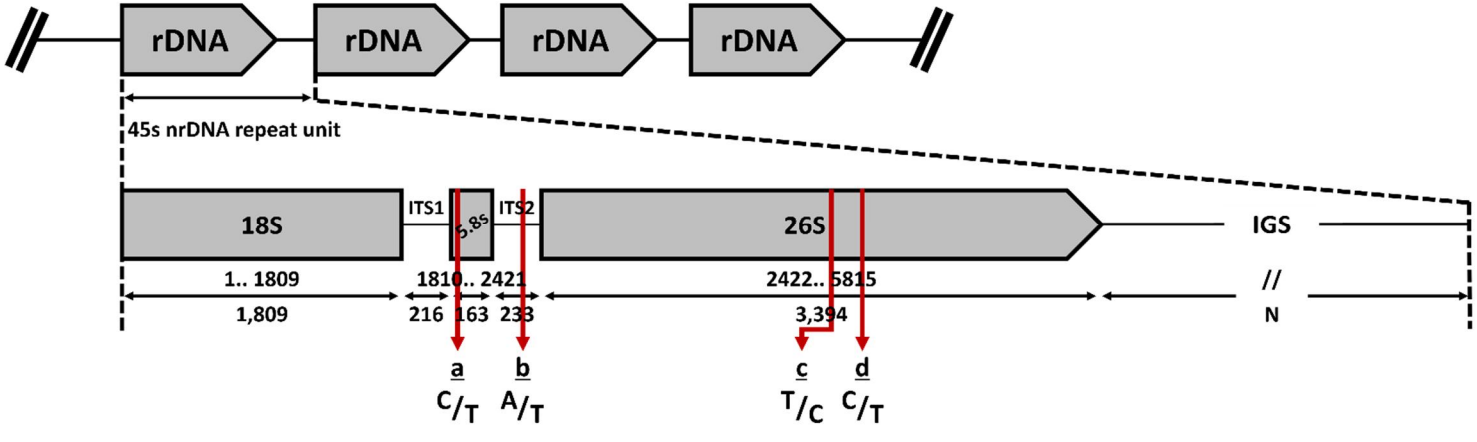


Figure 2-4. The structure of *P. japonicum* intra-specific 45s nrDNA variations. a~d refers to the position of four SNPs located within 45S nrDNA sequence.

Phylogenetic analysis of *P. japonicum* accessions

To understand the relationship between the types of *P. japonicum* and the order of diversification within the accessions. Phylogenetic analysis on the plastid genome of nine *P. japonicum* accessions was executed (Figure 2-5). To ensure the accuracy of the phylogenetic relationship between each accession plastome and 45s nrDNA, closely related species *D. takeshimana*, *L. seseloides,* and *S. montanum* were selected. Selected species are closely emplacing the following topological order of the phylogenetic analysis presented in Chapter I. *P. japonicum* accessions are clearly grouped into a clade with the L and S-types. The L-types clade is very closely related to one another, while within the S-types clade, Pj-7 and Pj-9 show a closer relationship than Pj-6. Among nine *P. japonicum* accessions, Pj-8 presented different topological relationships than other *P. japonicum* accessions. The phylogenetic relationship on Pj-8 accession exposed a speciation event before separating the L and S-types. Such analysis seems to support the differences in the number of the informative site under Pj-8 accessions representing are a far greater number than other accessions.

Phylogenetic relationship regarding nrDNA sequence, clustered and cleaned sequence using MAFFT and Gblock program. Each accession is grouped into four clades comprised of: Pj-1nr; Pj-2nr; Pj-8nr, Pj-7nr, Pj-3nr; Pj-4nr, Pj-5nr; Pj-6nr; Pj-9nr accessions, each clade representing a closer relationship to one another. The phylogenetic relationship under *P. japonicum* accessions based on nrDNA has represented a mixture between L and S-type accessions. For example, the S-type *P. japonicum*; Pj-8 displayed isolated accession from plastome topology, whereas its

nrDNA sequence Pj-8nr represented a clade with Pj-1nr and Pj-2nr of the L-type *P. japonicum* accessions from NCBI Genbank database.

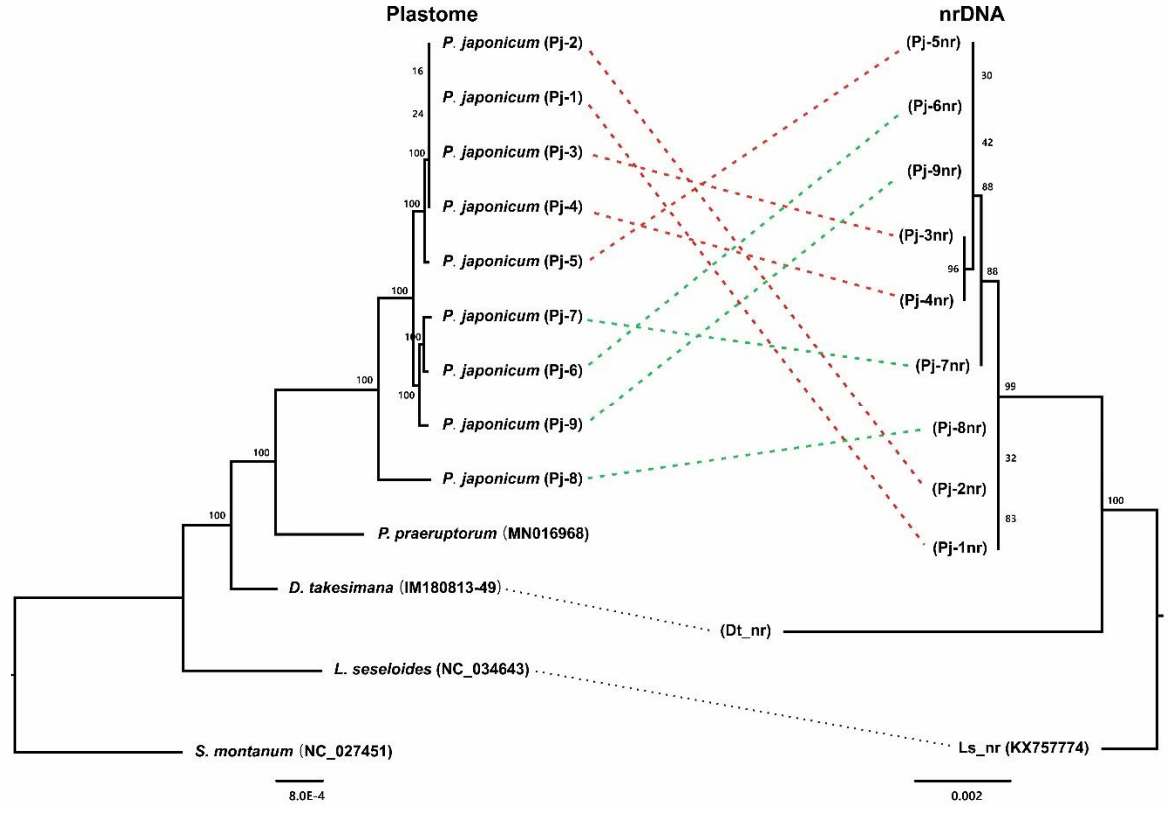
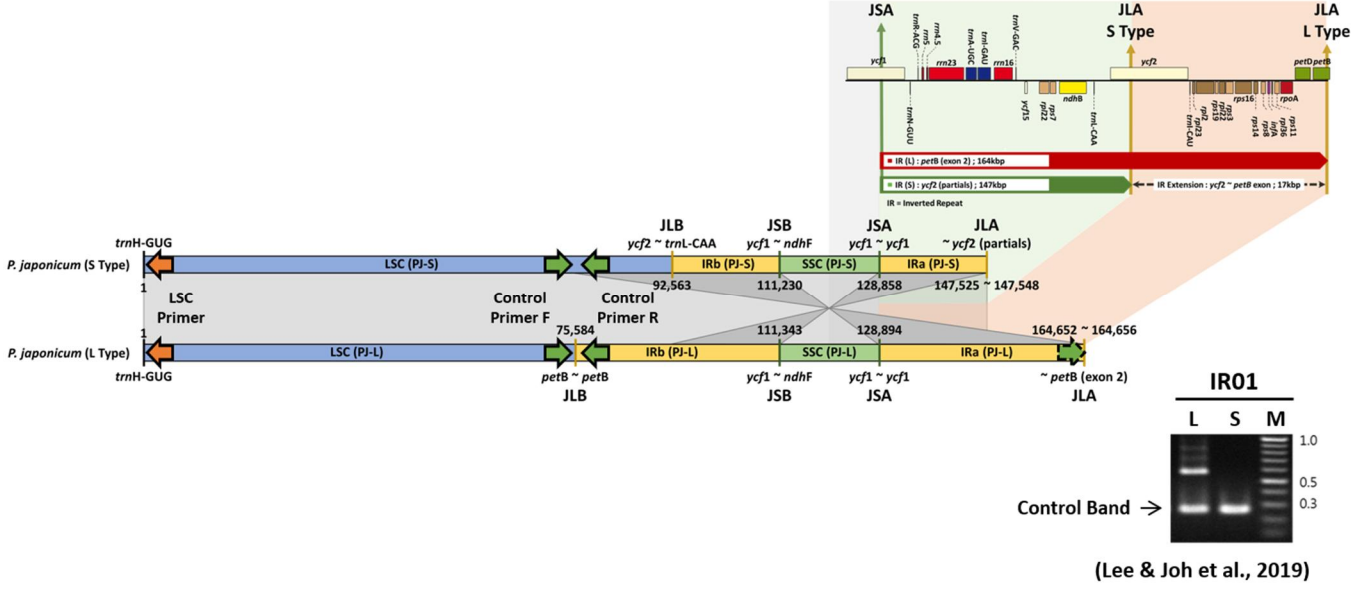


Figure 2-5. Phylogenetic analysis on nine *P. japonicum* accessions based on plastome and 45S nrDNA. The phylogenetic tree was constructed using maximum likelihood (ML). The line of the color represents the type of *P. japonicum* (Red: L-type, Green: S-type).

L / S-type plastome variations in *P. japonicum*

To evaluate the intraspecific diversity among *P. japonicum* germplasm, the IR variation marker from our previous study was first tested to verify the types of *P. japonicum* that exist within 38 germplasm collections [12]. From nine *P. japonicum* accessions, the size and IR junction region of four *P. japonicum* accessions (Pj-6, Pj-7, Pj-8, and Pj-9) are similar to both *L. seseloides* and *G. littoralis*, using IR01 marker was able to differentiate nine accessions and 38 germplasms with long and short IR regions through Polymerase Chain Reaction (PCR) and gel electrophoresis. The IR variation markers have categorized 23 *P. japonicum* germplasms from Geumodo region; Hamyang region; Goheung region; Ulleung Island; Keuneong, Oedolgae in Jeju island; SNU Pharmaceutical garden; Hantaek Botanical garden as the L-type (Table 2-8 No. 1~23). Whereas 15 *P. japonicum* germplasms from Haenam region; Wando region; Jeju Island; Marado Island; Oedo Island; Bakripo beach; SNU Pharmaceutical herb garden; Hantaek Botanical garden as the S-type (Table 2-8 No. 24~38). *P. japonicum* germplasms collected within Jeju island have demonstrated a mixture of L and S-type, which signifies the existence of the varieties in *P. japonicum* plastome.



(Lee & Joh et al., 2019)

Figure 2-6. The plastome structure difference between *P. japonicum* L and S-type. The following figure represents plastome variation of both L and S-types caused by LSC/IR (JLB) junction changes. The differences between each type are indicated on the top right corner, where gene contents of the IR region are shown. The colored area represents a similar pattern shown in the plastome map in Figures 2-1. The orange and green color arrow figure represents the molecular marker IR01 from the previous study by Lee and Joh et al. [12]. An identical IR01 marker from a previous study authenticated L and S-type *P. japonicum*.

Table 2-8. Information of IR types on 38 *P. japonicum* germplasms

No.	Collected Location	Type of *P. japonicum*	Assembled plastome	No.	Collected Location	Type of *P. japonicum*	Assembled plastome
1	Geumodo region	L-type	Pj-3	20	Ulleung island	L-type	
2	SNU Pharmaceutical herb garden	L-type		21	SNU Pharmaceutical herb garden	L-type	
3	Hantaek botanical garden	L-type		22	Hantaek botanical garden	L-type	Pj-4
4	Hamyang region	L-type		23	Hantaek botanical garden	L-type	
5	Goheung region	L-type		24	Jeju island	S-type	Pj-6
6	Hantaek botanical garden	L-type		25	Haenam region	S-type	Pj-9
7	Keuneong region in Jeju island	L-type	Pj-5	26	SNU Pharmaceutical herb garden	S-type	
8	Oedolgae region in Jeju island	L-type		27	Wando region	S-type	Pj-7
9	Ulleung island	L-type		28	Hantaek botanical garden	S-type	
10	Ulleung island	L-type		29	SNU Pharmaceutical herb garden	S-type	
11	Ulleung island	L-type		30	Bakripo beach	S-type	
12	Ulleung island	L-type		31	Oedo island	S-type	
13	Ulleung island	L-type		32	SNU Pharmaceutical herb garden	S-type	
14	Ulleung island	L-type		33	Seopjikoji beach in Jeju island	S-type	
15	Ulleung island	L-type		34	Marado island	S-type	
16	Ulleung island	L-type		35	SNU Pharmaceutical herb garden	S-type	Pj-8
17	Ulleung island	L-type		36	SNU Pharmaceutical herb garden	S-type	
18	Ulleung island	L-type		37	Hantaek botanical garden	S-type	
19	Ulleung island	L-type		38	Bakripo beach	S-type	

Application of intraspecific InDel marker to the *P. japonicum* germplasms

Based on the intra-specific polymorphic data from nine accessions, 17 intraspecific markers was developed to navigate the significance of intraspecific variation between L and S-type among 38 *P. japonicum* germplasm (Figure 2-7). Among 17 markers, Pj14 and Pj17 have demonstrated a distinctive difference between the L-types and the S-types (Figure 2-7A). Considering each amplified band in the gel figure as a genotype, the individual germplasm on the L-type illustrated a uniformed genotype (Figure 2-7A 1~23). In comparison, individual germplasm on the S-type illustrated dynamic genotype among other S-types (Figure 2-7A 24~38). However, there are some genotypes distinguished from other genotypes. No. 31 germplasm of the S-type *P. japonicum* collected from Oedo island have demonstrated unique genotype in Pj04 marker which no other germplasm has presented. No. 4 germplasm of the L-type *P. japonicum* collected from SNU Pharmaceutical herb garden, and No. 24 germplasm of the S-type *P. japonicum* collected from Jeju island demonstrated a unique band type in Pj08 marker. Some germplasms on type L have illustrated similar genotypes as the S-types.

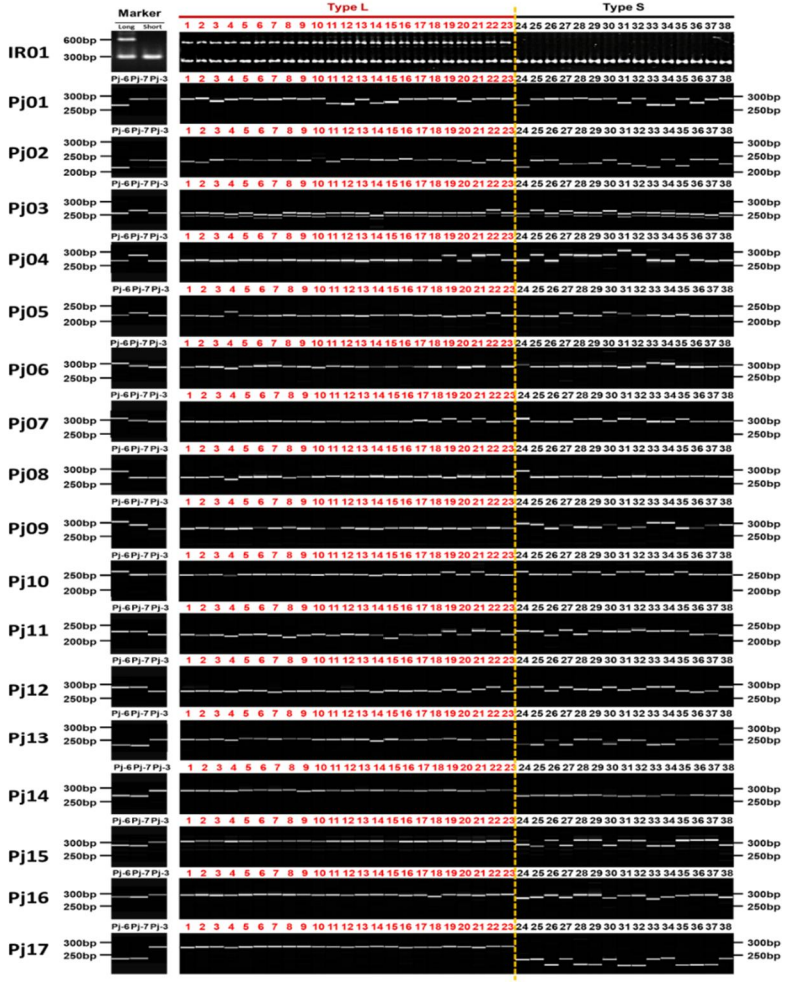
To validate the similarities between each *P. japonicum* germplasms UPGMA method uses (Figure 2-7B). Among the list of the 17 intraspecific markers in Table 2-9, few germplasms of L and S-type *P. japonicum* have illustrated a close relationship to each other. From gel electrophoresis, four *P. japonicum* germplasm (Figure 2-7A No. 15, 19, 21, 22) of long IR type presented genotype variation that displayed similar genotypes in the S-type *P. japonicum* germplasm (Figure 2-7A No. 25, 27, 28, 29, 30,

31, 32, 33, 34, 35). The dendrogram based on the UPGMA method also presented germplasms, No. 19, 21 from the L-type and No. 26, 36, 37 from the S-type, with a closer relationship to opposite types (Figure 2-7B). Aside from those germplasms, most germplasms are closely grouped, following the IR types.

Table 2-9. Information on InDel markers used for intra-specific diversity analysis

Marker ID	InDel		Primer Sequence (5'-3')	InDel Size	InDel Motif (Motif size)	Product size [a] (Pj-3 / Pj-6 / Pj-7)	Gene position
Pj01	Delete	F	ATACTGGACTGGAAAGGAATTTGAG	21	TAACCTAAGTGCAAAAATAGA (21)	285/264/285	*ycf3*/*trn*S-GGA
		R	CTTTCTTTTGAGGTTAGGCGGATA				
Pj02	Delete	F	ATAAGGATGAAATCACGCTCTGTAG	22	TTATATATATA (11)	233/211/233	*trn*W-CCA/*trn*P-UGG
		R	ATTACTTCTTCGTTGAGCAGTAACG				
Pj03	Insert	F	CCACATATTATGGGTAACGTAGGAG	9	AATGTGAAT (9)	248/248/257	*rpl2*/*rpl2*
		R	GCAGCAAGTGATTGAGTTCAGTAGT				
Pj04	Insert	F	CCTTATCTCCCTCTAGGGGTATTTT	18	TAGTGA(T)GATAT(T)GATG(C) (18) TAGTGA(C)GATAT(T)GATG(C) (18) TAGTGA(C)GATAT(C)GATG(C) (18) TAGTGA(C)GATAT(C)GATC(G) (18)	261/261/279	*ycf2*/
		R	TATCCATAGTTAGCAGTTCCAGCTC				
Pj05	Insert	F	TTTGGGTAATACCACCTTCA	9	ATTTATTGT (9)	215/215/224	*ycf1*/
		R	TCCCGTTTTTCTTAAGACTG				
Pj06	Insert	F	GCCTAAAGAAACGAAAGAATCG	10	AAAAAATATA (10)	/287/300/291	*atpF*/*atpH*
		R	ATTAGCATACTCACTCGCTTTCATC				
Pj07	Insert	F	ACATCTAGGGGTAAAGCACTGTTTC	10	AATAAAATA (9)	288/298/288	*rrn23*/
		R	GTGAGCTATTACGCACTCTTTCAA				
Pj08	Insert	F	CAATTCTAACTAGCCCTAATGGTCA	21	TAAAAAAAGTAGTAAATTAAT (21)	266/287/266	*ycf1*/*ndhF*
		R	GCGCATCTCTTCTTATCTTTTCTTG				
Pj09	Insert	F	GCCCTAACCATATTTCGACTTGTAA	25	TTAAACTAATTTCCT (15)	274/299/289	*ndhE*/*ndhG*
		R	GGGGTTATCACATTTATTTCCCTTC				
Pj10	Insert	F	ATCAGTGAGCTATTACGCACTCTTT	10	AATAAAATA (9)	288/298/288	*rrn23*/
		R	CAAATCGAGGCAAACTCTGAATACT				
Pj11	Delete	F	CTTTTCATTAGGCAATGGGTTC	11	AGGGTTCTTAA (11)	217/228/228	*ndhF*/*rpl32*
		R	GTTGGATGTGAAAGACATCTATTGG				
Pj12	Delete	F	TAATCTACTTCTGCAGTGGCTCATA	13	TTTATGAA (8)	265/278/278	*psaC*/*ndhE*
		R	CGCTATTGGACTAGCTATTGTTTCA				
Pj13	Insert	F	TGGCCATAATAGTGTAGTGATGCTA	17	TTCTTTTCCTCCCACCCC (18)	247/230/230	*trn*G-GCC/*trn*fM-CAU
		R	GCCCATTCAACTCTTGTTTGTCTAT				
Pj14	Insert	F	TGTTCGGATCTATTATGACATAGGC	18	ATTAGCCATTACTAATAG (18)	286/268/268	*psaA*/*ycf3*
		R	GAATAAAAGAGACCCTCGAATCACT				
Pj15	Insert	F	AAGCGAAGTTTACGGAAGTAGTGTA	15	(T)ATTCTATATATATT (15) (C)TATATATATATATT (15)	295/282/280	*rps4*/*trn*T-UGU
		R	CGATAGCCGGCTTTTCTCTATTTAT				
Pj16	Insert	F	TTGGTATTCTCCGTGTATTCTTACG	10	TCCCCTTTTT (10)	289/279/279	*rps11*/*rpl36*
		R	GTGAAAAATGTCGACTAATCCGTAG				
Pj17	Insert	F	ATTTCATAGGGAACCCCAAAGT	44	(CTATTTCTTTTCTATATATGTT) (22)	277/233/233	*trn*N-GUU/*ycf1*
		R	TCACCACAAACCTCCTTTTTCT				

a Amplicon size expected for each *Peucedanum japonicum* Accessions. Pj-3: Geumodo; Pj-6: Jeju Island; Pj-7: Wando region


Marker
Long
Short
Type L
Type S
IR01
Pj01
Pj02
Pj03
Pj04
Pj05
Pj06
Pj07
Pj08
Pj09
Pj10
Pj11
Pj12
Pj13
Pj14
Pj15
Pj16
Pj17
600bp
300bp
250bp
200bp
Pj-6 Pj-7 Pj-3

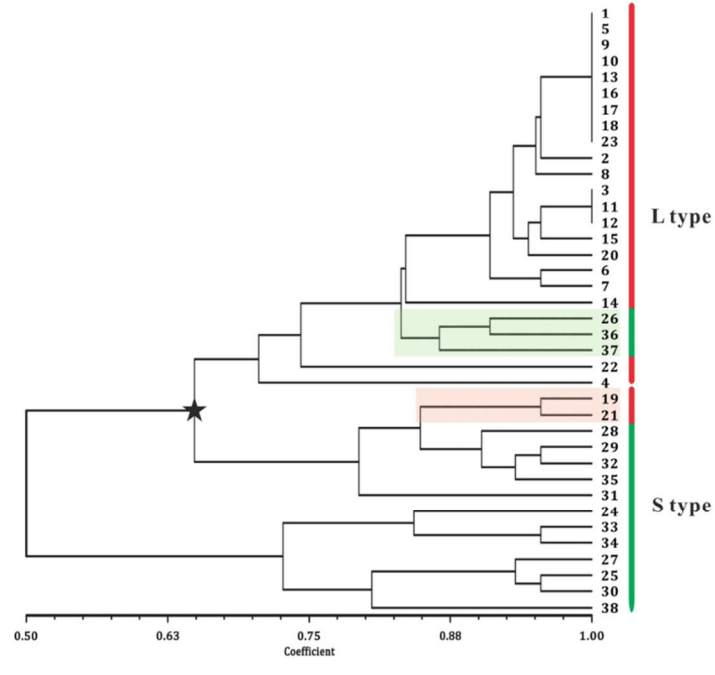

L type
S type
0.50
0.63
0.75
0.88
1.00
Coefficient

Figure 2-7. Application of molecular markers to 38 *Peucedanum japonicum* germplasms. 17 InDel marker analysis on 38 *P. japonicum* germplasms. Numbers (1~38) indicate plant no. as described in Table 2-8. For IR01 marker analysis, electrophoresis using 1% agarose gel. M represented a 100 bp DNA ladder. For 17 InDel markers, analysis was performed using Fragment Analyzer (Advanced Analytical Technologies Inc., USA) following the manufacturer's instructions. Pj-3, Pj-6 and Pj-7 indicate the completed assembly on *P. japonicum* accessions from Chapter 1 (Table 2-2). B. Cladogram analysis on *P. japonicum* germplasm on 17 InDel marker using UPGMA method. Each amplified band of the molecular marker was annotated into genotype using binary code. Then, the cladogram was constructed using the UPGMA method (NTSYS 2.0). Numbers in the cladogram represent each of 38 *P. japonicum* germplasms.

Development of KASP markers and validation

Eight KASP markers were applied to 38 *P. japonicum* germplasms (Table 2-10). Under Eight markers successfully amplified all 38 germplasms and formed two clusters of genotypes. The X and Y axes represent the FAM (465-510nm) and HEX (533-580nm) values, respectively, which represent the alleles of the SNP (Figure 2-8). The result of the KASP marker regarding 38 germplasms were two types under *P. japonicum* (L/S-types) have represented a polymorphic variation genotype under eight genic marker regions. Each marker was grouped into the types for each germplasm and separated each individual by the cladogram tree analysis results (Figure 2-9). As it turns out, the L-types clustered into two groups among each marker, whereas the S-types were clustered into four groups.

Most L-type germplasms presented the same genotype variation under eight markers. However, the germplasms under Keuneong (#7) and Oedolgae (#8) region on Jeju island demonstrated different genotypes under *rpo*C2 and *ycf1* marker, with the same genotype as the S-type germplasms (Table 2-11). Among the S-type germplasms, genotypes were separated into four different groups. Germplasms from SNU (#26, #29, #32, #36), HT (#28, #35, #37), Oedo (#31) region represented a same genotype as as the L-type germplasms under *rpoA, ycf4 and rpl2* region (Table 2-11), refer to the group as S1. Germplasms from Haenam (#25), Marado (#34), Bakripo (#30, #38) region illustrated a similar genotype with S1, except the genotype appeared under the *rpoA region*; refer to the group as S2 (Table 2-11). The Wando (#27) region represented the same genotype as S2 except under the *psbK* region, referred to as S3 (Table 2-11). Lastly, Jeju (#24) and Seopjikoji (#33) region under Jeju island provided even further

differences on the genotype under *ycf4*, *psbK,* and *rpl2* region, then S2, and S3; referring as an S4 (Table 2-11). UPGMA analysis based on the KASP marker genotype result has demonstrated the separation of both L and S-types while illustrating the L-type clustered into two groups and the S-type is clustered into four groups (Figure 2-9). Various genotypes provided by the KASP marker and the grouping under the S-type germplasms support evidence that the S-type is more varied than the L-type.

Table 2-10. Information on KASP markers used for intra-specific diversity analysis

Marker ID	FAM Allele	HEX Allele	Primer Seq Allele X	Primer Seq Allele Y	Primer Seq common
PJ_KASP_rpoA	G	T	GCATTTATCAATTGATTTACCGAAAAAGTC	AAAGCATTTATCAATTGATTTACCGAAAAAGTA	GAGTTGTGCTAAAGATTCAAAACCCATTTT
PJ_KASP_ycf4	A	C	AAATCGTCGCATCTTCCTCCGATTA	CGTCGCATCTTCCTCCGATTC	CTTCTATTCTGACGGACTGAATATCCTTT
PJ_KASP_psaA	G	A	CCTTATCCGTATCTAGCTACTGAC	CCCTTATCCGTATCTAGCTACTGAT	GTGTGAACAATGACAGTTGTGTACCATA
PJ_KASP_psbK	G	A	CATTTGTCTTAATTATGCCTTTTATTCGAG	GCATTTGTCTTAATTATGCCTTTTATTCGAA	CGGGCAATTTGGCGAAGAAAAGACTA
PJ_KASP_rpl2	T	C	GGACAAGTGGGGAATGTTGGGA	GACAAGTGGGGAATGTTGGGG	GCTTAGATCCGGCTCTACCCAAATT
PJ_KASP_rpoC2	G	A	CATTTCCATATGTAAATTCAGGTGCATG	AACATTTCCATATGTAAATTCAGGTGCATA	GAAATGCACTGGAATACCGACGTGTA
PJ_KASP_rps15	C	T	GGATACGGAGACTTACTTCACATTTG	AGGATACGGAGACTTACTTCACATTTA	GCAGACCTCTCTGAGATAAATAGTCTTT
PJ_KASP_ycf1	A	G	TTTCTTTTAATATCCCTGAACAGATGAAATTT	CTTTTAATATCCCTGAACAGATGAAATTC	TTCTGATCCTTTACCCATCCAATTTCTAAA

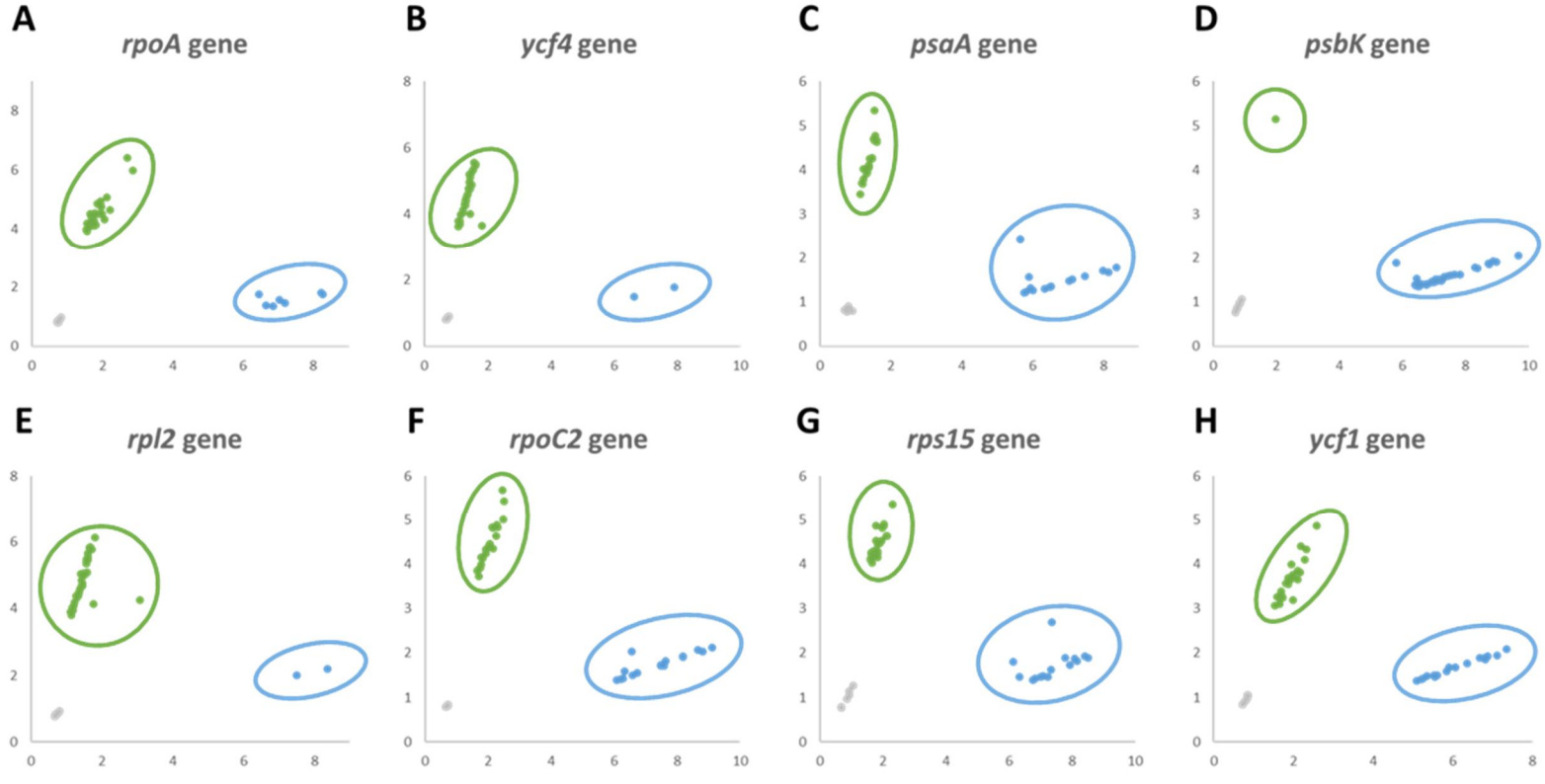


Figure 2-8. Genotype plot of eight KASP (A~H) markers targeting SNP variants on 38 *P. japonicum* germplasms with FAM (Green) and HEX (Blue) fluorophores. In this case, 38 germplasms were genotyped into two distinct clusters.

Table 2-11. Genotype data from eight KASP marker

Accession	IR	rpoA	ycf4	psaA	psbK	rpl2	rpoC2	rps15	ycf1
1	L	Y	Y	Y	X	Y	Y	Y	Y
2	L	Y	Y	Y	X	Y	Y	Y	Y
3	L	Y	Y	Y	X	Y	Y	Y	Y
4	L	Y	Y	Y	X	Y	Y	Y	Y
5	L	Y	Y	Y	X	Y	Y	Y	Y
6	L	Y	Y	Y	X	Y	Y	Y	Y
9	L	Y	Y	Y	X	Y	Y	Y	Y
10	L	Y	Y	Y	X	Y	Y	Y	Y
11	L	Y	Y	Y	X	Y	Y	Y	Y
12	L	Y	Y	Y	X	Y	Y	Y	Y
13	L	Y	Y	Y	X	Y	Y	Y	Y
14	L	Y	Y	Y	X	Y	Y	Y	Y
15	L	Y	Y	Y	X	Y	Y	Y	Y
16	L	Y	Y	Y	X	Y	Y	Y	Y
17	L	Y	Y	Y	X	Y	Y	Y	Y
18	L	Y	Y	Y	X	Y	Y	Y	Y
19	L	Y	Y	Y	X	Y	Y	Y	Y
20	L	Y	Y	Y	X	Y	Y	Y	Y
21	L	Y	Y	Y	X	Y	Y	Y	Y
22	L	Y	Y	Y	X	Y	Y	Y	Y
23	L	Y	Y	Y	X	Y	Y	Y	Y
7	L	Y	Y	Y	X	Y	X	Y	X
8	L	Y	Y	Y	X	Y	X	Y	X
26	S	Y	Y	X	X	Y	X	X	X
28	S	Y	Y	X	X	Y	X	X	X
29	S	Y	Y	X	X	Y	X	X	X
31	S	Y	Y	X	X	Y	X	X	X
32	S	Y	Y	X	X	Y	X	X	X
35	S	Y	Y	X	X	Y	X	X	X
36	S	Y	Y	X	X	Y	X	X	X
37	S	Y	Y	X	X	Y	X	X	X
25	S	X	Y	X	X	Y	X	X	X
30	S	X	Y	X	X	Y	X	X	X
34	S	X	Y	X	X	Y	X	X	X
38	S	X	Y	X	X	Y	X	X	X
27	S	X	Y	X	Y	Y	X	X	X
24	S	X	X	X	X	X	X	X	X
33	S	X	X	X	X	X	X	X	X

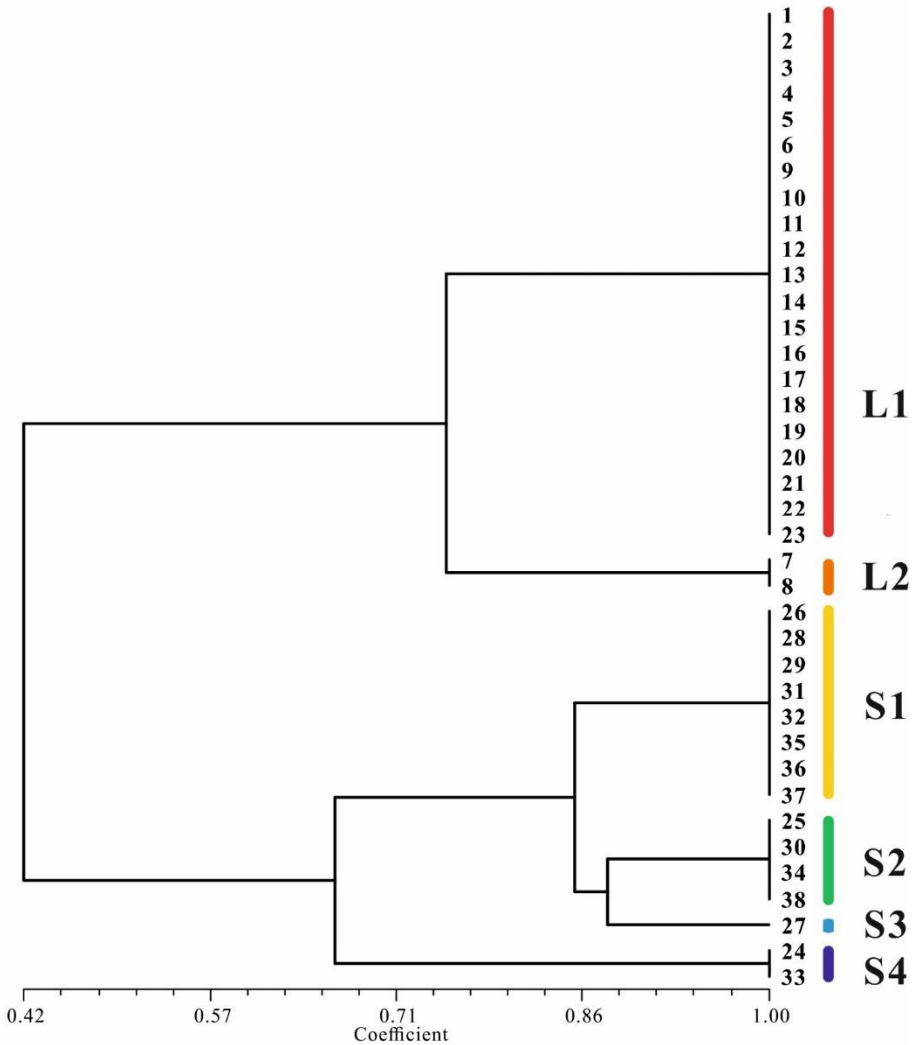


Figure 2-9. Cladogram of 38 *P. japonicum* germplasm collections based on plastome eight KASP markers. The cladogram was constructed using the UPGMA method (NTSYS 2.0). Numbered L and S-types represent the divided groups based on Table 2-11.

DISCUSSION

Overview of a major difference between two types of *P. japonicum*

In this study, *P. japonicum* accessions were sequenced and annotated to further understand intraspecific genetic diversity. Complete assembled *P. japonicum* plastome provided intraspecific variation and discovered a sub-genotype within *P. japonicum* accessions (Figure 2-1, Table 2-4). Between nine *P. japonicum* accessions, the most distinctive characteristic observed is the difference among the plastome size (164 kbp, 147 kbp) induced from the IR regions. The intraspecific IR junction was visualized in the plastome map drawn using OGDRAW (Figure 2-1) and categorized the two *P. japonicum* plastomes into L and S-types reflected from intraspecific IR junctions. IR junction has shifted between the two types from *petB* intron to near 3' end *ycf2* gene. The IR length of the L-type (35,759~35,760 bp) appears to be nearly two times longer than the S-type (18,606~18,667 bp) (Table 2-2). Besides the length and the junction changes in the IR region, both L and S-types *P. japonicum* plastome exposed no changes on gene contents nor the gene order (Figure 2-1, 2-4). Comparative analysis on nine plastome sequences, the number of polymorphic variations represented a more significant number of InDel and SNP in the S-type (< 55; < 126), whereas between the L-type, polymorphic variation remains small (< 2; < 9) on both InDels and SNPs respectively (Table 2-6). Such postulated a broad genetic diversity appearance within S-type accessions from these outputs compared to the L-type accessions.

Phylogenetic relationship between *P. japonicum* accessions

The phylogenetic analysis through the plastome sequence demonstrated separation between L and S-type as a clade. Within S-type clade, Pj-7 (Wando) and Pj-9 (Haenam) accessions demonstrate closely clustered than Pj-6 (Jeju island) (Table 2-1). The region difference, upon which the Wando and Haenam region location is much closer than Jeju island, has implied such results. Pj-8 (SNU) accession illustrated a speciation event before L and S-type diverges among nine accessions (Figure 2-5). Such output can be verified by the number of polymorphisms within Pj-8 accessions appears higher than other *P. japonicum* accessions combined (Table 2-6). The outcome of the IR junction in the most Group 3 of Apioideae species in Chapter 1 remains within the *ycf2* region and the early speciation of Pj-8 accession in Chapter 2, illustrates the recently divergence of the *P. japonicum* L-type from the *P. japonicum* S-type. On the other hand, phylogenetic analysis based on nrDNA sequences revealed a mixture of both L and S-types: Pj-1, Pj-2, Pj-8 accessions and Pj-5, Pj-6, Pj-9 accessions are closely clustered (Figure 2-5). The coexistence on both L and S-type polymorphism under 45S nrDNA sequence seems possible due to bi-parental inheritance features that led the possibility of crossing between L and S-types. The parsimony information regarding the nrDNA sequence (4 SNP) is too small to define the answer phylogeny differences between plastome and nrDNA phylogeny in *P. japonicum* (Figure 2-4, Table 2-7).

Among nine *P. japonicum* accession, Pj-8 has revealed early speciation compared to other accessions from plastome-based phylogenetic analysis. The number of polymorphisms and phylogenetic analysis showed Pj-8 is far more significant than other *P. japonicum* accessions, implying Pj-8 accession belongs under different

Peucedanum species other than *P. japonicum*. Despite the identical nrDNA sequence length on nine *P. japonicum* accessions, the small number of SNP polymorphisms discovered between each accession and the same genotype under 45S nrDNA with Pj-1 and Pj-2 demonstrate that Pj-8 accessions belong under the *P. japonicum*.

Broad intraspecific diversity within *P. japonicum* population

The intraspecific diversity between nine *P. japonicum* accessions revealed 73 InDel polymorphism and 315 SNPs. The 73 InDels are composed of repeat sequences located within the intergenic region. The evaluation of 17 InDel based markers on 38 *P. japonicum* germplasm collections from gel electrophoresis provided evidence of higher genetic diversity in the S-type. The discovered 315 SNPs are mostly located within the plastome's LSC region, followed by SSC then the IR region (Figure 2-2, Table S2-2). Among nine *P. japonicum* accessions, SNPs are significantly represented more in Pj-8. Even amongst the S-type accessions, the number of polymorphisms on Pj-8 was the highest (Table 2-6). The evaluation of the eight SNP-based KASP markers on 38 *P. japonicum* germplasm collections, the KASP markers have demonstrated a broad diversity of the S-type *P. japonicum* populations with sub-divided four groups (S1, S2, S3, S4), which Pj-8 accession is clustered under the S1 group. UPGMA analysis based upon the InDel marker exhibits a *P. japonicum* germplasm cluster with different types. In detail, the S1 type germplasms (#26, #36, #37) is closely clustered with the L1 type germplasms (#14, #22) (Figure 2-10). The S1 type germplasm is a common lineaege of *P. japonicum*, as evidenced by the greater number of polymorphisms discovered under

Pj-8 accession, which are closely related to L-type accession Pj-1 and Pj-2 under 45S nrDNA and UPGMA analysis, indicating a closer relationship between S1 and L1 type germplasms. While the application of eight KASP markers was able to detect the narrow genetic diversity on the L-type has diversified into another L-type (Figure 2-9, Table 2-11).

The study on intraspecific diversity is necessary as it reflects the evolutionary history of the species population with identifying the potential for adapting to environmental changes [31, 32]. Changes in the environment affect the phenotypic features responding to new environmental conditions, which leads to the genetic variants of plant species. Despite so, most evolution-related studies are focused more toward the inter-genus level of comparative analysis; thus, intraspecific analysis is still poorly described in plant studies. The recent study on intraspecific genetic diversity on *Ulva* species demonstrates 36 variant sites [33], which in comparison *P. japonicum*, provide broader intraspecific diversity with 315 variant sites. As six new distinct L and S-types appeared from the eight SNP-based markers, continuous development of the other molecular markers from 315 SNPs would provide even more accurate distinguishable markers to navigate broad genetic diversity within *P. japonicum*. No strong morphological relationship has appeared within *P. japonicum*. Plastome-based developed molecular marker polymorphic site would target the specific morphological features related as a specific trait under *P. japonicum* that would benefit the agricultural background on *P. japonicum* breeding in future studies.

Overall, this chapter, demonstrated a broader intraspecific diversity within the *P. japonicum* germplasms collections from this study. The comparative intraspecific

analysis has demonstrated that the two types are based on the 17kbp length differences from the IR junction changes discovered in Chapter 1. But even further, the intraspecific polymorphisms (InDels and SNPs) among the L and S-type *P. japonicum* shares different genetic diversity rates. The L-type *P. japonicum* represented low genetic diversities, which corresponds to the result from the previous study with an identical plastome [12]. On the contrary, the S-type *P. japonicum* illustrated high genetic diversities, especially the Pj-8 accessions represented with an early speciation event through phylogenetic analysis. Molecular marker analysis through 38 germplasm populations provided four subgroups on the S-type and two subgroups on the L-type. Such marker-based analysis would define *P. japonicum* germplasm collection with specific morphological features in future studies.

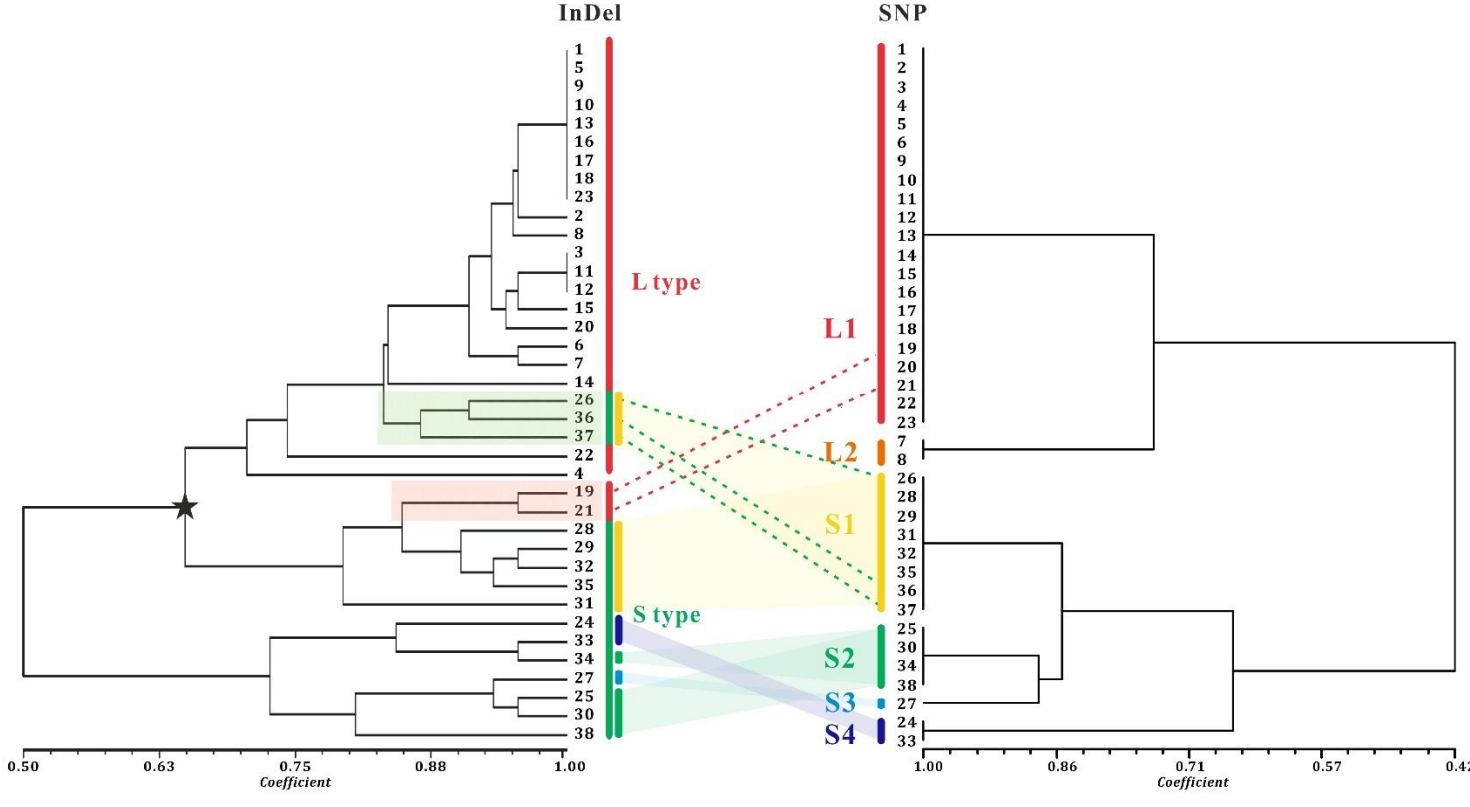


Figure 2-10. Illustration of comparison on UPGMA result based on InDel and SNP based markers. Each colored bar represents the type of 38 *P. japonicum* germplasm collections (Red: Type L1; Orange: Type L2; Yellow: Type S1; Green: Type S2; Blue: Type S3; Purple: Type S4

REFERENCES

1. Sarkhail, P. (2014). Traditional uses, phytochemistry and pharmacological properties of the genus Peucedanum: A review. *Journal of ethnopharmacology* 156, 235-270.

2. Nukitrangsan, N., Okabe, T., Toda, T., Inafuku, M., Iwasaki, H., and Oku, H. (2012). Anti-obesity activity of *Peucedanum japonicum* Thunb extract in obese diabetic animal model C57BL/6J Ham Slc-ob/ob mice. *International Journal of Life Science and Medical Research* 2(2), 7.

3. Chun, JM., Lee, AR., Kim, HS., Lee, AY., Gu, GJ., Moon, BC., and Kwon, BI. (2018). *Peucedanum japonicum* extract attenuates allergic airway inflammation by inhibiting Th2 cell activation and production of pro-inflammatory mediators. *Journal of ethnopharmacology* 211, 78-88.

4. Aida, Y., Kasama, T., Takeuchi, N., Chiba, M., and Tobinaga, S. (1998). Pharmacological activities of khellactones, compounds isolated from *Peucedanum japonicum* THUNB. and *Peucedanum praeruptorium* DUNN. *Methods and findings in experimental and clinical pharmacology* 20(4), 343-352.

5. Ikeshiro, Y., Mase, I., and Tomita, Y. (1992). Dihydropyranocoumarins from roots of *Peucedanum japonicum*. *Phytochemistry* 31(12), 4303-4306.

6. Jun, NJ., Kim, SC., Song, EY., Jang, KC., Lee, DS., and Cho, SK. (2014). Isolation of anticancer compounds from *Peucedanum japonicum* Thunb. roots. *Korean*

Journal of Plant Resources 27(3), 215-222.

7. Cho, YH., Seong, NS., Ham, IH., and Choi, HY. (2004). A comparative study on the immunization and anti-cancer effect of the root and the aerial part of *Peucedanum japonicum* Thunb. *The Korea Journal of Herbology* 19(3), 137-137.

8. Kim, JK., Kang, HM., Jang, DC., Na, JK., and Choi, KY. (2020). Effect of Light Intensity and Temperature on the Growth and Functional Compounds in the Baby Leaf Vegetable Plant *Peucedanum japonicum* Thunb. *Horticultural Science and Technology* 38(6), 822-829

9. Sugiura, M. (1992). The chloroplast genome. *10 Years plant molecular biology*, 149-168.

10. Wicke, S., Schneeweiss, GM., Depamphilis, CW., Müller, KF., and Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant molecular biology* 76(3), 273-297.

11. Hollingsworth, PM., Graham, SW., and Little, DP. (2011). Choosing and using a plant DNA barcode. *PloS one* 6(5), e19254.

12. Lee, HO., Joh, HJ., Kim, K., Lee, SC., Kim, NH., Park, JY., ... and Yang, TJ. (2019). Dynamic chloroplast genome rearrangement and DNA barcoding for three apiaceae species known as the medicinal herb “Bang-Poong”. *International journal of molecular sciences* 20(9), 2196.

13. Allen, GC., Flores-Vergara, MA., Krasynanski, S., Kumar, S., and Thompson, WF. (2006). A modified protocol for rapid DNA isolation from plant tissues using

cetyltrimethylammonium bromide. *Nature protocols* 1(5), 2320-2325.

14. Kim, K., Lee, SC., Lee, J., Lee, HO., Joh, HJ., Kim, NH., ... and Yang, TJ. (2015). Comprehensive survey of genetic diversity in chloroplast genomes and 45S nrDNAs within Panax ginseng species. *PloS one* 10(6), e0117159.

15. Kim, K., Lee, SC., Lee, J., Yu, Y., Yang, K., Choi, BS., ... and Yang, TJ. (2015). Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of Oryza AA genome species. *Scientific reports* 5(1), 1-13.

16. Kurtz, S., Phillippy, A., Delcher, AL., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, SL. (2004). Versatile and open software for comparing large genomes. *Genome biology* 5(2), 1-9.

17. Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, ES., Fischer, A., Bock, R., and Greiner, S. (2017). GeSeq–versatile and accurate annotation of organelle genomes. *Nucleic acids research* 45(W1), W6-W11.

18. Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic acids research*, 47(W1), W59-W64.

19. Lagesen, K., Hallin, P., Rødland, EA., Stærfeldt, HH., Rognes, T., and Ussery, DW. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic acids research*, 35(9), 3100-3108.

20. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, TL. (2009). BLAST+: architecture and applications. *BMC*

bioinformatics 10(1), 1-9.

21. Katoh, K., and Standley, DM. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* 30(4), 772-780.

22. Frazer, KA., Pachter, L., Poliakov, A., Rubin, EM., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic acids research* 32(suppl_2), W273-W279.

23. Hall, T., Biosciences, I., and Carlsbad, C. (2011). BioEdit: an important software for molecular biology. *GERF Bull Biosci* 2(1), 60-61.

24. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, BC., Remm, M., and Rozen, SG. (2012). Primer3—new capabilities and interfaces. *Nucleic acids research* 40(15), e115-e115.

25. Rohlf, FJ. (1992). *NTSYS-pc: numerical taxonomy and multivariate analysis system*. Applied Biostatistics.

26. Semagn, K., Babu, R., Hearne, S., and 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.

27. Salamov AA., Solovyev VV. (2000) Ab initio gene finding in Drosophila genomic DNA. *Genome Res* 10(4), 516-22.

28. Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after

removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic biology* 56(4), 564-577.

29. Silvestro, D., and Michalak, I. (2012). raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity & Evolution* 12(4), 335-337.

30. Rambaut, A., and Drummond, AJ. (2012). FigTree version 1.4. 0.

31. Pauls, SU., Carsten N., Miklós B., and Markus P. (2013). The Impact of Global Climate Change on Genetic Diversity within Populations and Species. *Molecular ecology* 22(4), 925-46.

32. Paz-Vinas, I., Géraldine L., Virgilio H., Charlotte V., Nicolas P., Gaël G., and Simon B. (2018). Systematic Conservation Planning for Intraspecific Genetic Diversity. *Proceedings of the Royal Society B: Biological Sciences* 285(1877), 20172746.

33. Fort, A., Marcus M., Kevin C., Philippe P., Björn U., Michael DG., and Ronan S. (2021). Foliose Ulva Species Show Considerable Inter-Specific Genetic Diversity, Low Intra-Specific Genetic Variation, and the Rare Occurrence of Inter-Specific Hybrids in the Wild. *Journal of phycology* 57(1), 219-33.

SUPPLEMENTARY DATA

Table S1-1. Information of Plastome sequence of Araliaceae family.

No.	Family/ Sub-Family	Species	Length (bp)	LSC (bp)	SSC (bp)	IRs (bp)	IR junction Position	G+C %	Gene	CDS	tRNA	rRNA	Clades/Tribes	ID
1	Araliaceae	*Aralia atropurpurea*	156226	86231	18068	25963	*rps19*	38.14	132	87	37	8	Aralia	NC_056999
2	Araliaceae	*Aralia continentalis*	155999	86049	18052	25949	*rps19*	38.15	132	87	37	8	Aralia	NC_041648
3	Araliaceae	*Aralia cordata*	156087	86101	18052	25967	*rps19*	38.13	133	87	37	8	Aralia	NC_041648
4	Araliaceae	*Aralia elata*	156220	86263	18111	25923	*rps19*	38.08	132	87	37	8	Aralia	KT153023
5	Araliaceae	*Aralia undulata*	156333	86029	18092	26105	*rps19*	38.1	132	87	37	8	Aralia	NC_022810
6	Araliaceae	*Brassaiopsis hainla*	156459	86567	18020	25936	*rps19*	37.92	132	87	37	8	Brassaiopsis	NC_022811
7	Araliaceae	*Cheirodendron bastardianum*	156698	86571	18191	25975	*rps19*	38	132	87	37	8	Cheirodendron	NC_049884
8	Araliaceae	*Dendropanax dentiger*	156687	86680	18247	25880	*rps19*	37.96	132	87	37	8	Dendropanax	NC_026546
9	Araliaceae	*Dendropanax morbifer*	156366	86475	18125	25883	*rps19*	37.99	132	87	37	8	Dendropanax	NC_027607
10	Araliaceae	*Dendropanax oligodontus*	156403	86440	18075	25944	*rps19*	37.99	132	87	37	8	Dendropanax	NC_053618
11	Araliaceae	*Diplopanax stachyanthus*	157522	87640	18182	25850	*rps19*	37.78	132	87	37	8	Diplopanax	NC_029750
12	Araliaceae	*Eleutherococcus brachypus*	156981	86921	18184	25938	*rps19*	37.95	132	87	37	8	Eleutherococcus	NC_050832
13	Araliaceae	*Eleutherococcus gracilistylus*	156770	86729	18175	25933	*rps19*	37.95	132	87	37	8	Eleutherococcus	KT153020
14	Araliaceae	*Eleutherococcus senticosus*	156768	86755	18153	25930	*rps19*	37.95	132	87	37	8	Eleutherococcus	NC_016430
15	Araliaceae	*Eleutherococcus sessiliflorus*	156730	86603	18213	25957	*rps19*	37.95	132	87	37	8	Eleutherococcus	KT153019
16	Araliaceae	*Fatsia japonica*	155613	86488	17867	25629	*rps19*	37.91	132	87	37	8	Fatsia	NC_027685
17	Araliaceae	*Hydrocotyle sibthorpioides*	152880	84064	18690	25063	*rps19*	37.51	132	87	37	8	Hydrocotyle	NC_035502
18	Araliaceae	*Hydrocotyle verticillata*	153207	84352	18739	25058	*rps19*	37.59	132	87	37	8	Hydrocotyle	NC_015818

19	Araliaceae	*Kalopanax septemlobus*	156413	86467	18118	25914	*rps19*	37.95	132	87	37	8	Kalopanax	NC_022814
20	Araliaceae	*Metapanax delavayi*	156343	86361	18130	25926	*rps19*	37.93	132	87	37	8	Metapanax	NC_022812
21	Araliaceae	*Panax major*	156402	86188	18007	26103	*rps19*	38.07	132	87	37	8	Panax	NC_053713
22	Araliaceae	*Panax notoginseng*	156466	86190	18004	26136	*rps19*	38.08	132	87	37	8	Panax	NC_026447
23	Araliaceae	*Panax quinquefolius*	156088	86095	17993	26000	*rps19*	38.07	132	87	37	8	Panax	NC_027456
24	Araliaceae	*Panax stipuleanatus*	156064	86116	18174	25887	*rps19*	38.03	132	87	37	8	Panax	NC_030598
25	Araliaceae	*Panax trifolius*	156157	86322	18047	25894	*rps19*	38.08	132	87	37	8	Panax	NC_037994
26	Araliaceae	*Raukaua anomalus*	156664	86602	18153	25975	*rps19*	38.05	132	87	37	8	Raukaua	NC_049886
27	Araliaceae	*Raukaua simplex*	156713	86598	18202	25958	*rps19*	38.02	132	87	37	8	Raukaua	NC_049888
28	Araliaceae	*Schefflera actinophylla*	156675	86576	18157	25968	*rps19*	37.9	132	87	37	8	Schefflera	NC_049889
29	Araliaceae	*Schefflera delavayi*	156341	86123	18146	26036	*rps19*	37.83	132	87	37	8	Schefflera	NC_022813
30	Araliaceae	*Schefflera heptaphylla*	156685	86610	18147	25964	*rps19*	37.93	132	87	37	8	Schefflera	NC_029764

Table S1-2. Samples information and summary of plastomes sequence downloaded from NCBI Genbank database

No	Tribe/Clade	Scientific Name	Korean Name	IR Junction	Plastome Size	LSC	SSC	IRs	G+C %	Gene	CDS	tRNA	rRNA	ID
1	Araliaceae	*Panax ginseng cultivar Chunpoong*	인삼	*rps19*	156,248	86,128	18,084	26,018	38.07	130	85	37	8	KM088019
2	Araliaceae	*Aralia elata*	두릅나무	*rps19*	156220	86263	18,111	25923	38.08	132	87	37	8	KT153023
3	Araliaceae	*Hydrocotyle verticillata*		*rps19*	153207	84352	18739	25058	37.59	132	87	37	8	NC_015818
4	Chamaesium clade	*Chamaesium delavayi*		*rps19*	154,684	85,029	17,415	26,120	38.34	130	85	37	8	MN119367
5	Komarovia clade	*Chuanminshen violaceum*		*rps19*	154530*	84172*	17,800	26,279	37.79	130	85	37	8	KU921430.2
6	Physospermopsis clade	*Hansenia weberbaueriana*	중국강활	*rps19*	158,625	88,260	18,237	26,064	37.68	130	85	37	8	NC_035053.1
7	Oenantheae	*Cicuta virosa*		*rps19*	154,569	84,177	17,578	26,407	37.52	130	85	37	8	NC_037711
8	Oenantheae	*Tiedemannia filiformis*		*rps19*	154,737	84,585	17,140	26,506	37.32	130	85	37	8	HM596071.1
9	Arcuatopterus Clade	*Sillaphyton insolens*	덕우기름나물	*rps19*	156912*	84,492	17,417	27508*	37.68	130	85	37	8	NC_033344.1
10	Acronema Clade	*Ligusticum delavayi*		*rps19*	155,623	85,066	16,741	26,908	37.55	130	85	37	8	NC_049052
11	Acronema Clade	*Pterygopleurum neurophyllum*	서울개발나물	*rps19*	154,369	84,411	21,210	24,374	37.61	130	85	37	8	NC_033345.1
12	Torilidinae	*Caucalis platycarpos*	헤지호그파슬리	*rps19*	171083*	85042*	17,553	34,244	37.84	130	85	37	8	KX832334.1
13	Daucinae	*Cuminum cyminum*		*rps19*	157,839	83,927	17,598	28,157	37.79	130	85	37	8	NC_046879
14	Daucinae	*Daucus carota*	당근	*rps19*	155,911	84,242	17,571	27,049	37.66	130	85	37	8	NC_008325.1
15	Careae	*Carum carvi*	캐러웨이	*rps3*	155,449	83,672	17,549	27,114	37.77	125	80	37	8	NC_029889.1
16	Pyramidoptereae	*Crithmum maritimum*	록셈파이어	*rpl16*	158,355	85,230	17,139	27,993	37.55	133	88	37	8	NC_015804.1
17	Cachrys clade	*Prangos trifida*		*rpl2*	153,510	86,482	17,446	24,791	37.75	129	84	37	8	NC_037852.1
18	Apieae	*Anethum graveolens*	딜	*rpl2*	153,357	86,506	17,517	24,667	37.65	129	84	37	8	NC_029470.1
19	Apieae	*Foeniculum vulgare*	회향	*rpl2*	153628*	86,659	17,470	24750*	37.65	129	84	37	8	NC_029469.1
20	Apieae	*Petroselinum crispum*	파슬리	*rpl2*	152,890	86,116	17,506	24,634	37.78	129	84	37	8	NC_015821.1

21	**Selineae**	***Angelica acutiloba***	**일당귀**	***ycf2***	**147,074**	**93,367**	**17,573**	**18,067**	**37.53**	**126**	**82**	**36**	**8**	**NC_029391.1**
22	**Selineae**	***Angelica decursiva***	**바디나물**	***ycf2***	**146,719**	**93,256**	**17,497**	**17,983**	**37.56**	**127**	**82**	**37**	**8**	**KT781591.1**
23	**Selineae**	***Angelica nitida***		***ycf2***	**146,512**	**93,298**	**17,950**	**17,632**	**37.48**	**126**	**82**	**36**	**8**	**MF594405.1**
24	**Selineae**	***Arracacia xanthorrhiza***	**아라카차**	***ndhB***	**143,989**	**94,820**	**17,439**	**15,865**	**37.48**	**124**	**81**	**35**	**8**	**NC_032364.1**
25	**Selineae**	***Peucedanum praeruptorum***	**백화전호**	***ycf2***	**147,197**	**92,161**	**17,610**	**18,713**	**37.58**	**125**	**82**	**35**	**8**	**MN016968**
26	**Selineae**	***Peucedanum terebinthaceum***	**기름나물**	***ycf2***	**147,925**	**93,368**	**17,571**	**18,493**	**37.48**	**126**	**82**	**36**	**8**	**NC_053641**
27	**Selineae**	***Glehnia littoralis***	**갯방풍**	***ycf2***	**147,477**	**93,496**	**17,555**	**18,213**	**37.51**	**126**	**82**	**36**	**8**	**NC_034645.1**
28	**Selineae**	***Ledebouriella seseloides***	**방풍**	***ycf2***	**147,830**	**93,202**	**17,324**	**18,652**	**37.53**	**126**	**82**	**36**	**8**	**NC_034643.1**
29	**Selineae**	***Seseli montanum***		***ycf2***	**147,823**	**92,621**	**17,480**	**18,861**	**37.57**	**126**	**82**	**36**	**8**	**NC_027451.1**

Table S1-3. Information regarding *Angelica decursiva* intraspecific plastome polymorphisms

SNP InDels	NCBI	IM
NCBI	0	112
IM	35 (40)	0

Table S2-1a. Summary of *P. japonicum* Intra-specific InDel Polymorphic information

TR:	SSR:	Unique InDel:	Total InDel
51	4	18	73

Table S2-1b. Summary of *P. japonicum* Intra-specific tandem repeat Polymorphism

No.	InDel	Gene position	Quadripartite region	InDel Motif (bp)	Specific	InDel	Pj-1	Pj-2	Pj-3	Pj-4	Pj-5	Pj-6	Pj-7	Pj-8	Pj-9
1	TR	*trn*H-GUG-*psbA*	LSC	AAATATAA (8)	Pj-8	Ins	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**1**
2	TR	*matK-trn*K-UUU	LSC	GTTTATGTTC (10)	Pj-8	Ins	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**1**
3	TR	*rps16-trn*Q-UUG	LSC	ACACAGATATACATATATTTA (21)	Pj-9	Ins	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**2**
4	TR	*rps16-trn*Q-UUG	LSC	AGATATGATTCATAAA (16)	Pj-8	Ins	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**3**	**2**
5	TR	*psbK-psbI*	LSC	ATATAAGGTTAAGG (14)	Pj-8	Ins	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**1**
6	TR	*trn*G-UCC-*trn*R-UCU	LSC	TATTTATA (8)	Pj-8	Del	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**1**	**2**
7	TR	*atpH-atpI*	LSC	Others: ATTACG (6) PJS1: ATTCTG (6)	Pj-8	Ins	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**1**
8	TR	*rpoB-trn*C-GCA	LSC	TGAAAAT (7)	Pj-8	Ins	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**1**
9	TR	*petN-psbM*	LSC	TATAT (5)	Pj-8	Ins	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**1**
10	TR	*petN-psbM*	LSC	TAAAAATAATAGAT (14)	Pj-8	Ins	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**1**
11	TR	*trn*T-GGU-*psbD*	LSC	AAATAGAAC (9)	Pj-8	Del	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**1**	**2**
12	TR	*trn*T-GGU-*psbD*	LSC	TTTTATAA (8)	Pj-8	Del	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**1**	**2**
13	TR	*trn*S-UGA-*psbZ*	LSC	ATATAGGAAGTA (12)	Pj-8	Ins	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**1**
14	TR	*trn*S-UGA-*psbZ*	LSC	TATTTATTAA (10)	Pj-8	Del	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**1**	**2**

15	TR	*ndhJ-ndhK*	LSC	TATTTG (6)	Pj-8	Ins	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**1**
16	TR	*psbE-petL*	LSC	AAGGAAGTACTC (12)	Pj-8	Del	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**1**	**2**
17	TR	*psaJ-rpl33*	LSC	TTTTAGAATAAG (12)	Pj-9	Ins	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**1**
18	TR	*clpP-clpP*	LSC	ATTTGATGTATTCGTGCAT	Pj-8	Ins	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**1**
19	TR	*petD-petD*	LSC	CATATAT (7)	Pj-7	Ins	**2**	**2**	**2**	**2**	**2**	**2**	**3**	**2**	**2**
20	TR	*petD-rpoA*	LSC	TTATCTA (7)	Pj-8	Ins	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**1**
21	TR	*infA/rps8*	LSC	TTTAAA (6)	Pj-8	Del	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**2**	**3**
22	TR	*rpl16/rpl16*	LSC	GGAAAGA (7)	Pj-8	Del	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**1**	**2**
23	TR	*ycf2-trn*L-CAA	LSC	AAATTAT (7)	Pj-8	Ins	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**1**
24	TR	*trn*L-CAA-ndhB	LSC	TAATCC (6)	Pj-8	Ins	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**3**	**2**
25	TR	*ycf15-trn*V-GAC	LSC	TTTTATATCCGTA (13)	Pj-8	Ins	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**3**	**2**
26	TR	*trn*I-GAU-*trn*I-GAU	LSC	ACCGAAAGGAAAAGCGTGAA (20)	Pj-8	Del	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**1**	**2**
27	TR	*rps16/trn*Q-UUG	LSC	AATCAA (6) AATAAA (6)	Pj-7	Ins	**2**	**2**	**2**	**2**	**2**	**2**	**3**	**2**	**3**
28	TR	*atpF/atpH*	LSC	AAAAAATATA (10)	Pj-6	Ins	**1**	**1**	**1**	**1**	**1**	**2**	**1**	**1**	**1**
29	TR	*psaA/ycf3*	LSC	ATTAGCCATTACTAATAG (18)	L-type	Ins	**2**	**2**	**2**	**2**	**2**	**1**	**1**	**1**	**1**
30	TR	*ycf3/trn*S-GGA	LSC	TAACCTAAGTGCAAAAATAGA (21)	Pj-6	Del	**2**	**2**	**2**	**2**	**2**	**1**	**2**	**2**	**2**
31	TR	*rps4/trn*T-UGU	LSC	**(T)**ATTCTATATATATT (15) **(C)**TATATATATATATT (15)	L-type	Ins	**T:2 (C:1)**	**T:2 (C:1)**	**T:2 (C:1)**	**T:2 (C:1)**	**T:2 (C:1)**	**T:1 (C:1)**	**T:1 (C:1)**	**T:1 (C:1)**	**T:1 (C:1)**
32	TR	*ndhC/trn*V-UAC	LSC	CTTAATA (7)	L-type	Ins	**2**	**2**	**2**	**2**	**2**	**1**	**1**	**1**	**1**
33	TR	*trn*W-CCA/*trn*P-UGG	LSC	TTATATATATA (11)	Pj-6	Del	**3**	**3**	**3**	**3**	**3**	**1**	**3**	**2**	**3**
34	TR	*petD/petD*	IRB/LSC	TCATATA (7)	Pj-7	Ins	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**3**	**2**
35	TR	*rps11/rpl36*	IRB/LSC	TCCCCTTTTT (10)	L-type	Ins	**2**	**2**	**2**	**2**	**2**	**1**	**1**	**1**	**1**

36	TR	*rps11/rpl36*	IRB/LSC	ACCTTTTTTA	Pj-8	Ins	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**1**
37	TR	*rpl2/rpl2*	IRB/LSC	AATGTGAAT (9)	Pj-7	Ins	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**1**	**2**
38	TR	*ycf2/*	IRB	******(1) *****(2)****(3) TAGTGA(T)GATAT(T)GATG(C) TAGTGA(C)GATAT(T)GATG(C) TAGTGA(C)GATAT(C)GATG(C) TAGTGA(C)GATAT(C)GATC**(G)**	Pj-7	Ins	**TTC : 2 CTC : 2 CCC : 0 CCG : 1**	**TTC : 2 CTC : 2 CCC : 0 CCG : 1**	**TTC : 2 CTC : 2 CCC : 0 CCG : 1**	**TTC : 2 CTC : 2 CCC : 0 CCG : 1**	**TTC : 2 CTC : 2 CCC : 0 CCG : 1**	**TTC: 2 CTC: 2 CCC: 0 CCG : 1**	**TTC : 2 CTC : 3 CCC : 0 CCG : 1**	**TTC : 0 CTC : 2 CCC : 3 CCG : 1**	**TTC : 2 CTC : 3 CCC : 0 CCG : 1**
39	TR	*ycf2/*	IRB	AAATAGAATGAAATATAT (18)	Pj-8	Ins	1	1	1	1	1	1	1	2	1
40	TR	*ycf2*	IRB	CATACT (6)	Pj-8	Ins	1	1	1	1	1	1	1	2	1
41	TR	*rrn23/*	IRB	ATAAAATAA (9)	Pj-6	Ins	**2**	**2**	**2**	**2**	**2**	**3**	**2**	**3**	**2**
42	TR	*trn*N-GUU/*ycf1*	IRB	TTCTATTTCTTTTCTATATATG (22)	L-type	Ins	**4**	**4**	**4**	**4**	**4**	**2**	**2**	**1**	**2**
43	TR	*ycf1/ndhF*	SSC	TAAAAAAAGTAGTAAATTAAT (21)	Pj-6	Del	**1**	**1**	**1**	**1**	**1**	**2**	**1**	**1**	**1**
44	TR	*ndhF-rpl32*	SSC	GGGTTCTTAAA (11)	S-type	Ins	1	1	1	1	1	2	2	2	2
45	TR	*rpl32-trn*L-UAG	SSC	TAAATATGAATAT (13)	Pj-8	Ins	1	1	1	1	1	1	1	2	1
46	TR	*psaC/ndhE*	SSC	TTTATGAA (8)	S-type	Ins	**1**	**1**	**1**	**1**	**1**	**2**	**2**	**1**	**2**
47	TR	*psaC/ndhE*	SSC	ATATG (5)	L-type	Del	**2**	**2**	**2**	**2**	**2**	**3**	**3**	**3**	**3**
48	TR	*ndhE/ndhG*	SSC	TTAAACTAATTTCCT (15)	S-type	Ins	**1**	**1**	**1**	**1**	**1**	**2 + (10)bp**	**2**	**1**	**2**
49	TR	*ndhA-ndhA*	SSC	AATAAAAAA (9)	Pj-8	Del	2	2	2	2	2	2	2	1	2
50	TR	*ycf1/*	SSC	ATTTATTGT (9)	Pj-7	Ins	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**1**	**1**
51	TR	*rps15-ycf1*	SSC	ATTTTATAG ATTTTATAC	Pj-8	Ins	1	1	1	1	1	1	1	2	1

Table S2-1c. Summary of *P. japonicum* Intra-specific simple repeat polymorphism

No.	InDel	Gene position	CP region	InDel Motif (bp)	Specific	InDel	Pj-1	Pj-2	Pj-3	Pj-4	Pj-5	Pj-6	Pj-7	Pj-8	Pj-9
1	SSR	*trn*R-UCU/*atpA*	LSC	AT (2)	Pj-6	Ins/Del	6	6	6	6	6	7	6; (GT:1)	5	6
2	SSR	*atpF*/*atpH*	LSC	AT (2)	Pj-6/Pj-7	Ins	5	5	5	5	7	7	7	7	7
3	SR	*rps4*/*trn*T-UGU	LSC	TA (2)	Pj-6	Ins	0	0	0	0	0	1	0	0	0
4	SSR	*ndhC*/*trn*V-UAC	LSC	TA (2)	S-type	Ins/Del	3	3	3	3	3	3	4	6	4

Table S2-1d. Summary of *P. japonicum* Intra-specific unique insert polymorphism

No.	InDel	Gene position	CP region	InDel Motif (bp)	Specific	InDel	Pj-1	Pj-2	Pj-3	Pj-4	Pj-5	Pj-6	Pj-7	Pj-8	Pj-9
1	Unique	JLA/*trn*H-GUG	LSC	AATTAATATTCATTTTTTGATTTGTTTGTAACAGGA (36)	Pj-8	Ins	0	0	0	0	0	0	0	1	0
2	Unique	JLA/*trn*H-GUG	LSC	TCGGGGCCTTT (11)	L-type	Del	0	0	0	0	0	1	1	1	1
3	Unique	*trn*H-GUG/*psbA*	LSC	TTTATTTATTAATTCTTTTA (20)	Pj-8	Del	1	1	1	1	1	1	1	0	1
4	Unique	*trn*S-GCU-*trn*G-UCC	LSC	ATTTTCATG (9)	Pj-8	Ins	0	0	0	0	0	0	0	1	0
5	Unique	*trn*T-GGU-*psbD*	LSC	AGT (3)	Pj-8	Ins	0	0	0	0	0	0	0	1	0
6	Unique	*trn*T-GGU-*psbD*	LSC	CTAAATTC (8)	Pj-8	Ins	0	0	0	0	0	0	0	1	0
7	Unique	*trn*T-GGU-*psbD*	LSC	AATAATAAAAAAAAGAAGG (19)	Pj-8	Del	1	1	1	1	1	1	1	0	1
8	Unique	*trn*T-UGU-*trn*L-UAA	LSC	ATATACTT (8)	Pj-8	Del	1	1	1	1	1	1	1	0	1
9	Unique	*ndhC*-*trn*V-UAC	LSC	AGGCTGAG (8)	Pj-8	Ins	0	0	0	0	0	0	0	1	0
10	Unique	*ndhC*-*trn*V-UAC	LSC	ATATATATATATATAA (16)	Pj-8	Ins	1	1	1	1	1	1	1	2	1
11	Unique	*petA*-*psbJ*	LSC	TATTATAT (8)	Pj-8	Del	1	1	1	1	1	1	1	0	1

12	Unique	*clpP-clpP*	LSC	AAAAATAA (8)	Pj-8	Ins	0	0	0	0	0	0	0	1	0
13	Unique	*psbT-psbN*	LSC	AGTTCTAAC (8)	Pj-8	Ins	0	0	0	0	0	0	0	1	0
14	Unique	*ycf2-trn*L-CAA	LSC	ATTAGATAG (9)	Pj-8	Del	1	1	1	1	1	1	1	0	1
15	Unique	*rpl32-trn*L-UAG	LSC	AATAATCTAC (10)	Pj-8	Del	1	1	1	1	1	1	1	0	1
16	Unique	*trn*G-GCC/*trn*fM-CAU	LSC	TTCTTTTCCTCCCACCCC (18)	S-type	Del	1	1	1	1	1	0	0	1	0
17	Unique	*ndhF*/*rpl32*	SSC	AGGGTTCTTAA (11)	L-type	Del	0	0	0	0	0	1	1	0	0
18	Unique	*rpl32*/*trn*L-UAG	SSC	TAAAT (5)	L-type	Del	0	0	0	0	0	1	1	1	1

Table S2-2a. Summary of number of intraspecific SNP on nine *P. japonicum* accessions

Genic region			Intergenic region
Synonymous Mutation	Non-Synonymous Mutation	rRNA	
56	99	1	159

Table S2-2b. Summary of *P. japonicum* Intra-specific SNP information

No.	Pj-1	Pj-2	Pj-3	Pj-4	Pj-5	Pj-6	Pj-7	Pj-8	Pj-9	Gene Pos	Genic/IGS	Specific	quadripartite	Syn/NonSyn
1	G	G	G	G	G	G	G	A	G	*psbA*	Genic	Pj-8	LSC	S
2	T	T	T	T	T	T	T	G	T	*psbA*	Genic	Pj-8	LSC	S
3	G	G	G	G	G	G	A	G	G	*psbA*	Genic	Pj-7	LSC	NS
4	C	C	C	C	C	C	C	A	C	*matK*	Genic	Pj-8	LSC	NS
5	G	G	G	G	G	G	A	G	A	*matK*	Genic	Pj-9/Pj-7	LSC	S
6	C	C	C	C	C	C	T	C	T	*matK*	Genic	Pj-9/Pj-7	LSC	NS
7	A	A	A	A	A	A	A	T	A	*matK*	Genic	Pj-8	LSC	NS
8	G	G	G	G	G	T	G	G	G	*matK*-*trn*K-UUU	IGS	Pj-6	LSC	
9	T	T	T	T	T	T	T	C	T	*trn*K-UUU-*trn*K-UUU	IGS	Pj-8	LSC	
10	A	A	A	A	A	A	C	A	A	*trn*K-UUU-*rps16*	IGS	Pj-7	LSC	
11	A	A	A	A	A	A	A	C	A	*trn*K-UUU-*rps16*	IGS	Pj-8	LSC	
12	A	A	A	A	A	A	A	C	A	*trn*K-UUU-*rps16*	IGS	Pj-8	LSC	
13	C	C	C	C	C	C	C	T	C	*rps16*-*rps16*	IGS	Pj-8	LSC	
14	A	A	A	A	A	A	A	C	A	*rps16*-*rps16*	IGS	Pj-8	LSC	
15	T	T	T	T	A	T	T	T	T	*rps16*-*trn*Q-UUG	IGS	Pj-5	LSC	
16	G	G	G	G	G	G	G	G	A	*rps16*-*trn*Q-UUG	IGS	Pj-9	LSC	

No.	Pj-1	Pj-2	Pj-3	Pj-4	Pj-5	Pj-6	Pj-7	Pj-8	Pj-9	Gene Pos	Genic/IGS	Specific	quadripartite	Syn/NonSyn
17	T	T	T	T	T	T	G	T	T	*rps16-trn*Q-UUG	IGS	Pj-7	LSC	
18	T	T	T	T	T	T	T	G	T	*rps16-trn*Q-UUG	IGS	Pj-8	LSC	
19	T	T	T	T	T	C	C	C	C	*rps16-trn*Q-UUG	IGS	LTYPE	LSC	
20	A	A	A	A	A	A	A	T	A	*trn*Q-UUG-*psbK*	IGS	Pj-8	LSC	
21	A	A	A	A	A	A	A	G	A	*trn*Q-UUG-*psbK*	IGS	Pj-8	LSC	
22	G	G	G	G	G	G	A	G	G	*psbK*-	Genic	Pj-7	LSC	NS
23	C	C	C	C	C	C	C	T	C	*psbK*-	Genic	Pj-8	LSC	NS
24	G	G	G	G	G	G	G	A	G	*psbK-psbI*	IGS	Pj-8	LSC	
25	C	C	C	C	C	C	C	A	C	*psbK-psbI*	IGS	Pj-8	LSC	
26	G	G	G	G	G	G	G	C	G	*psbK-psbI*	IGS	Pj-8	LSC	
27	A	A	A	A	A	A	A	C	A	*psbK-psbI*	IGS	Pj-8	LSC	
28	A	A	A	A	A	A	A	C	A	*psbI-trn*S-GCU	IGS	Pj-8	LSC	
29	G	G	G	G	G	G	A	G	G	*trn*S-GCU-*trn*G-UCC	IGS	Pj-7	LSC	
30	G	G	G	G	G	A	G	G	G	*trn*S-GCU-*trn*G-UCC	IGS	Pj-6	LSC	
31	T	T	T	T	T	C	C	C	C	*trn*G-UCC-*trn*G-UCC	IGS	LTYPE	LSC	
32	A	A	A	A	A	A	G	A	A	*trn*R-UCU-*atpA*	IGS	Pj-7	LSC	
33	C	C	C	C	C	C	C	T	C	*atpA*-	Genic	Pj-8	LSC	S
34	C	C	C	C	C	A	C	C	C	*atpA-atpF*	IGS	Pj-6	LSC	
35	T	T	T	T	T	T	G	T	G	*atpF*-	Genic	Pj-9/Pj-7	LSC	NS
36	G	G	G	G	G	G	G	A	G	*atpF-atpF*	IGS	Pj-8	LSC	
37	A	A	A	A	A	A	A	C	A	atpF-*atpF*	IGS	Pj-8	LSC	
38	G	G	G	G	G	G	G	G	A	*atpF-atpF*	IGS	Pj-9	LSC	
39	T	T	T	T	T	G	T	T	T	*atpF-atpH*	IGS	Pj-6	LSC	

No.	Pj-1	Pj-2	Pj-3	Pj-4	Pj-5	Pj-6	Pj-7	Pj-8	Pj-9	Gene Pos	Genic/IGS	Specific	quadripartite	Syn/NonSyn
40	C	C	C	C	C	C	C	T	C	*atpH-atpI*	IGS	Pj-8	LSC	
41	C	C	C	C	C	C	C	G	C	*atpH-atpI*	IGS	Pj-8	LSC	
42	T	T	T	T	T	T	T	C	T	*atpH-atpI*	IGS	Pj-8	LSC	
43	A	A	A	A	A	A	A	C	A	*atpH-atpI*	IGS	Pj-8	LSC	
44	C	C	C	C	C	C	C	T	C	*atpH-atpI*	IGS	Pj-8	LSC	
45	G	G	G	G	G	T	T	T	T	*atpI-*	Genic	LTYPE	LSC	S
46	C	C	C	C	C	C	C	T	C	*atpI-*	Genic	Pj-8	LSC	NS
47	T	T	T	T	T	T	T	T	C	*rps2-*	Genic	Pj-9	LSC	NS
48	C	C	C	C	C	T	C	C	C	*rps2-*	Genic	Pj-6	LSC	S
49	G	G	G	G	G	G	G	T	G	*rpo*C2-	Genic	Pj-8	LSC	NS
50	C	C	C	C	C	T	T	C	T	*rpo*C2-	Genic	Pj-9/Pj-7/Pj-6	LSC	NS
51	A	A	A	A	A	A	A	C	A	*rpo*C2-	Genic	Pj-8	LSC	NS
52	T	T	T	T	T	C	C	C	C	*rpo*C2-	Genic	LTYPE	LSC	S
53	T	T	T	T	T	T	T	G	T	*rpo*C2-	Genic	Pj-8	LSC	NS
54	A	A	A	A	A	A	A	C	A	*rpo*C2-	Genic	Pj-8	LSC	NS
55	A	A	A	A	G	G	G	G	G	*rpo*C2-	Genic	Pj-5/STYPE	LSC	NS
56	A	A	A	A	A	A	A	G	A	*rpo*C2-	Genic	Pj-8	LSC	NS
57	T	T	T	T	T	T	T	T	C	*rpo*C2-	Genic	Pj-9	LSC	NS
58	G	G	G	G	G	G	G	A	G	*rpo*C2-	Genic	Pj-8	LSC	NS
59	C	C	C	C	C	C	T	C	C	*rpo*C2-	Genic	Pj-7	LSC	NS
60	T	T	T	T	T	T	T	C	T	*rpo*C1 exon 2-	Genic	Pj-8	LSC	S
61	G	G	G	G	G	G	G	A	G	*rpo*C1 exon 2-	Genic	Pj-8	LSC	NS
62	C	C	C	C	C	C	A	C	A	*rpo*C1-*rpo*C1	IGS	Pj-9/Pj-7	LSC	

No.	Pj-1	Pj-2	Pj-3	Pj-4	Pj-5	Pj-6	Pj-7	Pj-8	Pj-9	Gene Pos	Genic/IGS	Specific	quadripartite	Syn/NonSyn
63	T	T	T	T	T	A	T	T	T	*rpo*C1-*rpo*C1	IGS	Pj-6	LSC	
64	T	T	T	T	T	T	T	G	T	*rpo*C1-*rpo*C1	IGS	Pj-8	LSC	
65	T	T	T	T	T	T	T	T	C	*rpo*C1-*rpo*C1	IGS	Pj-9	LSC	
66	C	C	C	C	C	C	C	T	C	*rpo*C1 exon 1-	Genic	Pj-8	LSC	S
67	C	C	C	C	T	C	C	C	C	*rpo*C1-*rpo*B	IGS	Pj-5	LSC	
68	C	C	C	C	C	C	C	G	C	*rpo*B-	Genic	Pj-8	LSC	NS
69	T	T	T	T	A	T	T	T	T	*rpo*B-	Genic	Pj-5	LSC	NS
70	T	T	T	T	T	G	T	T	T	*rpo*B-	Genic	Pj-6	LSC	NS
71	C	C	C	C	C	C	C	T	C	*rpo*B-*trn*C-GCA	IGS	Pj-8	LSC	
72	C	C	C	C	C	A	A	C	A	*rpo*B-*trn*C-GCA	IGS	Pj-9/Pj-7/Pj-6	LSC	
73	A	A	A	A	A	A	A	G	A	*rpo*B-*trn*C-GCA	IGS	Pj-8	LSC	
74	T	T	T	T	T	T	T	A	T	*rpo*B-*trn*C-GCA	IGS	Pj-8	LSC	
75	C	C	C	C	C	A	A	A	A	*rpo*B-*trn*C-GCA	IGS	LTYPE	LSC	
76	C	C	C	C	C	C	C	T	C	*rpo*B-*trn*C-GCA	IGS	Pj-8	LSC	
77	G	G	G	G	G	T	T	G	T	*rpo*B-*trn*C-GCA	IGS	Pj-9/Pj-7/Pj-6	LSC	
78	A	A	A	A	A	G	G	G	G	*trn*C-GCA-*petN*	IGS	LTYPE	LSC	
79	C	C	C	C	C	C	T	C	C	*trn*C-GCA-*petN*	IGS	Pj-7	LSC	
80	C	C	C	C	C	C	C	A	C	*petN*-*psbM*	IGS	Pj-8	LSC	
81	T	T	T	T	T	T	T	G	T	*petN*-*psbM*	IGS	Pj-8	LSC	
82	G	G	G	G	G	G	G	A	G	*petN*-*psbM*	IGS	Pj-8	LSC	
83	G	G	G	G	G	G	G	T	G	*petN*-*psbM*	IGS	Pj-8	LSC	
84	T	T	T	T	T	T	T	C	T	*petN*-*psbM*	IGS	Pj-8	LSC	
85	C	C	C	C	C	C	C	T	C	*petN*-*psbM*	IGS	Pj-8	LSC	

No.	Pj-1	Pj-2	Pj-3	Pj-4	Pj-5	Pj-6	Pj-7	Pj-8	Pj-9	Gene Pos	Genic/IGS	Specific	quadripartite	Syn/NonSyn
86	A	A	A	A	A	A	-	C	A	*petN-psbM*	IGS	Pj-8	LSC	
87	A	A	A	A	A	A	A	C	A	*psbM-trn*E-UUC	IGS	Pj-8	LSC	
88	A	A	A	A	A	A	A	C	A	*psbM-trn*E-UUC	IGS	Pj-8	LSC	
89	G	G	G	G	G	G	G	A	G	*trn*D-GUC-*trn*T-GGU	IGS	Pj-8	LSC	
90	G	G	G	G	G	G	A	G	G	*trn*T-GGU-*psbD*	IGS	Pj-7	LSC	
91	T	T	T	T	T	G	G	G	G	*trn*T-GGU-*psbD*	IGS	LTYPE	LSC	
92	T	T	T	T	T	T	T	G	T	*trn*T-GGU-*psbD*	IGS	Pj-8	LSC	
93	T	T	T	T	T	T	T	G	T	*trn*T-GGU-*psbD*	IGS	Pj-8	LSC	
94	C	C	C	C	T	C	C	C	C	*psbC-*	Genic	Pj-5	LSC	S
95	C	C	C	C	C	C	C	C	T	*psbC-*	Genic	Pj-9	LSC	S
96	G	G	G	G	G	G	G	A	G	*trn*S-UGA-*psbZ*	IGS	Pj-8	LSC	
97	A	A	A	A	A	A	A	C	A	*psbZ-trn*G-GCC	IGS	Pj-8	LSC	
98	C	C	C	C	C	G	G	G	G	*psbZ-trn*G-GCC	IGS	LTYPE	LSC	
99	G	G	G	G	G	G	G	G	A	*trn*G-GCC-*trn*fM-CAU	IGS	Pj-9	LSC	
100	C	C	C	C	C	A	A	C	A	*trn*G-GCC-*trn*fM-CAU	IGS	Pj-9/Pj-7/Pj-6	LSC	
101	A	A	A	A	A	T	T	A	T	*trn*G-GCC-*trn*fM-CAU	IGS	Pj-9/Pj-7/Pj-6	LSC	
102	T	T	T	T	T	A	A	T	A	*trn*G-GCC-*trn*fM-CAU	IGS	Pj-9/Pj-7/Pj-6	LSC	
103	C	C	C	C	C	C	C	T	C	*rps14-psaB*	IGS	Pj-8	LSC	
104	A	A	A	A	A	A	A	G	A	*psaB-*	Genic	Pj-8	LSC	S
105	T	T	T	T	T	T	T	C	T	*psaB-*	Genic	Pj-8	LSC	S
106	G	G	G	G	G	G	G	A	G	*psaB-*	Genic	Pj-8	LSC	NS
107	G	G	G	G	G	G	G	A	G	*psaB-*	Genic	Pj-8	LSC	S
108	A	A	A	A	A	G	G	G	G	*psaA-*	Genic	LTYPE	LSC	S

No.	Pj-1	Pj-2	Pj-3	Pj-4	Pj-5	Pj-6	Pj-7	Pj-8	Pj-9	Gene Pos	Genic/IGS	Specific	quadripartite	Syn/NonSyn
109	G	G	G	G	G	G	G	G	A	*psaA*-	Genic	Pj-9	LSC	S
110	T	T	T	T	T	C	C	C	C	*psaA*-	Genic	LTYPE	LSC	NS
111	A	A	A	A	A	G	A	A	A	*psaA-ycf3*	IGS	Pj-6	LSC	
112	C	C	C	C	C	C	C	A	C	*psaA-ycf3*	IGS	Pj-8	LSC	
113	T	T	T	T	T	T	T	A	T	*psaA-ycf3*	IGS	Pj-8	LSC	
114	G	G	G	G	G	G	T	G	T	*psaA-ycf3*	IGS	Pj-9/Pj-7	LSC	
115	C	C	C	C	C	C	C	A	C	*ycf3-trn*S-GGA	IGS	Pj-8	LSC	
116	C	C	C	C	C	C	C	T	C	*trn*S-GGA-*rps4*	IGS	Pj-8	LSC	
117	C	C	C	C	C	C	T	C	C	*rps4*-	Genic	Pj-7	LSC	S
118	A	A	A	A	A	A	A	C	A	*rps4-trn*T-UGU	IGS	Pj-8	LSC	
119	T	T	T	T	T	C	C	C	C	*rps4-trn*T-UGU	IGS	LTYPE	LSC	
120	T	T	T	T	T	T	T	A	T	*rps4-trn*T-UGU	IGS	Pj-8	LSC	
121	T	T	T	T	T	T	T	G	T	*rps4-trn*T-UGU	IGS	Pj-8	LSC	
122	G	G	G	G	G	G	G	T	G	*trn*T-UGU-*trn*L-UAA	IGS	Pj-8	LSC	
123	T	T	T	T	T	T	T	G	T	*trn*T-UGU-*trn*L-UAA	IGS	Pj-8	LSC	
124	G	G	G	G	G	G	G	T	G	*trn*T-UGU-*trn*L-UAA	IGS	Pj-8	LSC	
125	C	C	C	C	C	C	C	T	C	*trn*T-UGU-*trn*L-UAA	IGS	Pj-8	LSC	
126	T	T	T	T	T	T	T	A	T	*trn*F-GAA-*ndhJ*	IGS	Pj-8	LSC	
127	T	T	T	T	T	T	T	A	T	*trn*F-GAA-*ndhJ*	IGS	Pj-8	LSC	
128	T	T	T	T	T	C	C	C	C	*ndhJ*-	Genic	LTYPE	LSC	NS
129	C	C	C	C	C	A	C	C	C	*ndhK*-	Genic	Pj-6	LSC	NS
130	T	T	T	T	T	C	C	C	C	*ndhK*-	Genic	LTYPE	LSC	NS
131	G	G	G	G	G	G	G	T	G	*ndhC*-	Genic	Pj-8	LSC	NS

No.	Pj-1	Pj-2	Pj-3	Pj-4	Pj-5	Pj-6	Pj-7	Pj-8	Pj-9	Gene Pos	Genic/IGS	Specific	quadripartite	Syn/NonSyn
132	T	T	T	T	T	T	T	G	T	*ndhC-trn*V-UAC	IGS	Pj-8	LSC	
133	C	C	C	C	C	C	C	A	C	*ndhC-trn*V-UAC	IGS	Pj-8	LSC	
134	T	T	T	T	T	T	T	C	T	*ndhC-trn*V-UAC	IGS	Pj-8	LSC	
135	C	C	C	C	C	C	C	C	A	*ndhC-trn*V-UAC	IGS	Pj-9	LSC	
136	C	C	C	C	C	C	C	T	C	*trn*V-UAC-*trn*V-UAC	IGS	Pj-8	LSC	
137	C	C	C	C	C	C	C	A	C	*atpB-rbcL*	IGS	Pj-8	LSC	
138	C	C	C	C	C	C	C	C	A	*atpB-rbcL*	IGS	Pj-9	LSC	
139	A	A	A	A	A	G	G	G	G	*atpB-rbcL*	IGS	LTYPE	LSC	
140	C	C	C	C	C	C	C	T	C	*atpB-rbcL*	IGS	Pj-8	LSC	
141	C	C	C	C	T	C	C	C	C	*atpB-rbcL*	IGS	Pj-5	LSC	
142	G	G	G	G	G	G	G	T	G	*accD-*	Genic	Pj-8	LSC	NS
143	C	C	C	C	C	C	C	T	C	*accD-*	Genic	Pj-8	LSC	S
144	A	A	A	A	A	A	A	C	A	*accD-*	Genic	Pj-8	LSC	NS
145	A	A	A	A	A	G	G	G	G	*accD-*	Genic	LTYPE	LSC	NS
146	G	G	G	G	G	G	G	T	G	*accD-psaI*	IGS	Pj-8	LSC	
147	T	T	T	T	T	T	T	C	T	*accD-psaI*	IGS	Pj-8	LSC	
148	A	A	A	A	A	A	A	G	A	*psaI-ycf4*	IGS	Pj-8	LSC	
149	T	T	T	T	T	T	T	C	T	*ycf4-*	Genic	Pj-8	LSC	NS
150	C	C	C	C	C	A	C	C	C	*ycf4-*	Genic	Pj-6	LSC	NS
151	G	G	G	G	G	G	G	T	G	*ycf4-*	Genic	Pj-8	LSC	NS
152	G	G	G	G	G	G	G	A	G	*cemA-*	Genic	Pj-8	LSC	NS
153	T	T	T	T	T	A	T	T	T	*petA-psbJ*	IGS	Pj-6	LSC	
154	T	T	T	T	T	A	T	T	T	*petA-psbJ*	IGS	Pj-6	LSC	

No.	Pj-1	Pj-2	Pj-3	Pj-4	Pj-5	Pj-6	Pj-7	Pj-8	Pj-9	Gene Pos	Genic/IGS	Specific	quadripartite	Syn/NonSyn
155	C	C	C	C	C	A	C	C	C	*petA-psbJ*	IGS	Pj-6	LSC	
156	T	T	T	T	T	T	T	C	T	*psbE-*	Genic	Pj-8	LSC	S
157	C	C	C	C	C	C	A	C	A	*psbE-petL*	IGS	Pj-9/Pj-7	LSC	
158	T	T	T	T	T	T	T	A	T	*petG-trn*W-CCA	IGS	Pj-8	LSC	
159	A	A	A	A	A	G	G	G	G	*petG-trn*W-CCA	IGS	LTYPE	LSC	
160	A	A	A	A	A	G	G	G	G	*trn*P-UGG-*psaJ*	IGS	LTYPE	LSC	
161	C	C	C	C	C	C	C	A	C	*rpl33-rps18*	IGS	Pj-8	LSC	
162	T	T	T	T	T	T	T	C	T	*rps18-*	Genic	Pj-8	LSC	NS
163	T	T	T	T	T	C	C	C	C	*rps18-rpl20*	IGS	LTYPE	LSC	
164	G	G	G	G	G	A	G	G	G	*clpP* exon3-	Genic	Pj-6	LSC	S
165	C	C	C	C	C	C	C	T	C	*clpP* exon3-	Genic	Pj-8	LSC	S
166	T	T	T	T	T	T	T	A	T	*clpP* exon2-*clpP* exon1	IGS	Pj-8	LSC	
167	T	T	T	T	T	T	T	A	T	*clpP* exon2-*clpP* exon1	IGS	Pj-8	LSC	
168	T	T	T	T	T	T	T	A	T	*clpP* exon2-*clpP* exon1	IGS	Pj-8	LSC	
169	A	A	A	A	A	A	A	C	A	*clpP* exon2-*clpP* exon1	IGS	Pj-8	LSC	
170	G	G	G	G	G	G	G	T	G	*clpP-psbB*	IGS	Pj-8	LSC	
171	G	G	G	G	G	G	T	G	G	*psbB-*	Genic	Pj-7	LSC	NS
172	G	G	G	G	G	G	G	A	G	*psbB-*	Genic	Pj-8	LSC	S
173	G	G	G	G	T	G	G	G	G	*psbB-*	Genic	Pj-5	LSC	NS
174	T	T	T	T	T	T	T	C	T	*psbT-*	Genic	Pj-8	LSC	S
175	A	A	A	A	A	A	G	A	A	*petB-petB*	IGS	Pj-7	LSC/IRB	
176	G	G	G	G	G	G	G	A	G	*petB-petB*	IGS	Pj-8	LSC/IRB	
177	G	G	G	G	G	G	G	A	G	*petB-petB*	IGS	Pj-8	LSC/IRB	

No.	Pj-1	Pj-2	Pj-3	Pj-4	Pj-5	Pj-6	Pj-7	Pj-8	Pj-9	Gene Pos	Genic/IGS	Specific	quadripartite	Syn/NonSyn
178	C	C	C	C	C	T	C	C	C	*petB* exon2-	Genic	Pj-6	LSC/IRB	S
179	A	A	A	A	A	A	A	G	A	*petB* exon2-	Genic	Pj-8	LSC/IRB	S
180	C	C	C	C	C	C	T	C	C	*petD-petD*	IGS	Pj-7	LSC/IRB	
181	A	A	A	A	A	A	A	G	A	*petD-petD*	IGS	Pj-8	LSC/IRB	
182	C	C	C	C	C	C	C	A	C	*petD-rpoA*	IGS	Pj-8	LSC/IRB	
183	T	T	T	T	T	G	G	T	G	*rpoA-*	Genic	Pj-9/Pj-7/Pj-6	LSC/IRB	NS
184	T	T	T	T	T	T	T	G	T	*rpoA-*	Genic	Pj-8	LSC/IRB	S
185	T	T	T	T	T	T	T	C	T	*rpoA-*	Genic	Pj-8	LSC/IRB	NS
186	C	C	C	C	C	C	C	T	C	*rps11-rpl36*	IGS	Pj-8	LSC/IRB	
187	C	C	C	C	C	C	C	A	C	*rps11-rpl36*	IGS	Pj-8	LSC/IRB	
188	T	T	T	T	T	C	C	T	C	*rps11-rpl36*	IGS	Pj-9/Pj-7/Pj-6	LSC/IRB	
189	A	A	A	A	A	A	A	G	A	*infA-*	Genic	Pj-8	LSC/IRB	S
190	A	A	A	A	A	A	A	T	A	*rps8-*	Genic	Pj-8	LSC/IRB	NS
191	T	T	T	T	T	T	T	C	T	*rpl16* exon2-	Genic	Pj-8	LSC/IRB	S
192	C	C	C	C	C	C	C	A	C	*rpl16-rpl16*	IGS	Pj-8	LSC/IRB	
193	A	A	A	A	A	A	A	C	A	*rpl16-rpl16*	IGS	Pj-8	LSC/IRB	
194	G	G	G	G	G	G	G	A	G	*rpl16-rpl16*	IGS	Pj-8	LSC/IRB	
195	G	G	G	G	G	T	T	T	T	*rpl16-rpl16*	IGS	LTYPE	LSC/IRB	
196	C	C	C	C	C	C	C	T	C	*rpl16-rpl16*	IGS	Pj-8	LSC/IRB	
197	C	C	C	C	C	T	T	C	T	*rpl16-rpl16*	IGS	Pj-9/Pj-7/Pj-6	LSC/IRB	
198	G	G	G	G	G	G	G	A	G	*rpl16-rpl16*	IGS	Pj-8	LSC/IRB	
199	T	T	T	T	T	T	T	C	T	*rps3-*	Genic	Pj-8	LSC/IRB	S
200	C	C	C	C	C	C	C	A	C	*rps3-*	Genic	Pj-8	LSC/IRB	NS

No.	Pj-1	Pj-2	Pj-3	Pj-4	Pj-5	Pj-6	Pj-7	Pj-8	Pj-9	Gene Pos	Genic/IGS	Specific	quadripartite	Syn/NonSyn
201	A	A	A	A	A	C	A	A	A	*rps19*-	Genic	Pj-6	LSC/IRB	S
202	C	C	C	C	C	T	C	C	C	*rpl2* exon2-	Genic	Pj-6	LSC/IRB	NS
203	C	C	C	C	C	T	C	C	C	*rpl2* exon2-	Genic	Pj-6	LSC/IRB	NS
204	T	T	T	T	T	G	T	T	T	*rpl2-rpl2*	IGS	Pj-6	LSC/IRB	
205	C	C	C	C	C	C	C	A	C	*rpl23*-	Genic	Pj-8	LSC/IRB	NS
206	C	C	C	C	C	C	C	T	C	*ycf2*-	Genic	Pj-8	LSC/IRB	S
207	T	T	T	T	T	T	T	G	T	*ycf2*-	Genic	Pj-8	LSC/IRB	S
208	G	G	G	G	G	G	G	A	G	*ycf2*-	Genic	Pj-8	LSC/IRB	NS
209	T	T	T	T	T	T	T	C	T	*ycf2*-	Genic	Pj-8	LSC/IRB	S
210	G	G	G	G	G	A	G	G	G	*ycf2*-	Genic	Pj-6	LSC/IRB	NS
211	C	C	C	C	C	C	A	C	C	*ycf2*-	Genic	Pj-7	LSC/IRB	NS
212	G	G	G	G	G	A	G	G	G	*ycf2*-	Genic	Pj-6	LSC/IRB	S
213	C	C	C	C	C	C	C	T	C	*ycf2*-	Genic	Pj-8	LSC/IRB	NS
214	T	T	T	T	T	C	T	T	T	*ycf2*-	Genic	Pj-6	LSC/IRB	S
215	T	T	T	T	T	G	G	G	G	*ycf2*-	Genic	LTYPE	LSC/IRB	NS
216	A	A	A	A	A	A	A	C	A	*ycf2*-	Genic	Pj-8	LSC/IRB	NS
217	A	A	A	A	A	A	A	G	A	*ycf2*-	Genic	Pj-8	LSC/IRB	NS
218	G	G	G	G	G	A	A	G	A	*ycf2*-	Genic	Pj-9/Pj-7/Pj-6	LSC/IRB	NS
219	C	C	C	C	C	C	C	T	C	*ycf2*-	Genic	Pj-8	LSC/IRB	S
220	A	A	A	A	A	A	A	G	A	*ycf2*-	Genic	Pj-8	LSC/IRB	NS
221	A	A	A	A	A	A	A	G	A	*ycf2*-	Genic	Pj-8	LSC/IRB	S
222	A	A	A	A	A	A	A	G	A	*ycf2*-	Genic	Pj-8	LSC/IRB	NS
223	C	C	C	C	C	C	C	T	C	*ycf2*-	Genic	Pj-8	LSC/IRB	NS

No.	Pj-1	Pj-2	Pj-3	Pj-4	Pj-5	Pj-6	Pj-7	Pj-8	Pj-9	Gene Pos	Genic/IGS	Specific	quadripartite	Syn/NonSyn
224	A	A	A	A	A	A	A	G	A	*ycf2*-	Genic	Pj-8	LSC/IRB	S
225	C	C	C	C	C	C	C	T	C	*ycf2*-	Genic	Pj-8	LSC/IRB	NS
226	T	T	T	T	T	T	T	C	T	*ycf2*-	Genic	Pj-8	LSC/IRB	S
227	T	T	T	T	T	T	T	C	T	*ycf2*-	Genic	Pj-8	LSC/IRB	S
228	T	T	T	T	T	T	T	C	T	*ycf2*-	Genic	Pj-8	LSC/IRB	S
229	T	T	T	T	T	T	T	C	T	*ycf2*-	Genic	Pj-8	LSC/IRB	S
230	C	C	C	C	C	C	C	A	C	*ycf2*-	Genic	Pj-8	LSC/IRB	NS
231	A	A	A	A	A	A	A	C	A	*ycf2*-	Genic	Pj-8	LSC/IRB	NS
232	C	C	C	C	C	A	A	C	A	*ycf2*-	Genic	Pj-9/Pj-7/Pj-6	LSC/IRB	NS
233	G	G	G	G	G	G	G	C	G	*ycf2*-	Genic	Pj-8	LSC/IRB	NS
234	A	A	A	A	A	A	A	C	A	*ycf2*-	Genic	Pj-8	LSC/IRB	S
235	T	T	T	T	T	T	T	G	T	*ycf2*-	Genic	Pj-8	LSC/IRB	NS
236	G	G	G	G	G	A	A	G	A	*ycf2*-	Genic	Pj-9/Pj-7/Pj-6	IRB	NS
237	A	A	A	A	A	A	A	C	A	*ycf2*-	Genic	Pj-8	IRB	NS
238	G	G	G	G	G	T	T	T	T	*ycf2*-*trn*L-CAA	IGS	LTYPE	IRB	
239	C	C	C	C	C	C	C	A	C	*ycf2*-*trn*L-CAA	IGS	Pj-8	IRB	
240	T	T	T	T	T	T	T	C	T	*ycf2*-*trn*L-CAA	IGS	Pj-8	IRB	
241	A	A	A	A	A	C	C	C	C	*ycf2*-*trn*L-CAA	IGS	LTYPE	IRB	
242	T	T	T	T	T	T	T	C	T	*ycf2*-*trn*L-CAA	IGS	Pj-8	IRB	
243	A	A	A	A	A	A	A	G	A	*trn*L-CAA-*ndhB*	IGS	Pj-8	IRB	
244	C	C	C	C	C	C	C	T	C	*ndhB* exon2-	Genic	Pj-8	IRB	NS
245	A	A	A	A	A	A	A	C	A	*ndhB*-*rps7*	IGS	Pj-8	IRB	
246	G	G	G	G	G	A	A	G	A	*rps7*-	Genic	Pj-9/Pj-7/Pj-6	IRB	NS

No.	Pj-1	Pj-2	Pj-3	Pj-4	Pj-5	Pj-6	Pj-7	Pj-8	Pj-9	Gene Pos	Genic/IGS	Specific	quadripartite	Syn/NonSyn
247	T	T	T	T	T	G	G	G	G	*rps12-ycf15*	IGS	LTYPE	IRB	
248	T	T	T	T	T	T	T	G	T	*rrn16-*	rRNA	Pj-8	IRB	X
249	T	T	T	T	T	T	T	G	T	*trn*A-UGC-*trn*A-UGC	IGS	Pj-8	IRB	
250	A	A	A	A	A	A	A	C	A	*trn*N-GUU-*ycf1*	IGS	Pj-8	IRB	
251	G	G	G	G	G	G	G	T	G	*ndhF-*	Genic	Pj-8	SSC	NS
252	C	C	C	C	C	C	T	C	T	*ndhF-*	Genic	Pj-9/Pj-7	SSC	S
253	C	C	C	C	C	C	A	C	A	*ndhF-*	Genic	Pj-9/Pj-7	SSC	NS
254	C	C	C	C	C	C	A	T	C	*ndhF-*	Genic	Pj-8/Pj-7	SSC	NS
255	C	C	C	C	C	C	C	T	C	*ndhF-*	Genic	Pj-8	SSC	NS
256	A	A	A	A	A	A	A	T	A	*ndhF-*	Genic	Pj-8	SSC	NS
257	T	T	T	T	T	T	T	C	T	*ndhF-*	Genic	Pj-8	SSC	NS
258	G	G	G	G	G	G	G	C	G	*ndhF-*	Genic	Pj-8	SSC	NS
259	G	G	G	G	G	G	G	A	G	*ndhF-*	Genic	Pj-8	SSC	NS
260	G	G	G	G	G	G	G	A	G	*ndhF-*	Genic	Pj-8	SSC	NS
261	A	A	A	A	A	A	A	C	A	*ndhF-*	Genic	Pj-8	SSC	NS
262	T	T	T	T	T	T	T	G	T	*ndhF-*	Genic	Pj-8	SSC	S
263	G	G	G	G	G	G	G	A	G	*ndhF-*	Genic	Pj-8	SSC	S
264	C	C	C	C	C	C	C	A	C	*ndhF-*	Genic	Pj-8	SSC	NS
265	C	C	C	C	C	T	T	T	T	*ndhF-*	Genic	LTYPE	SSC	NS
266	G	G	G	G	G	G	G	C	G	*ndhF-rpl32*	IGS	Pj-8	SSC	
267	T	T	T	T	T	T	T	A	T	*ndhF-rpl32*	IGS	Pj-8	SSC	
268	C	C	C	C	C	C	C	A	C	*ndhF-rpl32*	IGS	Pj-8	SSC	
269	G	G	G	G	G	G	G	T	G	*ndhF-rpl32*	IGS	Pj-8	SSC	

No.	Pj-1	Pj-2	Pj-3	Pj-4	Pj-5	Pj-6	Pj-7	Pj-8	Pj-9	Gene Pos	Genic/IGS	Specific	quadripartite	Syn/NonSyn
270	G	G	G	G	G	G	G	A	G	*ndhF-rpl32*	IGS	Pj-8	SSC	
271	A	A	A	A	A	A	A	C	A	*ndhF-rpl32*	IGS	Pj-8	SSC	
272	C	C	C	C	C	C	C	T	C	*rpl32-trn*L-UAG	IGS	Pj-8	SSC	
273	T	T	T	T	T	T	T	G	T	*rpl32-trn*L-UAG	IGS	Pj-8	SSC	
274	C	C	C	C	C	C	T	C	C	*rpl32-trn*L-UAG	IGS	Pj-7	SSC	
275	A	A	A	A	A	A	C	A	A	*trn*L-UAG-*ccsA*	IGS	Pj-7	SSC	
276	T	T	T	T	T	T	T	G	T	*trn*L-UAG-*ccsA*	IGS	Pj-8	SSC	
277	C	C	C	C	C	C	C	T	C	*ccsA-*	Genic	Pj-8	SSC	NS
278	C	C	C	C	C	C	C	A	C	*ccsA-*	Genic	Pj-8	SSC	NS
279	C	C	C	C	C	C	C	T	C	*ccsA-*	Genic	Pj-8	SSC	S
280	G	G	G	G	G	G	G	T	G	*ccsA-ndhD*	IGS	Pj-8	SSC	
281	G	G	G	G	G	G	G	A	G	*ndhD-*	Genic	Pj-8	SSC	S
282	G	G	G	G	G	G	A	G	G	*ndhD-*	Genic	Pj-7	SSC	NS
283	A	A	A	A	A	A	A	C	A	*ndhD-*	Genic	Pj-8	SSC	S
284	C	C	C	C	C	A	A	C	A	*ndhE-ndhG*	IGS	Pj-9/Pj-7/Pj-6	SSC	
285	G	G	G	G	G	G	T	G	G	*ndhG-*	Genic	Pj-7	SSC	NS
286	A	A	A	A	A	A	A	C	A	*ndhG-ndhI*	IGS	Pj-8	SSC	
287	C	C	C	C	C	C	C	A	C	*ndhI-*	Genic	Pj-8	SSC	NS
288	A	A	A	A	A	A	A	G	A	*ndhI-*	Genic	Pj-8	SSC	NS
289	C	C	C	C	C	C	C	T	C	*ndhA-ndhA*	IGS	Pj-8	SSC	
290	C	C	C	C	C	C	A	C	C	*ndhA-ndhA*	IGS	Pj-7	SSC	
291	G	G	G	G	G	G	G	A	G	*ndhA-ndhA*	IGS	Pj-8	SSC	
292	A	A	A	A	A	G	G	G	G	*ndhA-ndhA*	IGS	Pj-8	SSC	

No.	Pj-1	Pj-2	Pj-3	Pj-4	Pj-5	Pj-6	Pj-7	Pj-8	Pj-9	Gene Pos	Genic/IGS	Specific	quadripartite	Syn/NonSyn
293	G	G	G	G	G	G	G	T	G	*ndhA* exon1-	Genic	Pj-8	SSC	NS
294	C	C	C	C	C	C	C	A	C	*ndhA* exon1-	Genic	Pj-8	SSC	NS
295	A	A	A	A	A	A	A	C	A	*ndhH*-	Genic	Pj-8	SSC	NS
296	G	G	G	G	G	G	G	T	G	*ndhH*-	Genic	Pj-8	SSC	NS
297	G	G	G	G	G	G	G	T	G	*ndhH*-	Genic	Pj-8	SSC	NS
298	G	G	G	G	A	A	A	A	A	*ndhH*-	Genic	Pj-5/STYPE	SSC	S
299	A	A	A	A	A	A	A	C	A	*rps15*-	Genic	Pj-8	SSC	NS
300	T	T	T	T	T	C	C	C	C	*rps15*-	Genic	LTYPE	SSC	S
301	G	G	G	G	G	G	G	T	G	*rps15-ycf1*	IGS	Pj-8	SSC	
302	T	T	T	T	T	T	T	C	T	*rps15-ycf1*	IGS	Pj-8	SSC	
303	T	T	T	T	T	G	G	G	G	*ycf1*a-	Genic	LTYPE	SSC	NS
304	G	G	G	G	G	G	G	T	G	*ycf1*a-	Genic	Pj-8	SSC	S
305	T	T	T	T	T	T	T	G	T	*ycf1*a-	Genic	Pj-8	SSC	NS
306	G	G	G	G	G	G	G	T	G	*ycf1*a-	Genic	Pj-8	SSC	NS
307	A	A	A	A	A	A	A	C	A	*ycf1*a-	Genic	Pj-8	SSC	S
308	A	A	A	A	A	A	A	C	A	*ycf1*a-	Genic	Pj-8	SSC	NS
309	T	T	T	T	T	T	T	A	T	*ycf1*a-	Genic	Pj-8	SSC	NS
310	A	A	A	A	A	C	C	C	C	*ycf1*a-	Genic	LTYPE	SSC	NS
311	G	G	G	G	G	G	G	G	A	*ycf1*a-	Genic	Pj-9	SSC	NS
312	C	C	C	C	C	C	C	C	T	*ycf1*a-	Genic	Pj-9	SSC	S
313	C	C	C	C	C	T	T	T	T	*ycf1*a-	Genic	LTYPE	SSC	S
314	C	C	C	C	C	C	C	C	A	*ycf1*a-	Genic	Pj-9	SSC	NS
315	G	G	G	G	A	A	A	A	A	*ycf1*a-	Genic	Pj-5/STYPE	SSC	S

국문 초록

색소체 (혹은 엽록체)는 빛을 필수 에너지원으로 변환하는 광합성 경로의 핵심 기관으로 식물의 생존과 관련된 중요한 기관 중 하나이다. 섹소체의 유전체 (Plastid genome; Plastome)은 원형구조를 가지며, 내부에는 두 개의 단일 복사 영역 (Single-copy region)과 두 개의 반전 반복 영역 (Inverted repeat region) 의 4개의 구역으로 구성된다. Plastome은 단일 세포에서 대량의 복제물의 생산이 가능하며, 적당한 변이 비율 그리고 한쪽 부모에서 내려오는 보수적인 유전적 특징 때문에 종간 다형성보다 종내 다형성이 낮게 발견되는 특징을 가진다. 이러한 특성을 활용하여, 식물 게놈 연구에서는 유전자 바코드 마커로서 엽록체 유전자 서열을 사용한다. 분자 바코딩 기술은 식물 식별실험을 쉽고 빠른 처리를 가능하게 하였지만, 그럼에도 불구하고 오류가 발생하는 경우가 잦다. 이러한 결함과 오류를 극복하기 위해, 최근에는 '슈퍼 바코딩'이라는 이름의 섹소체 유전체 전체를 이용하는 분석을 진행하는 방법이 바코딩의 연구의 새로운 트렌드가 되고 있다.

미나리아과 (Apioideae)는 약 380개 속 (Genus)와 3200여 종 (Species)를 보유하는 식물 과 (Family) 중 하나로, 경제적, 그리고 약초적으로 중요한 식물로 구성되어 있으며, 전 세계에 분포되어 있다. 미나리아과에 속하는 식물 종들의 계통발생학적 관계를 이해하는 것 연구는 오래전부터 진행되어온 중요한 연구 과제이다. 색소체 유전체의 유전자 염기서열과 핵 게놈 리보솜 DNA (Nuclear ribosomal DNA) 염기서열을 활용한 연구에서는 미나리아과에 속하는 식물종의 분기군과 종족을 분류하는 것에는 성공했지만, 여전히 몇몇 종들 사이에는 계통발생학적으로 해석하기 어려운 애매

한 부분이 발견되었다. 이러한 부분을 해결하기 위해, 최근 NGS 기술이 발전함에 따라 슈퍼바코딩 기술을 기반으로 구성된 계통발생학적 분석 연구의 수가 증가하였지만, 미나리아과 종에 관한 계통발생학적 연구는 대부분 종속간 (intra-genus) 또는 종간 (inter-species)비교에 초점을 맞추고 있다. 본 논문은, 색소체 유전체의 구조를 바탕으로 계통발생학적 관계도와 종간 및 종내 패턴의 변화를 관찰하기 위해 국내 자생하는 미나리아과 식물 종들을 대상으로 연구를 진행하였다.

본 연구는 두 개의 장으로 구성되었다. 첫 번째 장에서는, 계통발생학적 분석을 통해 미나리아과 식물 종들 간의 상호특이적 관계를 조사했다. 한국에서 자생하는 미나리아과 식물종을 31개체를 대상으로, 색소체 유전체의 완전장을 dnaLCW 방식으로 생성하였다. 그 외 추가로 NCBI 데이터베이스에서 29개의 미나리아과 식물과 외집단 (Outgroup)으로 선정된 3개의 두릅나무과 (Araliaceae)종의 색소체 유전체 완전장 염기서열 정보를 포함하여, 단백질 코딩 시퀀스 (CDS)를 대상으로, 계통수를 Maximum Likelihood 방법을 1000개의 Bootstrap replication 조건으로 제작하였다. 본 연구에서 그려진 계통수로 분석된 미나리아과 식물의 계통발생학적 관계는 다른 연구에서도 제시한 종족 및 계통군을 따랐다. 선택된 미나리아과 식물 종의 위상 관계를 이해하기 위해, 색소체 유전체의 구조, 특히 LSC (Large Single Copy)와 IR (Inverted Repeat) 영역 사이의 접합 (Junction)에 초점을 맞추었다. 비교 분석을 통해, 우리는 미나리아과 종들이 크게 세 그룹으로 시각화 되는 것을 확인하였다. 외집단으로 선정된 두릅나무과를 포함하여 그룹 1에 속하는 종들은 *rps19* 유전자 근처에 LSC/IR 접합부를 가지고 있다. 그룹 2에 속하는 종은 *rpl16* ~ *rpl2* 유전자 사이에서 위치한

LSC/IR 접합부를 가지고 있다. 그룹 3에 속하는 종들은 *ycf2* 유전자 근처에 LSC/IR 접합부를 가지고 있다는 것을 확인하였다. 그룹 2 그리고 그룹 3에 속한 종들은 LSC/IR 접합부가 이동함에 따라 색소체 유전체의 IR영역의 크기에서 축소 및 증폭의 변화가 일어났다는 것을 증명하였으며, 그것에 대한 결과를 바탕으로 네가지 변화과정을 설명하였다. 또한 각 미나리아과 그룹 사이의 분리 지점에 대해 분화 시간 추정은 미나리아과 식물 계열에 따른 진화 역사에 대한 추가 정보를 제공한다. 본 연구에서 얻은 정보는 미나리아과 식물들의 색소체 유전체의 차이에 대한 추가적인 이해를 제공한다.

두 번째 장에서는, 색소체 유전체의 완전장을 비교 분석을 통해 갯기름나물 (*Peucedanum japonicum*)의 종내 다양성을 조사하였다. 갯기름나물은 한국에서 자생하는 식물로 한약 및 쌈 채소 외 다른 분야에도 활용가능한 식물이다. 갯기름나물에서는 같은 종내에서 색소체유전체의 게놈 크기에서 차이를 보이는 두개 타입의 색소체 유전체 (L타입, S타입)이 발견되었다. 그외 비교분석을 통해 발견된 다형성으로부터 17 InDel (Insertion Deletion) 마커와 8개의 Single Nucleotide Polymorphism (SNP)를 대상으로 한 KASP (Kompetitive Allele Specific PCR) 마커가 개발되었고, 갯기름나물의 38개의 채집된 집단에 대한 각 분자 마커를 테스트하여 갯기름나물의 광범위한 유전적 다양성을 제공하였다. 비교 분석 및 마커 평가 연구를 통해 본 연구는 갯기름나물 종내에서 광범위한 종내 다양성에 대한 정보를 제공한다.

요약하자면, 본 연구는 색소체의 다양성이 미나리아과 식물 종에 대한 종간 및 종내 분석을 통해 미나리아과 식물들의 애매모호한 계통발생학적 패턴의 이해와 종간 그리고 종내 다양성에 대한 정보 제공하고 또한

갯기름나물의 육종에 유용한 정보와 기술을 제공할 것이다.

주요 단어: 미나리아과, 종내 종간 유전 다양성, 갯기름나물, 색소체 유전체, DNA 마커

학번: 2014-22947