scientific reports

Check for updates

OPEN

A recent large-scale intraspecific IR expansion and evolutionary dynamics of the plastome of *Peucedanum japonicum*

Ho Jun Joh[1,2,9], Young Sang Park[1,9], Jong-Soo Kang[1], Jin Tae Kim[3], Jickerson P. Lado[1,6], Sang Il Han[4], Young-Won Chin[5], Hyun-Seung Park[1,7], Jee Young Park[1] & Tae-Jin Yang[1,3,8]✉

Peucedanum japonicum **(PJ), a member of the Apiaceae family, is widely distributed and cultivated in East Asian countries for edible and functional foods. In this study, we compared the plastid genomes (plastomes) and 45S nuclear ribosomal DNA (45S nrDNA) simultaneously from 10 PJ collections. Plastome-based phylogenetic analysis showed that the PJ accessions were monophyletic within the genus *Peucedanum*. However, ten plastomes were classified into two different groups according to their length of inverted repeat (IR) block, the short-type (S-type) plastome group containing the 18.6 kbp of the original IR and the long-type (L-type) plastome group containing the 35.7 kbp of expanded IR by duplication of the 17.1 kbp of the large single copy region. A total of nine single nucleotide polymorphisms and eight insertions or deletions were identified among the five L-type plastomes, whereas large variations were identified among the five S-type plastomes. Calculation of synonymous substitution rates and divergence time estimation suggested that the 17 kbp IR expansion occurred recently. Molecular markers were developed and validated to classify the 55 PJ germplasm according to their plastome types. Our study would be useful for unraveling the dynamic evolution of plastomes in the Apiaceae family and for the molecular breeding of PJ.**

Keywords *Peucedanum japonicum*, Plastome, Phylogenetic analysis, Intra-species polymorphism, Inverted repeat, Kompetitive allele-specific PCR (KASP)

Peucedanum japonicum Thunb. (PJ) is a member of the family Apiaceae. This species is distributed in the continental coastal regions of East Asia, including mainland China and Korea[1]. Like other species of *Peucedanum*, PJ is rich in coumarin components, and its roots have been used as traditional oriental medicine to cure coughs, colds, and other symptoms[2]. As the demand for healthy products increase, the aerial parts of PJ, such as leaves and flowers, are also being utilized as herbal tea, diet products, and vegetables[3].

The discordance between phylogenetic relationships and infrageneric classification based on morphological characters has been reported in the genus *Peucedanum* as along with its subfamily Apioideae[4–6]. To resolve the discordance, previous studies utilizing the plastid genome (plastome), nuclear ribosomal DNA (nrDNA), and transcriptome data suggested that the discordance within the Apioideae was the consequence of chloroplast capture as a result of hybridization or introgression[4–8].

[1]Department of Agriculture, Forestry and Bioresources, Plant Genomics and Breeding Institute, Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea. [2]Department of Forest Bioresources, National Institute Forest Science, Suwon 16631, Republic of Korea. [3]Interdisciplinary Program in Agricultural Genomics, College of Agriculture and Life Sciences, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea. [4]Medicinal Plant Garden, College of Pharmacy, Seoul National University, Koyang 10257, Republic of Korea. [5]College of Pharmacy and Research Institute of Pharmaceutical Sciences, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea. [6]Genetics and Molecular Biology Division, Institute of Biological Sciences, College of Arts and Sciences, University of the Philippines Los Baños, 4031 Laguna, Philippines. [7]Department of Integrative Biological Sciences and Industry, Convergence Research Center for Natural Products, Sejong University, Seoul 05006, Republic of Korea. [8]Institute of Seed Biotechnology, Institutes of Green Bioscience and Technology, Seoul National University, Gangwon-do 25354, Republic of Korea. [9]Ho Jun Joh and Young Sang Park contributed equally to this work. ✉email: tjyang@snu.ac.kr

The plastome typically exhibits a quadripartite structure, comprising four parts: two single-copy (SC) regions, the large single copy (LSC) and the small single copy (SSC), separated by two copies of an inverted repeat (IR) region. In land plants, the length of the plastome generally ranged from 120 kbp to 160 kbp, with conserved gene content and order[9]. Plastome inheritance typically occurs uniparentally, primarily through the maternal line, which suggests limited genetic recombination in the genomic sources[10–12]. These genetic characteristics of the plastome contribute to its conserved features[10]. In addition to these conserved features, the relatively higher mutation rates of plastid genes compared to those of mitochondrial genes allow for the accumulation of moderate variants, making it the most widely used genome data for phylogenetic studies and DNA barcoding of land plants[13–16]. Although the plastome exhibits conserved characteristics, the degree of intraspecies variation differs considerably, ranging from mere single-digit to hundreds of variations among species[17–19]. The 45S nrDNA exists in a tandem repeat array of thousand copies with conserved features due to concerted evolution[20,21]. Despite the conserved features of the 45S nrDNA, internal transcribed spacers (ITS), including ITS1, ITS2, and 5.8S rDNA, exhibited relatively variable sequences[22]. In contrast with the plastome, as 45S nrDNA is in the nuclear genome, it is inherited biparentally. Therefore, ITS regions are also utilized for phylogenetic studies along with plastomes. Several assembly pipelines and tools have been developed to assemble plastome and 45S nrDNA simultaneously[17,23,24]. Utilizing complete plastomes and 45S nrDNA could yield much more informative sites than conventional barcoding methods[15].

In land plants, two IR regions are located between two SC regions, with lengths typically ranging from 15 kbp to 30 kbp. It has been reported that nucleotide substitution rates of genes located in the IR regions are relatively lower than those in the SC regions. Furthermore, boundary shifts of the IR region of closely related species generally occur on a small scale, resulting in a few gene number changes[25–27]. On the other hand, previous studies have reported large-scale boundary shifts of the IR region in various plants[28–32]. Previous studies elucidated the evolutionary relationships among Apioideae species with their IR boundaries[7,33]. Moreover, the IR expansion and contraction at the genus level were reported from two closely related genera, *Angelica* and *Peucedanum*, respectively[6,31]. In the case of *Angelica*, it was observed that the IR regions of *A. morii* and *A. tsinlingensis* exhibited expansions of 7 and 17 kbp, respectively. IR expansion of about 17 kbp was previously reported by our group[30]. However, no such expansion was found in another PJ plastome reported by Liu et al.[6,30].

Here, we assembled seven new complete PJ plastomes and 45S nrDNA. We conducted a comparative analysis of plastome characteristics associated with the recent large-scale boundary shift of the IR regions. We identified that the boundary shift occurred with some accessions having longer IR regions than others. To understand its evolutionary dynamics, we divided it into two parts: expanded IR (e-IR) and original IR (o-IR). Moreover, five Kompetitive allele-specific PCR (KASP) markers were developed from the polymorphisms identified in plastomes to distinguish our 55 germplasms. This study will expand our understanding of the large-scale structural variation observed within the same species and will be useful for molecular-assisted breeding and classification of PJ collections.

Results

Characteristics of plastomes and 45S nrDNAs among PJ collections

In this study, we assembled complete plastomes and 45S nrDNA sequences from seven PJ accessions. The total length of plastomes ranged from 147,525 bp to 164,660 bp. The ten PJ plastomes, including three from previous studies, were divided into two types based on their length: long-type (L-type) and short-type (S-type) (Table 1)[6,30]. Both the L-type and S-type plastomes presented a typical quadripartite structure, comprising two SC regions and two IR regions despite their differences in length. Five of the ten PJ accessions (Pj-1, Pj-2, Pj-3, Pj-4 and Pj-5) exhibited the L-type plastome, while the other five (Pj-6, Pj-7, Pj-8, Pj-9 and Pj-10) exhibited the S-type plastome. The L-type plastomes ranged from 164,652 bp to 164,660 bp in total length, while the S-type plastomes ranged from 147,525 bp to 147,592 bp. The LSC region and IR regions presented large differences in length based on the plastome type, with a range of 75,583 to 92,804 bp in the LSC region and 18,606 to 35,760 bp in the IR region. However, the SSC region was relatively conserved, with a range of 17,551 to 17,628 bp. Therefore, the difference between the L-type and S-type plastomes of PJ was observed to be a result of the boundary shift at the junction between the LSC and IRb region (JLB). The boundary shift was not identified in the SSC and IRa

Accession	Length (bp)	LSC (bp)	SSC (bp)	IR (bp)	Gene	CDS	tRNA	rRNA	GC content (%)	GenBank accession
Pj-1	164,653	75,584	17,551	35,759	141	96	37	8	37.46	KU866530
Pj-2	164,653	75,584	17,551	35,759	141	96	37	8	37.46	KU866531
Pj-3	164,652	75,583	17,551	35,759	141	96	37	8	37.45	PQ315775*
Pj-4	164,656	75,585	17,551	35,760	141	96	37	8	37.45	PQ315776*
Pj-5	164,660	75,588	17,552	35,760	141	96	37	8	37.45	PQ315777*
Pj-6	147,525	92,563	17,628	18,667	126	82	36	8	37.54	PQ315778*
Pj-7	147,548	92,631	17,605	18,656	126	82	36	8	37.54	PQ315779*
Pj-8	147,578	92,655	17,605	18,659	126	82	36	8	37.54	PQ315780*
Pj-9	147,592	92,804	17,576	18,606	126	82	36	8	37.55	PQ315781*
Pj-10	147,592	92,804	17,576	18,606	126	82	36	8	37.55	OK336477

Table 1. Information of PJ plastomes used in this study. *Newly assembled in this study.

region (JLA). Fourteen genes (*petD*, *rpoA*, *rps*11, *rpl*36, *infA*, *rps*8, *rpl*14, *rpl*16, *rps*3, *rpl*22, *rps*19, *rpl*2, *rpl*23, and *ycf*2) originally located in the LSC region of the S-type plastome shifted into the e-IR region, which led to the duplication of the 14 genes in the e-IR regions of the L-type.

The 45S nrDNA sequences of seven PJ accessions were assembled into a single contig consisting of 18S rDNA, ITS1, 5.8S rDNA, ITS2, and 26S rDNA. Although the plastome length was significantly different among PJ accessions, the 45S nrDNAs of the nine PJ accessions, including two from a previous study were consistent in length and structure (Supplementary Table S1)[30].

Intraspecific diversity in the plastomes and 45S nrDNAs of PJ collections

A comparison of the ten PJ plastomes, a total of 473 intraspecific polymorphisms, including 301 single nucleotide polymorphisms (SNPs) and 172 insertions/deletions (InDels), were identified (Fig. 1; Table 2). Among the 301 SNPs, 242 were parsimony informative sites, while 59 were singleton sites. A pairwise comparison revealed that

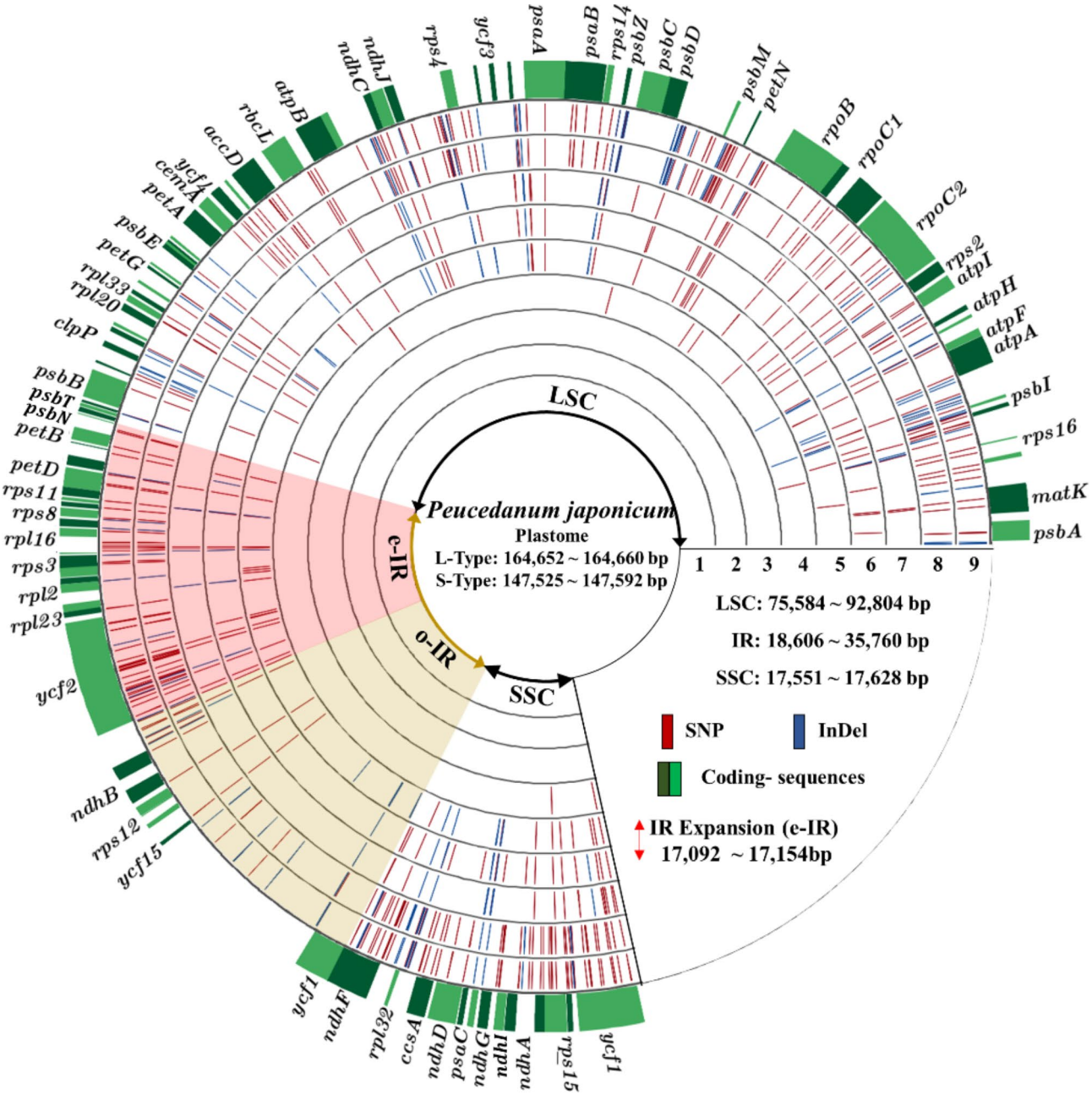

Fig. 1. Plastome map of PJ. Shadow region indicates the IR region. The yellow shadow region indicates the o-IR region. The red shadow region and red arrow indicate the e-IR region in L-type accessions but the LSC in S-type. Red line and blue line indicate SNP and InDel respectively. Variants were calculated by pairwise comparison using Pj-1 as a reference. 1: Pj-1 & Pj-2, 2: Pj-1 & Pj-3, 3: Pj-1 & Pj-4, 4: Pj-1 & Pj-5, 5: Pj-1 & Pj-6, 6: Pj-1 & Pj-7, 7: Pj-1 & Pj-8, 8: Pj-1 & Pj-9, 9: Pj-1 & Pj-10. Single indel (1 bp indel) were excluded.

Accession	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)	Pj-10 (S)
Pj-1 (L)	–	**0**	**0**	**0**	**9**	61	70	63	222	222
Pj-2 (L)	**0**	–	**0**	**0**	**9**	61	70	63	222	222
Pj-3 (L)	**1**	**1**	–	**0**	**9**	61	70	63	222	222
Pj-4 (L)	**2**	**2**	**2**	–	**9**	61	70	63	222	222
Pj-5 (L)	**6**	**6**	**6**	**5**	–	64	73	66	225	225
Pj-6 (S)	41	41	41	40	38	–	**47**	**40**	**219**	**219**
Pj-7 (S)	41	41	42	41	41	**38**	–	**34**	**227**	**227**
Pj-8 (S)	44	44	43	43	43	**38**	**19**	–	**222**	**222**
Pj-9 (S)	107	107	107	108	105	**113**	**115**	**113**	–	**0**
Pj-10 (S)	107	107	107	108	105	**113**	**115**	**113**	**0**	–
	InDels									
Total SNPs: 301 Total InDels: 173										

Table 2. Sequence variations among PJ plastomes. Significant values are in bold.

L-type plastomes had fewer variants than S-type plastomes. A comparison of the five L-type plastomes revealed nine SNPs and eight InDels. Among them, four L-type plastomes (Pj-1, Pj-2, Pj-3, and Pj-4) exhibited almost identical sequences with one or two InDel variants. Furthermore, Pj-5, which had nine SNPs compared to the other four L-type plastomes, had fewer variants when compared to the S-type plastomes. A comparison of the five S-type plastomes exhibited 262 SNPs and 152 InDels. In particular, two S-type plastomes (Pj-9 and Pj-10) exhibited identical sequences and showed numerous variants compared to other accessions, suggesting that they could be genetically distinct from both L-type and S-type plastomes.

The nucleotide diversity (pi value) was calculated in order to estimate divergence hotspots across the PJ plastomes (Fig. 2A). The pi value of the plastome was calculated as 0.00080, while the pi values of individual windows ranged from 0 to 0.00493. Eleven regions (*rpoB-petN*, *petN-psbM*, *rps4-trnL*, *rpl16*, two regions of *ycf2*, *ycf2-ndhB*, *ndhF-rpl32*, *ndhH* and two regions of *ycf1a*) were identified as divergence hotspots based on a pi value threshold of 0.0025, exhibiting higher pi values than other regions. Seven regions were identified as being located in the SC regions, while four regions (*rpl16*, two regions of *ycf2*, and *ycf2-ndhB*) were located in the e-IR region.

In order to investigate mutations in the protein-coding genes, a total of 76 commonly shared protein-coding genes in the ten PJ plastomes were analyzed. Forty-six of the shared genes contained variants, while the remaining 31 genes were identical among the PJ plastomes (Fig. 2B). Among the 45 genes, four genes (*rpoC2*, *rpoA*, *ycf2*, and *ndhF*) contained both SNPs and InDels, and the remaining 41 genes exhibited only SNPs. The InDels observed in three genes (*rpoC2*, *ycf2*, and *ndhF*) exhibited individual-specific InDels among the PJ plastomes including different copy numbers of poly-A or T and tandem repeats. The *rpoA* gene was contracted by a single SNP, creating an early termination of the gene (Supplementary Figure S1). The single SNP induced a non-synonymous substitution that changed TCA to TAA, resulting in the conversion of amino acid from serine to stop codon in the plastomes of Pj-1, Pj-2, Pj-3, Pj-4, Pj-5, Pj-9 and Pj-10.

A comparison of the 45S nrDNA sequences among nine PJ accessions revealed that four SNPs were included in 21 heterozygous nucleotide sites. This indicates that all four SNPs in the 45S nrDNA sequences were heterozygous sites in at least one accession (Supplementary Table S2).

Large structure variations in the plastomes of the PJ collections

The inversion of the *trnD-trnY-trnE* region has previously been reported in two *Peucedanum* and two *Angelica* species, namely *P. japonicum*, *P. medicum*, *A. gigas*, and *A. morii*[6,31]. The inversion was observed in all ten PJ accessions, including Pj-9 and Pj-10 (Fig. 3 and Supplementary Figure S2). The inversion length was about 490 bp in the four species, and around 40 bp of the flanking regions of the homologous regions were found in the Apioideae subfamily. The forward flanking region was adjacent to the *trnD* gene, while the backward flanking region was almost 70 bp away from the *trnE* gene. The inversion seems to have occurred in the same regions in the four species and appears to not be related to the phylogenetic relationships within the family (Fig. 4A).

The IR region of the L-type plastome underwent expansion, which led to a shift in the JLB and structural change (Fig. 3). A comparison of the IR lengths revealed that the IR region of the L-type was almost 17 kb longer than that of the S-type. The IR regions of the L-type plastomes ranged from 35,759 to 35,760 bp in length, whereas those of the S-type plastomes ranged from 18,606 to 18,667 bp. In the aspect of the genic regions, three genes (*ndhB*, *rps7*, and *ycf15*) existed as two copies in the o-IR region of the S-type plastomes. However, fourteen genes (*petD*, *rpoA*, *rps11*, *rpl36*, *infA*, *rps8*, *rpl14*, *rpl16*, *rps3*, *rpl22*, *rps19*, *rpl2*, *rpl23*, and *ycf2*) were duplicated in the e-IR region of the L-type plastome as a consequence of the IR expansion. Among the 17 duplicated genes, only *rpoA* and *ycf2* showed variations in length, whereas the remaining 15 genes were still consistent.

To investigate whether the IR expansion affected the reduction in mutation rates of the duplicated plastid genes, synonymous substitution rates for each plastome compartment were estimated (Supplementary Table S3). A pairwise comparison using *Panax ginseng* as an outgroup revealed that neither compartment exhibited

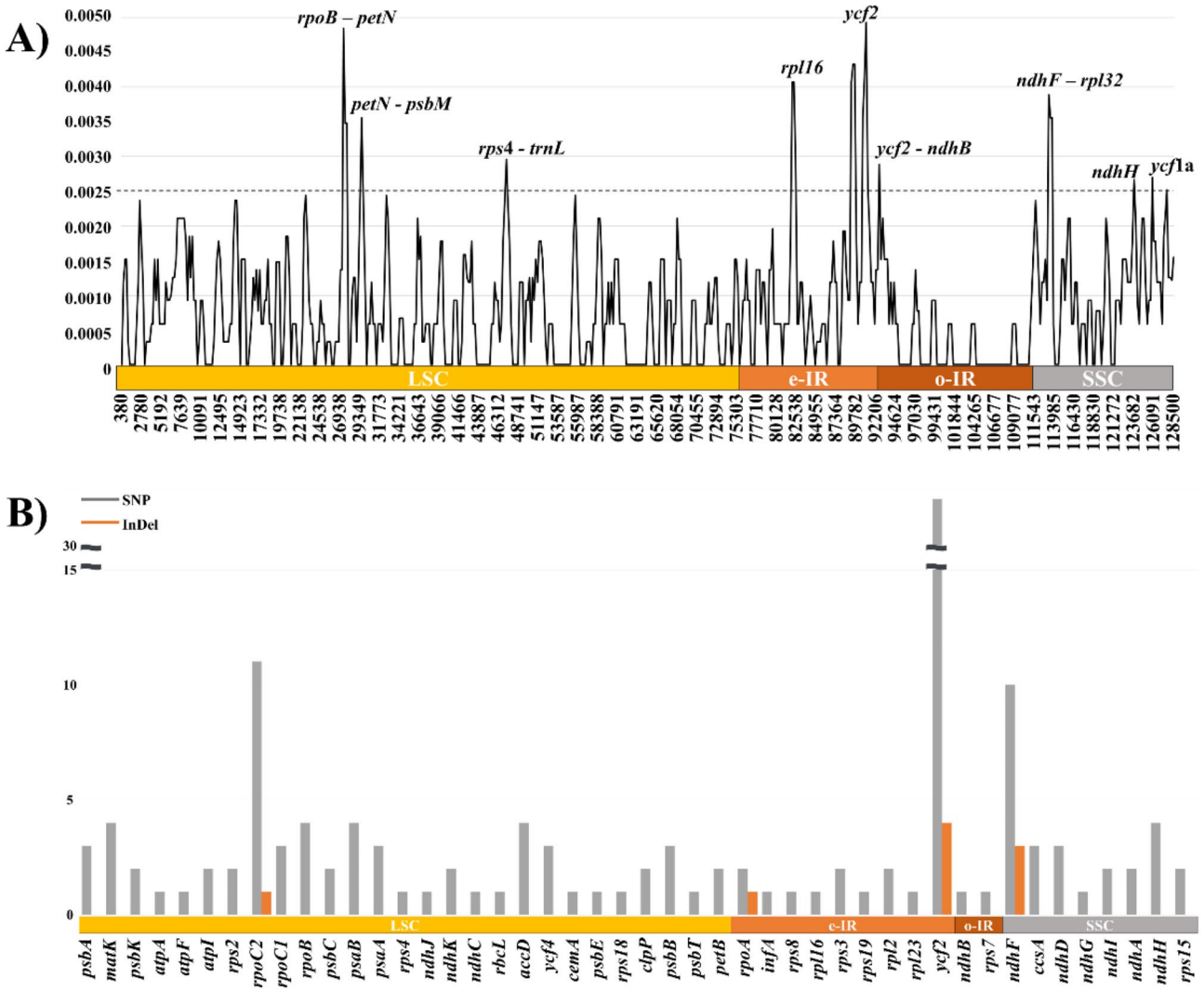

Fig. 2. Nucleotide diversity (pi value) and number of variants among protein-coding genes in PJ plastomes. (**A**) Total pi value calculated as 0.00080 by DnaSP with sliding window method (Window size: 600 bp, Sliding size: 200 bp). Eleven regions exhibiting higher pi value than other regions were estimated as divergence hotspots. (**B**) Gray and orange vertical bar indicate SNP and InDel respectively. Yellow, orange, dark orange, and gray horizontal bar under the graph indicate LSC, e-IR, o-IR and SSC regions respectively. Among 76 commonly shared genes in 10 PJ accessions, the identical 31 genes showed no variants among 10 PJ accessions were not marked.

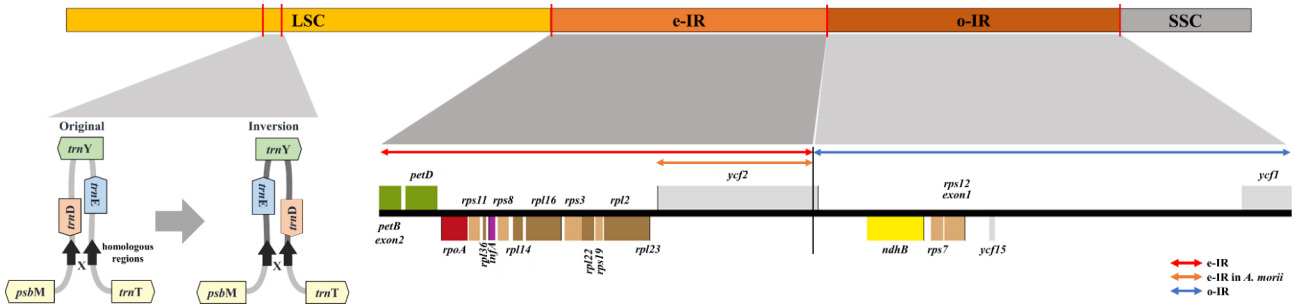

Fig. 3. Structure variations of the PJ plastomes. Illustration of IR expansion and tRNA inversion. The blue line indicates the o-IR. The red line indicates e-IR in L-type plastomes of PJ and *Angelica tsinlingensis*. The orange line indicates e-IR in the plastomes of *A. morii*.

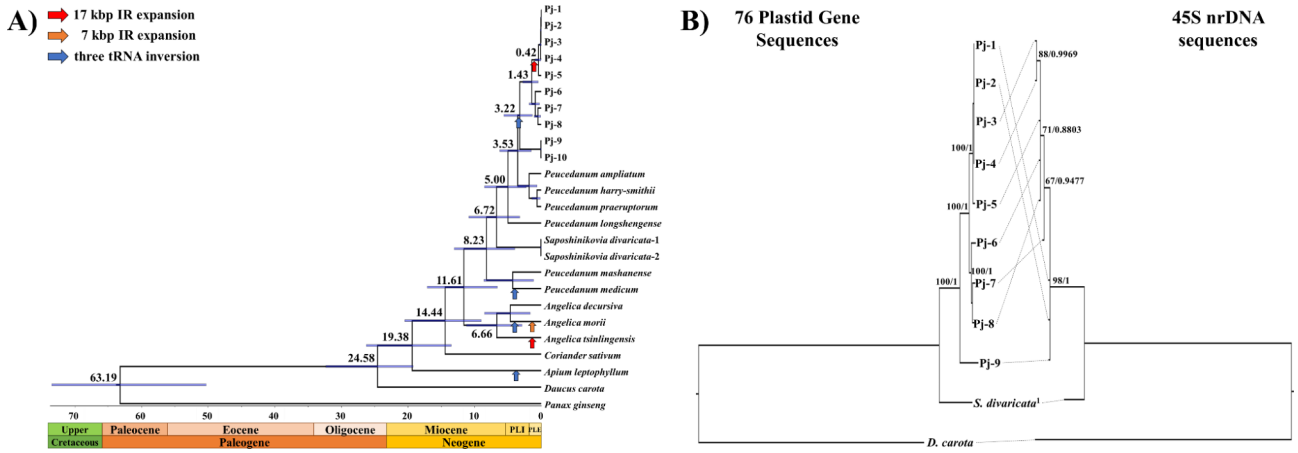

Fig. 4. Divergence time estimation and phylogenetic relationships of PJ with related species. (**A**) Estimated divergence times were indicated near the nodes. Unit for marked time is Mya. Structure variations occurred in PJ were marked on the branches. Red, orange, and blue arrows indicated 17 kbp IR expansion, 7 kbp IR expansion, and three tRNA inversion, respectively. Two *S. divaricata* were collected from different accessions in Genbank; *S.divaricata*-1: KU866529, *S. divaricata*-2: MZ089852. (**B**) Phylogenetic relationships between plastome and 45S nrDNA sequences. Genbank ID of 45S nrDNA sequences of Pj-1, Pj-2, *S. divaricata*, and *D. carota* were KX757776, KX757777, KX757775, and MF185182 respectively. Bootstrap support value and posterior probabilities (BP/PP) are shown above the branches.

significantly accelerated or decelerated mutation rates in the PJ plastomes. Overall, it can be inferred that the IR expansion caused by the boundary shift in JLB did not affect the mutation rates.

Phylogenetic analysis and divergence time estimation

Phylogenetic analyses were conducted using newly assembled and publicly available plastomes and 45S nrDNA of PJ and its closely related species in the genus *Peucedanum* (Fig. 4 and Supplementary Figure S3). The plastome-based tree (PT-tree), constructed using Maximum likelihood (ML) and Bayesian inference (BI) methods, indicated an identical topology, with high bootstrap support values (BS) and posterior probabilities (PP). The topology of the PT-tree in this study was consistent with those of previous studies[6,7]. In the PT-tree, all ten PJ accessions were monophyletic (BS = 100, PP = 1), and the PJ clade was sister to the clade of three *Peucedanum* species, *P. ampliatum*, *P. harry-smithii* var. *grande*, and *P. praeruptorum* (BS = 100, PP = 1). Within the PJ clade, the five L-type accessions (Pj-1, Pj-2, Pj-3, Pj-4, and Pj-5) and three S-type accessions (Pj-6, Pj-7, and Pj-8) formed their own subclades, respectively, and the remaining S-type accessions, Pj-9 and Pj-10, were sister to the two subclades (BS = 82, PP = 0.986). These results suggest that Pj-9 and Pj-10 were distinct from the other eight PJ accessions, and the L-type plastome probably diverged from the S-type plastome in PJ (Fig. 4A).

The 45S nrDNA-based tree (NR-tree) was constructed including nine accessions of PJ (Pj-1 to Pj-9) and two species (*Saposhnikovia divaricata* and *Daucus carota*) (Fig. 4B and Supplementary Figure S3B). The phylogenetic relationships among the nine PJ accessions were inconsistent between the PT-tree and the NR-tree. The Pj-7 was sister to Pj-8 in the PT-tree, while it was sister to the other five accessions in the NR-tree. In addition, two L-type accessions, Pj-1 and Pj-2, showed a polytomy with one S-type accession (Pj-9), in the NR-tree (Fig. 4B).

We estimated the divergence times of PJ accessions with their related species in order to reveal the evolutionary history of their plastomes (Fig. 4A). As a result, PJ diverged from other *Peucedanum* species around 3.53 million years ago (Mya). Within PJ collections, the plastomes of Pj-9 and Pj-10 diverged from the other eight PJ accessions at 3.22 Mya, and the split between common ancestor of L-type and S-type plastomes occurred at 1.43 Mya.

Development of KASP markers for classification of PJ collections

We developed five KASP markers by using SNPs in the PJ plastomes to distinguish our 55 PJ germplasms, seven of them were used for plastome assembly. The regions of *psbK*, *rpoC2*, *psaA*, *ycf4*, and *rpoA* were utilized for developing the markers (Supplementary Table S4). The result of the five KASP markers divided our germplasm into six different types (A-F) (Fig. 5 and Supplementary Figure S4). The most common type observed in our collection was the A type ($n = 38$), which included Pj-1, Pj-2, and Pj-3. The B type ($n = 2$), C type ($n = 2$), and D type ($n = 1$) were present in one or two accessions, while E and F types were in four and eight accessions, respectively. Therefore, these marker sets would be useful for tracing maternal origin, identifying novel plastome types, and conserving rare accessions from the germplasm for the breeding program.

Discussion

Intra-species diversity of PJ plastomes

It is widely recognized that the structure, gene contents, and gene order of the plastome are highly conserved[9]. The conserved features and moderate variations between species make the plastome a valuable genetic resource

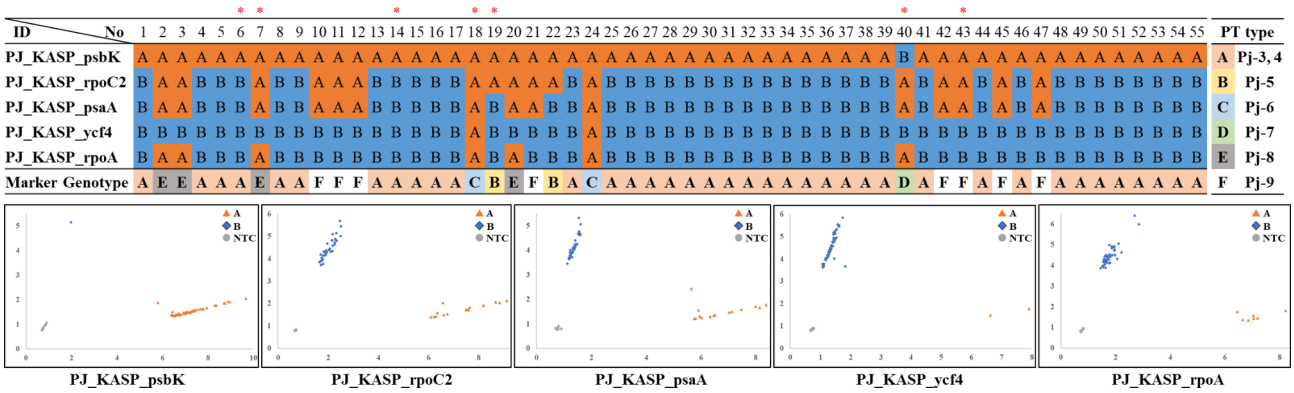

Fig. 5. Whole KASP marker validation for 55 PJ accessions and its genotyping results. A: X allele, B: Y allele. In the KASP plot, orange triangles, blue rectangles, and grey circles indicate A, B, and NTC, respectively. Detail maker validation results of each PCR marked on Supplementary Figure S4. Asterisks are the accession used for plastome assembly in this study; 6: Pj-3, 7: Pj-8, 14: Pj-4, 18: Pj-6, 19: Pj-5, 40: Pj-7, 43: Pj-9.

for species discrimination[15]. The degree of intra-species variation may be minimal or substantial, depending on the species[17–19]. A comparison of ten PJ plastomes revealed the genetic diversity within the PJ. The intra-species variations among ten PJ plastomes ranged from a small number to hundreds of variants. When comparing five L-type plastomes, there was a single-digit variation. Meanwhile, comparison of the L-type and S-type revealed more than a hundred variants. However, the significant difference among the ten PJ plastomes was the length of the plastome, which was caused by the boundary shift of JLB. The length difference between the L-type and S-type IR regions was about 17 kbp. Such a dynamic boundary shift in the IR region has been previously reported in other Apiaceae species[7,33]. Similar large-scale IR expansions have been reported at the genus level from *Ligusticum* and *Angelica* with expansion of 15 kbp and 17 kbp, respectively[31,34].

In addition to the boundary shift, an inversion involving three tRNA genes was identified when compared with its closely related species. Furthermore, all ten PJ plastomes shared the same inversion event. The sequence similarity of the flanking regions of the inversion implied that the inversion of three tRNA genes is likely caused by a hairpin loop structure[35,36]. The same inversion events involving the three tRNA genes had been reported in other species, including *P. medicum*, *A. morii*, and *A. leptophyllum*, but not in all species in the genera *Peucedanum* and *Angelica* shows this specific inversion, indicating that the inversion event seems to occur in these species independently[6,31].

The large-scale structural variation, IR expansion and inversion events of the plastomes, occurring independently in various species of the Apiaceae family indicate a dynamic plastome evolution in this family[7,26,30,31,33].

The recent boundary shifts for the IR expansion

Previous studies have frequently reported the dynamic shift of the JLB in PJ and its related species[7,26,30,31,33]. The expansion and contraction of the IR regions are known to be the primary factors of the plastome variations in length[26,37,38]. Previous studies reported boundary shift in IR regions and proposed a model for it[25,27,39,40]. Unlike previous studies that reported repeat-related expansion, simple sequence repeat (SSR) and tandem repeat (TR) were not detected in flanking regions of the IR (Supplementary Figure S5). Moreover, the coverage plot revealed that the plastome of PJ might not exist as heteroplasmy (Supplementary Figure S6). It was reported that the shifts of the JLB frequently occur in the Apiaceae family with diverse boundaries in specific taxa or species[7,30,31,33]. Therefore, by considering the diverse IR length variations among the Apiaceae family, it would be challenging to specify a single factor responsible for IR expansion. It is also possible that the boundary shifts of the IR region could affect the mutation rates of the genes that have been shifted from the SC to the IR regions and vice versa. Previous studies indicated that the mutation rates of the genes that shifted into the IR regions decreased[13,27,41]. However, it has also been reported that mutation rates of genes in the newly expanded IR regions remained unchanged between closely related species[37,42]. The divergence time estimation indicated that the split between the L-type ancestors and S-type ancestors of PJ occurred at 1.43 Mya, and the split within the L-type plastomes occurred at 0.42 Mya. This finding suggests that the IR expansion occurred between 0.42 and 1.43 Mya by the boundary shift in JLB. Therefore, it can be expected that the genes shifted into the e-IR region may not yet have sufficient time to show significant mutation rate differences with the homologous genes in the LSC region of other species.

Geological distribution and genetic diversity of PJ

Not only in South Korea, PJ resources are localized in lower latitude regions such as the Philippines Arc[1]. In this study, we revealed the diversity of PJ plastomes and found that there are two big different plastome types, L-type and S-type, among PJ collections from South Korea (Supplementary Figure S7 and Supplementary Table S5). As the PJ is distributed from Honshu Arc (excluding Hokkaido) to the Philippines Arc, it will be needed to collect diverse PJ resources localized in lower latitude regions such as Ryukyu Arc, Taiwan Arc, and Philippines Arc.

Peucedanum japonicum var. *latifolium* has been observed in the Ryukyu Islands, also known as the Nansei or the Ryukyu Arc, a chain of more than 150 Japanese islands that stretch southwest from Kyushu to Taiwan. According to a previous study, PJ resources from different islands in the Ryukyu Islands might belong to distinct infraspecific taxa[43]. Further, Seo and Hotta divided PJ from Japan into three varieties based on geographical variation in four morphological characteristics (stem diameter, stem length, flowering shoot length, and height) and two life history traits (flowering periods and reproduction mode)[44]. In Taiwan Arc, close to Yonaguni of the Ryukyu Islands, PJ is abundant in the coastal area of its northern part and the Lanyu-Lutao Islands, also known as the Orchid Islands[45]. Previous studies reported that khellactone coumarins isolated from the PJ in Taiwan showed biochemical structural differences between species from the northern region and the Lanyu Islands[46–49]. The Batanes Islands in the Philippines Arc are a unique biogeographic region. Several works noted how the Batanes Islands share similar indigenous plants with the Lanyu-Lutao Islands, the Ryukyu Islands, and the Babuyanes Islands north of Luzon[50,51]. According to palynological studies, warm-temperate plants, including the northern PJ, may have migrated southward during glacial ages[52–55]. The Batanes Islands maintained close geographical contact with the neighboring Lanyu-Lutao with similar annual temperatures and no distinct four seasons.

This study suggested that the L-type originated recently from an S-type collection. An S-type plastome sequence was recently reported from China[6]. With a notable distribution of PJ in these island groups of the Ryukyu Arc, Taiwan Arc, and Philippines Arc, these areas can be used to expand the genomic diversity, including its migration and gene flow. Additionally, it would be useful to understand the effects of geographical distribution and climate on flowering periods, morphological diversity, and the biochemical and molecular diversity of PJ. Therefore, further study is needed to elucidate the structure variations of the plastomes, evolutionary relationships, genomic diversity, and the effects of geographical distribution and climate on the PJ resources.

Phylogenetic analysis of plastomes and nrDNAs

The plastomes are mainly inherited from the maternal parent, while the nrDNAs are inherited from both parents[10]. Both plastomes and nrDNAs were simultaneously assembled using the whole genome sequence of nine PJ individuals and utilized publicly available data from two other species (*Saposhnikovia divaricata* and *Daucus carota*) to construct a phylogenetic tree (Fig. 4B and Supplementary Figure S3).

Within the PJ clade, the PT-tree was not consistent with the NR-tree. Phylogenetic discordance between plastome and nrDNA could be observed in major crops, such as two rice cultivars that were developed by inter-subspecies hybridization between indica and japonica rice[35]. The plastomes of five L-type accessions (Pj-1, Pj-2, Pj-3, Pj-4, and Pj-5) and three S-type accessions (Pj-6, Pj-7, and Pj-8) formed their own subclades. Meanwhile, the plastome of Pj-9 and Pj-10 is distinct from the other eight PJ accessions. The divergence time estimation of this study estimated the first divergence among PJ at 3.22 Mya and the second one at 1.43 Mya. Even though the genetic distance of Pj-9 and Pj-10 is relatively far from other PJ accessions, they formed their own subclade (PJ clade). This PJ clade was sister to the clade of the other three *Peucedanum*, consisting of *P. ampliatum*, *P. harry-smithii*, and *P. praeruptorum*, consistent with previous studies[6,7]. Additionally, the three tRNA inversion event was detected in all the accessions in the PJ clade, whereas its sister clade was not. Moreover, the NR-tree revealed that the nrDNA of Pj-9 was identical to that of Pj-1 and Pj-2. Four SNPs detected in the 45S nrDNA of 10 PJ accessions. One SNP was detected in the 5.8S region and the ITS2 region, respectively, and the other two were detected in the 26S region (Supplementary Table S2). All the PJ formed their own subclade in the NR-tree, the ITS tree, and the ITS1 tree, while Pj1, Pj-2, Pj-9, and *S. divaricata* formed their own subclade in the ITS2 tree (Fig. 4B and Supplementary Figure S3). Additionally, the Pj-1, Pj-2, and Pj-9 formed their own subclade distinct from other PJ accessions in the NR-tree and the ITS tree. This result implies that SNPs located in the ITS regions, both ITS1 and ITS2, affected the tree topology, dividing PJ with *S. divaricata* and *D. carota.* Conventionally, the ITS regions were utilized for identification of species delimitation as barcoding regions[22]. On the other hand, SNPs in the ITS2 region affected tree topology, dividing Pj-1, Pj-2, and Pj-9 from other PJ accessions. As detected in Table 2, Pj-9 and Pj-10 were genetically far from other PJ accessions. This result implies that the ITS tree reflects the genetic distance among PJ accessions. Detecting heterozygous positions in 45S nrDNA also revealed that all four SNPs included at least two heterozygous PJ accessions. Two S-type accessions, Pj-4 and Pj-7, had heterozygous positions in both of the two SNPs located in the ITS regions (5.8S and ITS2). Additionally, Pj-3 to Pj-7 had heterozygous positions in the SNPs located in the 26S regions (Supplementary Table S2). We assumed that the nuclear genome was frequently cross-hybridized with the others, such as Pj-1 and Pj-2, among populations, although the plastome of Pj-9 was quite different from the others. To validate the possibility of the hybridization between PJ and the phylogenetic location of the PJ accessions, further research is needed, such as artificial pollination, genetic analysis of lower latitude PJ germplasm, and the development of molecular markers derived from the nuclear genome.

Molecular markers for plastome type identification in PJ

We developed five KASP markers to identify the plastome types of PJ. The 55 PJ accessions collected from various regions in Korea were classified into six distinct types. Even though our collection was limited to South Korea, the plastome comparisons and marker validation revealed a high degree of diversity among PJ. The PJ habitats are found in the continental coastal regions of East Asia, including mainland China and Korea[1]. Therefore, it is anticipated that this plant presents a high level of genetic diversity, necessitating a comprehensive analysis to understand its genomic features. Along with the use of roots for traditional medicine, PJ has also been recently utilized as a vegetable[2,3]. Therefore, marker-assisted breeding is necessary for breeding programs aimed at cultivating this highly diverse species as a medicinal plant and vegetables. The markers developed in this study could detect novel plastome types in our collections, such as B and C types. Moreover, it would be

useful for tracing the maternal origin of the germplasm and providing basic molecular tools for the classification of diverse PJ collections.

Materials and methods

Plant materials and DNA extraction

A total of 55 PJ accessions were collected from various institutes and wild habitats in Korea (Supplementary Figure S7 and Supplementary Table S5). Nine and seven accessions were provided by Hantaek botanical garden and the Medicinal Plant Garden of the College of Pharmacy, SNU, respectively. Three accessions were purchased from seed companies. Thirty-six accessions were collected from the wild habitats of Korea, including Ulleung-gun, Wando-gun, Yeosu-si, Jeju island, Geoje-si, Hamyang-gun, Haenam-gun, Goheung-gun, and Taean-gun. According to the national legislation, there are no specific permissions for use of purchases and wild accessions in this study. Genomic DNA was extracted using the modified cetyltrimethylammonium bromide (CTAB) method and the Exgene Plant SV Midi Kit (Geneall Biotechnology, Seoul, Korea) following the manufacturer's protocol[56]. Using the Nanodrop 1000 spectrometer (Thermo Fisher Scientific, USA) and agarose gel electrophoresis, the quantity and quality of the extracted DNA were confirmed. The genomic DNA of seven accessions was used to generate paired-end (PE) libraries. Six libraries were sequenced by the Illumin Miseq platform and one library (Pj-4) was sequenced by the Illumina Hiseq platform. Library preparation and sequencing services were provided by Phyzen (www.phyzen.com, Seongnam, South Korea).

Assembly and annotation of plastome and 45S nrDNA sequences

The *de novo* assembly of the low coverage whole genome sequence (dnaLCW) method was used to assemble the seven new plastome and 45S nrDNA sequence of PJ[17,23]. In summary, raw Illumina reads were trimmed using an embedded trimming tool within the CLC assembly cell (ver.4.21, CLC Bio, Denmark). The trimmed reads were assembled into contigs by *de novo* assembly using the CLC assembly cell. Contigs which have similarity to the reference sequence (*P. japonicum*, KU866530) were extracted by MUMer and BLASTZ[57,58]. Finally, plastome assembly was completed by manual curation. Annotating the plastome of seven PJ accessions was conducted by GeSeq and manually curated by Artemis[59,60]. For assembly validation, raw reads were remapped to the assembled sequences using an embedded mapping tool within the CLC assembly cell. The mapping results were visualized by JBrowse2[61].

The 45S nrDNA was assembled in a similar manner. Contigs which have similarity to the reference (*P. japonicum*, KX757776) were extracted. The start and end positions of each subunit of 45S nrDNA were determined by reference (*P. japonicum*, KX757776), each subunit sequence of *Daucus carota* (LOC135147507, LOC135147597, and LOC135152502), RNAmmer, and BLASTN (https://blast.ncbi.nlm.nih.gov/)[62,63].

Comparative analysis

To understand plastome diversity among ten PJ accessions, whole plastome sequences without IRa region were aligned using PRANK with the +F option[64,65]. Using the alignment tool embedded in MEGA X, aligned sequences were manually curated[66]. Afterwards, plastome gene map with its variant positions was drawn by circos-0.69-9 (http://circos.ca/)[67]. Plastome gene map for comparing boundary regions of the IR region was drawn by OGDRAW[68]. The Pi value calculation to detect divergence hotspots was conducted by DnaSP v6 using sliding window method (Window size: 600 bp, Sliding size: 200 bp)[69]. MISA and Tandem Repeat Finder (TRF) were used for repeat analysis of the PJ accessions[70,71].

Coding-sequence analysis

To determine the CDS variants, CDS were extracted from each plastome using FeatureExtract[72]. Among all the CDS, a total of 76 commonly shared CDS in the ten PJ plastomes and other related species were extracted. Each CDS was aligned by PRANK using the +F and translate option, and the alignment tool embedded in MEGA X was used for manual curation[64–66].

Phylogenetic analysis

Phylogenetic analyses were conducted to understand phylogenetic relationships of PJ and its related species. The 76 aligned CDS of the plastome were concatenated into a single matrix for reconstructing the PT-tree. Afterwards, concatenated a single matrix was trimmed by Gblocks with default parameters[73].

To select the best substitution model for each data set, jModelTest version 2.1.10 was used with Akaike Information Criterion (AIC) analysis[74]. The ML trees for each dataset were drawn by RAxML version 8.2.12 using an ML through rapidbootstrap with 1,000 replicates[75]. The BI trees for each dataset were drawn by MrBayes v.3.2.7 (ngen = 1,000,000, samplefreq = 200, burninfrac = 0.25)[76]. The GTR + Γ + I was used for reconstructing PT-tree. The comparison between PT-tree and NR-tree was conducted in the same manner as PT-tree reconstruction using GTR + Γ.

Divergence time and nucleotide substitution rate estimations

MCMCtree and codeml embedded in the PAML version 4.9j were used to estimate divergence time and nucleotide substitution rates[77]. The calibration point between Apiaceae and *P. ginseng* in Araliaceae was 60.19 Mya according to the result in previous study[78]. The second calibration point between *D. carota* and other species in Apiaceae was 19.30 Mya as minimum age according to the result in a previous study[5]. Nucleotide substitution rates were calculated using pairwise comparison with tree topology of the PT-tree (runmode:-2, model: 0, and NSsites: 0). The d_S values were calculated using *P. ginseng* as the outgroup.

KASP marker development

To develop the KASP markers, we utilized SNP positions obtained through the comparisons of plastomes among PJ accessions. The candidate SNP and the 50 bp of both flanking sequences were sent to LGC (LGC genomics, Hoddesdon, UK) for primer development. Using the 55 PJ accessions, five developed markers were validated. PCR amplification was conducted with 5 μL of KASP master mix, 0.07 μL of KASP Assay mix, 2 μL (20 ng) of DNA, and 2.93 μL of distilled water. The PCR conditions were as follows: hot-start activation at 94 °C for 15 min, 10 cycles of touchdown (94 °C for 20 s and touchdown at 61 °C, decreasing 0.6 °C per cycle), and then 29 cycles of amplification (94 °C for 20 s and 55 °C for 60 s). PCR amplification and genotyping were conducted using a Roche LightCycler® 480.

Data availability

The data used in this study can be accessed through NCBI Genbank. Genbank accession numbers of plastomes and 45S nrDNAs, Pj-3 to Pj-9, are PQ315775~315781 and PQ288681~288687, respectively. Raw data used for the assembly can be accessed through NCBI SRA. SRA accession numbers, Pj-3 to Pj-9, are SRX26495067~26495073. The accessions of Plastome and 45S nrDNA are also described at Table 1 and Supplementary Table S1, respectively.

Received: 20 September 2024; Accepted: 24 December 2024
Published online: 02 January 2025

References

1. Menglan, S. et al. Apiaceae. In *Flora of China* Vol 14 (eds. Wu, Z. Y., Raven, P. H. & Hong, D. Y.) 182–192 (Science Press and Missouri Botanical Garden Press, 2005).
2. Chun, J. M. et al. *Peucedanum japonicum* extract attenuates allergic airway inflammation by inhibiting Th2 cell activation and production of pro-inflammatory mediators. *J. Ethnopharmacol.* **211**, 78–88 (2018).
3. Kim, J. K., Kang, H. M., Jang, D. C., Na, J. K. & Choi, K. Y. Effect of light intensity and temperature on the growth and functional compounds in the baby leaf vegetable plant *Peucedanum japonicum* Thunb. *Hortic. Sci. Technol.* **38**, 822–829 (2020).
4. Downie, S. R., Spalik, K., Katz-Downie, D. S. & Reduron, J. P. Major clades within Apiaceae subfamily Apioideae as inferred by phylogenetic analysis of nrDNA ITS sequences. *Plant Divers. Evol.* **128**, 111 (2010).
5. Wen, J. et al. A transcriptome-based study on the phylogeny and evolution of the taxonomically controversial subfamily Apioideae (Apiaceae). *Ann. Bot.* **125**, 937–953 (2020).
6. Liu, C. K., Lei, J. Q., Jiang, Q. P., Zhou, S. D. & He, X. J. The complete plastomes of seven *Peucedanum* plants: Comparative and phylogenetic analyses for the *Peucedanum* genus. *BMC Plant Biol.* **22**, 1–14 (2022).
7. Wen, J. et al. Backbone phylogeny and evolution of Apioideae (Apiaceae): New insights from phylogenomic analyses of plastome data. *Mol. Phylogenetics Evol.* **161**, 107183 (2021).
8. Okuyama, Y. et al. Nonuniform concerted evolution and chloroplast capture: Heterogeneity of observed introgression patterns in three molecular data partition phylogenies of Asian *Mitella* (Saxifragaceae). *Mol. Biol. Evol.* **22**, 285–296 (2005).
9. Wicke, S., Schneeweiss, G. M., Depamphilis, C. W., Müller, K. F. & Quandt, D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol.* **76**, 273–297 (2011).
10. Greiner, S., Sobanski, J. & Bock, R. Why are most organelle genomes transmitted maternally?. *BioEssays* **37**, 80–94 (2015).
11. Chung, K. P., Gonzalez-Duran, E., Ruf, S., Endries, P. & Bock, R. Control of plastid inheritance by environmental and genetic factors. *Nat. Plants* **9**, 68–80 (2023).
12. Sakamoto, W. & Takami, T. Plastid inheritance revisited: Emerging role of organelle DNA degradation in angiosperms. *Plant Cell Physiol.* **65**, 484–492 (2024).
13. Wolfe, K. H., Li, W. H. & Sharp, P. M. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci.* **84**, 9054–9058 (1987).
14. Drouin, G., Daoud, H. & Xia, J. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol. Phylogenet. Evol.* **49**, 827–831 (2008).
15. Li, X. et al. Plant DNA barcoding: From gene to genome. *Biol. Rev.* **90**, 157–166 (2015).
16. Smith, D. R. Mutation rates in plastid genomes: They are lower than you might think. *Genome Biol. Evol.* **7**, 1227–1234 (2015).
17. Kim, K. et al. Comprehensive survey of genetic diversity in chloroplast genomes and 45S nrDNAs within *Panax ginseng* species. *PLoS ONE* **10**, e0117159 (2015).
18. Park, Y. S. et al. Analysis of the complete plastomes and nuclear ribosomal DNAs from *Euonymus hamiltonianus* and its relatives sheds light on their diversity and evolution. *PLoS ONE* **17**, e0275590 (2022).
19. Lee, M. et al. Authentication of *Allium ulleungense*, *A. microdictyon* and *A. ochotense* based on super-barcoding of plastid genome and 45S nrDNA. *PLoS ONE* **18**, e0294457 (2023).
20. Kovarik, A. et al. Concerted evolution of 18–5.8–26S rDNA repeats in Nicotiana allotetraploids. *Biol. J. Linn. Soc.* **82**, 615–625 (2004).
21. Ganley, A. R. & Kobayashi, T. Highly efficient concerted evolution in the ribosomal DNA repeats: Total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome Res.* **17**, 184–191 (2007).
22. Álvarez, I. & Wendel, J. F. Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* **29**, 417–434 (2003).
23. Kim, K. et al. Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of Oryza AA genome species. *Sci. Rep.* **5**, 15655 (2015).
24. Jin, J. J. et al. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **21**, 1–31 (2020).
25. Goulding, S. E., Wolfe, K., Olmstead, R. & Morden, C. Ebb and flow of the chloroplast inverted repeat. *Mol. Gen. Genet.* **252**, 195–206 (1996).
26. Downie, S. R. & Jansen, R. K. 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. *Syst. Bot.* **40**, 336–351 (2015).
27. Zhu, A., Guo, W., Gupta, S., Fan, W. & Mower, J. P. Evolutionary dynamics of the plastid inverted repeat: The effects of expansion, contraction, and loss on substitution rates. *New Phytol.* **209**, 1747–1756 (2016).
28. Henriquez, C. L. et al. Complete chloroplast genomes of *Anthurium huixtlense* and *Pothos scandens* (Pothoideae, Araceae): Unique inverted repeat expansion and contraction affect rate of evolution. *J. Mol. Evol.* **88**, 562–574 (2020).
29. Asaf, S. et al. Expanded inverted repeat region with large scale inversion in the first complete plastid genome sequence of *Plantago ovata*. *Sci. Rep.* **10**, 1–16 (2020).

30. Lee, H. O. et al. Dynamic chloroplast genome rearrangement and DNA barcoding for three Apiaceae species known as the medicinal herb "Bang-Poong". *Int. J. Mol. Sci.* **20**, 2196 (2019).
31. Wang, M. et al. Phylogenomic and evolutionary dynamics of inverted repeats across *Angelica* plastomes. *BMC Plant Biol.* **21**, 1–12 (2021).
32. Wei, N. et al. Plastome evolution in the hyperdiverse genus *Euphorbia* (Euphorbiaceae) using phylogenomic and comparative analyses: Large-scale expansion and contraction of the inverted repeat region. *Front. Plant Sci.* **12**, 712064 (2021).
33. Plunkett, G. M. & Downie, S. R. Expansion and contraction of the chloroplast inverted repeat in Apiaceae subfamily Apioideae. *Syst. Bot.* **25**, 648–667 (2000).
34. Ren, T. et al. Molecular evolution and phylogenetic relationships of *Ligusticum* (Apiaceae) inferred from the whole plastome sequences. *BMC Ecol. Evol.* **22**, 1–14 (2022).
35. Kelchner, S. A. & Wendel, J. F. Hairpins create minute inversions in non-coding regions of chloroplast DNA. *Curr. Genet.* **30**, 259–262 (1996).
36. Knox, E. B. The dynamic history of plastid genomes in the Campanulaceae *sensu lato* is unique among angiosperms. *Proc. Natl. Acad. Sci.* **111**, 11097–11102 (2014).
37. Weng, M. L., Ruhlman, T. A. & Jansen, R. K. Expansion of inverted repeat does not decrease substitution rates in *Pelargonium* plastid genomes. *New Phytol.* **214**, 842–851 (2017).
38. Mower, J. P. & Vickrey, T. L. Structural diversity among plastid genomes of land plants. *Adv. Bot. Res.* **85**, 263–292 (2018).
39. Zhou, S., Ma, K., Mower, J. P., Liu, Y. & Zhou, R. Leaf variegation caused by plastome structural variation: An example from *Dianella tasmanica*. *Hortic. Res.* **11**, uhae009 (2024).
40. Wang, R. J. et al. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol. Biol.* **8**, 36 (2008).
41. Li, F. W., Kuo, L. Y., Pryer, K. M. & Rothfels, C. J. Genes translocated into the plastid inverted repeat show decelerated substitution rates and elevated GC content. *Genome Biol. Evol.* **8**, 2452–2458 (2016).
42. Guo, Y. Y., Yang, J. X., Bai, M. Z., Zhang, G. Q. & Liu, Z. J. The chloroplast genome evolution of *Venus slipper* (Paphiopedilum): IR expansion, SSC contraction, and highly rearranged SSC regions. *BMC Plant Biol.* **21**, 1–14 (2021).
43. Hotta, M. & Shiuchi, T. Notes on the flora of the Ryukyu Islands. 1. Two new varieties from the Tokara Islands, *Peucedanum japonicum* Thunb. Var. *latifolium* (Umbelliferae) and *Hydrangea involucrata* Sieb. Var. *tokarensis* (Hydrangeaceae). *J. Jap. Bot.* **71**, 183–187 (1996).
44. Seo, A. & Hotta, M. Taxonomical notes on plants of southern Japn V. Infraspecific variation of *Peucedanum japonicum* Thunb. (Umbelliferae) in Kyushu and Ryukyu Islands. *Acta Phytotax. Geobot.* **51**, 99–116 (2000).
45. Kao, M. T. Umbelliferae *in Flora of Taiwan*. Vol. 3 (ed. Huang, T. C.) 1010 (Department of Botany, National Taiwan University, Taipei, 1993).
46. Yamada, Y. et al. Two new khellactone esters from *Peucedanum japonicum* THUNB. *Tetrahedron Lett.* **29**, 2513–2516 (1974).
47. Chang-yih, D., Shang-Kwei, W. & Yang-Chang, W. Cytotoxic pyranocoumarins from the aerial parts of *Peucedanum japonicum*. *Phytochemistry* **30**, 2812–2814 (1991).
48. Chang-Yih, D., Shang-Kwei, W. & Yang-Chang, W. Cytotoxic pyranocoumarins from roots of *Peucedanum japonicum*. *Phytochemistry* **31**, 1829–1830 (1992).
49. Jong, T. T., Hwang, H. C., Jean, M. Y., Wu, T. S. & Teng, C. M. An antiplatelet aggregation principle and X-ray structural analysis of cis-khellactone diester from *Peucedanum japonicum*. *J. Nat. Prod.* **55**, 1396–1401 (1992).
50. Hatusima, S. An enumeration of the plants of Batan Island, N. Philippines. *Mem. Fac. Agric. Kagoshima Univ.* **5**, 13–70 (1966).
51. Madulid, D. A. & Agoo, E. M. G. Notes on the economic plants of Batanes: *Citrus species* and *Phoenix loureiroi* var. *loureiroi*. *Bull. Natl. Mus. Ethnol.* **34**, 191–205 (2009).
52. Kamei, T. Fauna and flora of the Japanese islands in the last glacial time. *Quat. Res.* **20**, 191–205 (1981).
53. Tsukada, M. A vegetation map in the Japanese Archipelago approximately 20,000 years BP. *Jpn. J. Ecol.* **34**, 203–208 (1984).
54. Kuroda, T. Nansei-shoto no shokuseishi. In *Nihon-retto shokuseishi* (eds.Yasuda, Y. & Miyoshi, N.) 162–175 (Asakura shoten, 1998).
55. Chiang, T. Y. & Schaal, B. A. Phylogeography of plants in Taiwan and the Ryukyu Archipelago. *Taxon* **55**, 31–41 (2006).
56. Allen, G. C., Flores-Vergara, M., Krasynanski, S., Kumar, S. & Thompson, W. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* **1**, 2320–2325 (2006).
57. Kurtz, S. et al. Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).
58. Schwartz, S. et al. Human–mouse alignments with BLASTZ. *Genome Res.* **13**, 103–107 (2003).
59. Carver, T., Harris, S. R., Berriman, M., Parkhill, J. & McQuillan, J. A. Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28**, 464–469 (2012).
60. Tillich, M. et al. GeSeq–versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **45**, W6–W11 (2017).
61. Diesh, C. et al. JBrowse 2: A modular genome browser with views of synteny and structural variation. *Genome Biol.* **24**, 74 (2023).
62. Lagesen, K. et al. RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108 (2007).
63. Camacho, C. et al. BLAST+: Architecture and applications. *BMC Bioinform.* **10**, 1–9 (2009).
64. Löytynoja, A. & Goldman, N. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* **320**, 1632–1635 (2008).
65. Löytynoja, A. Phylogeny-aware alignment with PRANK in *Multiple sequence alignment methods* (ed Russell, D. J.) 155–170 (Springer, 2014).
66. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
67. Krzywinski, M. et al. Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
68. Greiner, S., Lehwark, P. & Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **47**, W59–W64 (2019).
69. Rozas, J. et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **34**, 3299–3302 (2017).
70. Beier, S., Thiel, T., Münch, T., Scholz, U. & Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **33**, 2583–2585 (2017).
71. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
72. Wernersson, R. FeatureExtract—extraction of sequence annotation made easy. *Nucleic Acids Res.* **33**, W567–W569 (2005).
73. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
74. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* **9**, 772–772 (2012).
75. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
76. Ronquist, F. et al. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
77. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
78. Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L. & Hernández-Hernández, T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* **207**, 437–453 (2015).

Acknowledgements
This study was supported by "Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ017134)" Rural Development Administration, and the Korea Health Technology R&D Project (grant number: HI23C1420) through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea.

Author contributions
T.J.Y and H.J.J contributed to the original concepts and design. S.I.H, H.J.J, Y.W.C, J.P.L and J.T.K collected and conserved the materials. Data analysis was performed by H.J.J, Y.S.P, J.S.K and H.S.P. Experiments were designed by H.J.J, Y.S.P, and J.Y.P. Data visualization was performed by Y.S.P. The draft manuscript was written by H.J.J, Y.S.P, and T.J.Y. All authors read and approved the final manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-84540-8.

Correspondence and requests for materials should be addressed to T.-J.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.