diversity

MDPI

Article

Comparative Analysis and Phylogenetic Study of the Chloroplast Genome Sequences of Two Korean Endemic *Primula* Varieties

Sang-Chul Kim, Young-Ho Ha, Dong-Kap Kim, Dong Chan Son, Hyuk-Jin Kim * and Kyung Choi

Division of Forest Biodiversity, Korea National Arboretum, 509 Gwangneungsumogwon-ro, Soheul-eup, Pocheon-si 11186, Korea; majin01@korea.kr (S.-C.K.); yh0990@korea.kr (Y.-H.H.); botany7403@korea.kr (D.-K.K.); sdclym@korea.kr (D.C.S.); kchoi69@korea.kr (K.C.)

* Correspondence: jins77@korea.kr; Tel.: +82-31-540-8811

Abstract: *Primula* comprises many species of horticultural value. In Korea, six species grow in the wild. Yamazaki classified the variety *Primula modesta* var. *fauriei* into *P. modesta* var. *hannasanensis* and *P. modesta* var. *koreana* based on the differences in leaf morphology. We compared the chloroplast genome sequences of the two Korean endemic *Primula* varieties and found that both had the typical quadripartite structure of angiosperms. The chloroplast genome size of *P. modesta* var. *hannasanensis* is 154,772 bp, including an 85,238 bp large single-copy region and a 17,790 bp small single-copy region, whereas that of *P. modesta* var. *koreana* is 154,667 bp, including an 85,152 bp large single-copy region and a 17,771 bp small single-copy region. The inverted repeat region is 25,872 bp in both varieties. We predicted 129 genes—84 protein-coding genes, 8 rRNAs, and 37 tRNAs. We identified 536 single-nucleotide polymorphisms and 501 indels between the varieties. Phylogenetic analysis revealed that the two varieties formed a sister group in the clade *P. knuthiana–P. stenocalyx*. This study will contribute to phylogenetic, taxonomic, and evolutionary studies of the genus *Primula*; it will also contribute to the analysis of the genetic diversity of the two varieties, and to the development of identification markers.

Keywords: chloroplast genome; *Primula*; next-generation sequencing; comparative genomics; phylogenetics; repeat analysis

Citation: Kim, S.-C.; Ha, Y.-H.; Kim, D.-K.; Son, D.C.; Kim, H.-J.; Choi, K. Comparative Analysis and Phylogenetic Study of the Chloroplast Genome Sequences of Two Korean Endemic *Primula* Varieties. *Diversity* **2022**, *14*, 458. https://doi.org/10.3390/d14060458

Academic Editor: Michael Wink

Received: 13 April 2022
Accepted: 6 June 2022
Published: 7 June 2022

1. Introduction

The genus *Primula* L. comprises approximately 430 species worldwide. It is widespread in the northern hemisphere in the highlands of Asia, North America, Europe, and the Eastern Shino Himalayas [1]. Many *Primula* species are popular as garden species owing to their appealing flowers and long flowering duration. Molecular phylogenetic studies of *Primula* have been carried out primarily with nuclear and plastid genes and fragments in China, which has the largest distribution area of the species. This has considerably improved the comprehension of the phylogenetic history of *Primula*. However, the exact phylogenetic history remains uncertain given the occurrence of frequent hybridization events and interspecies morphological variations caused by translocation [2–4].

In Korea, there are six *Primula* species [5]. Among them, Yamazaki divided *Primula modesta* var. *fauriei* (Franch.) Takeda into *P. modesta* var. *koreana*, which grows wild in the mainland, and *P. modesta* var. *hannasanensis*, which grows wild in Jeju Island, based on variations in the morphology of the leaf base [6–9]. However, comparative genetic information analysis of the two varieties has not yet been conducted.

Chloroplasts (CPs) play important roles in photosynthesis. Plastid genomes are generally 115–165 kb long. They have a quadripartite molecular structure containing two inverted repeat (IR) regions (20–28 kb long) linking the large single-copy (LSC; 80–90 kb long) and small single-copy regions (SSC; 16–27 kb long; [10,11]). In most terrestrial plant

lineages, the CP genomes are similar in gene order, gene content, structure, and intron content. Owing to their relatively high substitution rates, CP genomes are an important source of genetic markers for the phylogenetic identification of species and population genetics [12].

In this study, we analyzed the CP genomes of two Korean endemic *Primula* variants, *P. modesta* var. *koreana* and *P. modesta* var. *hannasanensis*, which have not yet been reported. We compared the two variants to examine single nucleotide polymorphisms (SNPs), indels, and simple sequence repeat (SSR) polymorphisms to identify markers useful for DNA barcodes and phylogenetic analysis. Our findings may not only contribute to further studies on the evolutionary history, lineage, and taxonomy of the genus *Primula*, but also aid in helping to better understand its CP genome.

2. Materials and Methods

2.1. Material Collection, DNA Extraction, and Next-Generation Genome Sequencing

Samples of the two varieties (*P. modesta* var. *koreana* and *P. modesta* var. *hannasanensis*) were collected from Sinbul Mountain (Ulsan-si, Gyeongsangnam-do, Korea; N: 35°31′26.7″ E: 129°03′01.5″) and Halla Mountain (Jeju, Korea; N: 33°22′48.9″ E: 126°34′52.049″), respectively. The total genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen Inc., Valencia, CA, USA) following the manufacturer's instructions. DNA quality was evaluated using a NanoDrop 2000 microspectrophotometer (Thermo Fisher Inc., Waltham, MA, USA), and the quantity was confirmed using 1% agarose gel electrophoresis. Voucher specimens of the two *Primula* accessions have been deposited at the Herbarium of the Korea National Arboretum (*P. modesta* var. *koreana*, ESK20-040; *P. modesta* var. *hannasanensis*, ESK20-138). Next-generation paired-end sequencing was performed using the Illumina MiSeq platform (TruSeq DNA PCR-Free) according to the manufacturer's protocol (Macrogen Inc., Seoul, Korea).

2.2. CP Genome Assembly and Annotation

The CP DNA sequence data were filtered from the whole genome data using GetOrganelle [13]. The two CP genomes were assembled using Geneious Prime (Biomatters, Auckland, New Zealand) [14] and annotated using GeSeq [15]. Unannotated portions, such as exons and introns, were manually edited. Transfer (t)RNA sequences were confirmed using tRNAscan-SE v1.21 [16]. Genome maps were drawn using OrganellarGenomeDRAW (OGDRAW; [17]).

2.3. Genome Comparison

The complete CP genomes of the two *Primula* varieties were aligned using MAFFT [18] and compared using m-VISTA (http://genome.lbl.gov/vista/index.shtml, accessed on 1 August 2021) in the shuffle-LAGAN mode [19]. Genome junctions were visualized and compared using IRscope [20].

2.4. Divergent Hotspot Identification

CP genome polymorphisms were analyzed using DNA Sequence Polymorphism (DnaSP) v6 [21] to determine the nucleotide diversity (*Pi*) values and confirm highly variable sites. The sequences were aligned using MAFFT in Geneious Prime [14,18]. The variations (SNPs, insertions, and deletions (indels)) between the varieties were analyzed using Geneious Prime based on a minimum variant frequency criterium of 0.25. The compartments were separated into coding sequences (CDSs), tRNA, ribosomal (r)RNA, and intergenic spacers (IGSs).

2.5. Relative Synonymous Codon Usage Analysis

Relative synonymous codon usage (RSCU) was computed from the CDSs of the two *Primula* CP genomes. The DAMBE program was employed for the RSCU and codon frequency analyses [22].

2.6. SSR and Long Repeat Sequence Analysis

SSRs within the two *Primula* CP genomes were analyzed using the MISA Perl script (MIcroSAtellite; [23]) based on the following minimum repeats criteria: mononucleotide repeats, 10; dinucleotide repeats, 5; trinucleotide repeats, 4; and tetra-, penta-, and hexa-nucleotide repeats, 3. REPuter was used to locate four repeat types (forward, reverse, complementary, and palindromic) based on the following criteria: a minimum repeat size of 30 bp and sequence identity of 90% [24].

2.7. Phylogenetic Analysis

The complete CP genome sequences of 76 Primulaceae and 9 other Ericales (4 Sapotaceae, 4 Pentaphylacaceae, and 1 Polemoniaceae) species were obtained from the NCBI database and used for a maximum likelihood (ML) phylogenetic analysis (Table S1). In total, 77 CDSs from 85 species were aligned using MAFFT in PhyloSuite [25]. ModelFinder within the PhyloSuite program was used to determine the optimal alternative model [26]. The ML analysis was performed using IQ-Tree software (http://iqtree.cibiv.univie.ac.at/, accessed on 4 August 2021).

3. Results

3.1. Common Features of the CP Genomes

The CP genome lengths of the two variants were in similar ranges: *P. modesta* var. *koreana*, 154,667 bp (LSC; 85,152 bp, SSC; 85,152 bp, and IRs; 25,872 bp); *P. modesta* var. *hannasanensis*, 154,772 bp (LSC; 85,238 bp, SSC; 17,790 bp, and IRs; 25,872 bp). Moreover, they had typical quadripartite structures (Figure 1 and Table 1). The CP of *P. modesta* var. *hannasanensis* is the largest among the *Primula* CP genomes reported to date. The two CP genomes had identical gene contents (129 genes—84 protein-coding, 8 rRNA, and 37 tRNA genes; Table 2). Seventeen genes located in the IR regions included 6 protein-coding genes (*rpl2*, *rpl23*, *rps7*, *rps12*, *ndhB*, and *ycf2*) and 4 rRNA genes (*rrn4.5*, *rrn5*, *rrn16*, and *rrn23*). Eight tRNA genes (*trnA-UGC*, *trnG-UCC*, *trnK-UUU*, *trnI-GAU*, *trnI-CAU*, *trnL-UAA*, *trnN-GUU*, *trnR-ACG*, and *trnV-UAC*) and 10 protein-coding genes (*atpF*, *ndhA*, *ndhB*, *rpl2*, *rpl16*, *rps12*, *rps16*, *rpoC1*, *petB*, and *petD*) contained one intron, and two protein-coding genes (*clpP1* and *pafI*) contained two introns (Table 2). *rps12* was confirmed to be a trans-spliced gene consisting of three exons: exon 1, found in the LSC region, and exons 2 and 3, located in the IR regions.

Table 1. Information of the two Korean *Primula* CP genomes.

Feature	*P. modesta* var. *koreana*	*P. modesta* var. *hannasanensis*
Accession number	MZ779113	MZ779112
Genome size [GC(%)]	154,667 [37.0]	154,772 [36.9]
LSC [GC(%)]	85,152 [34.8]	85,238 [34.8]
SSC [GC(%)]	17,771 [30.3]	17,790 [30.3]
IR [GC(%)]	25,872 [42.7]	25,872 [42.7]

3.2. Comparison of the CP Genomes of the Two Primula Varieties

We compared the gene sequence and CP genome content between the varieties via m-VISTA, and found that the CP genomes were nearly identical with coding and IR regions more conserved than non-coding, LSC, and SSC regions (Figure 2). The boundary structures of the two genomes were compared with those of four other *Primula* CP genomes located in the same group in the ML tree (see Section 3.6) The overall identity of the CP genomes was confirmed at all junctions, including JLB (LSC/IRb), JSB (IRb/SSC), JSA (SSC/IRa), and JLA (IRa/LSC). However, *trnH-GUG* was located in the LSC region 3 bp away from the JLA junction in four *Primula* species, and 11 and 1 bp away in *P. pulchella* Franch. and *P. knuthiana* Pax, respectively. A 978 bp region of *ycf1* was located in IRa and a 7 bp region of *ndhF* was integrated into IRb. At the JLB junction, a 41 bp region of *rps19* was

present in the IR region of all five species, except that of *P. knuthiana* (100 bp in the IR) (Figure 3).

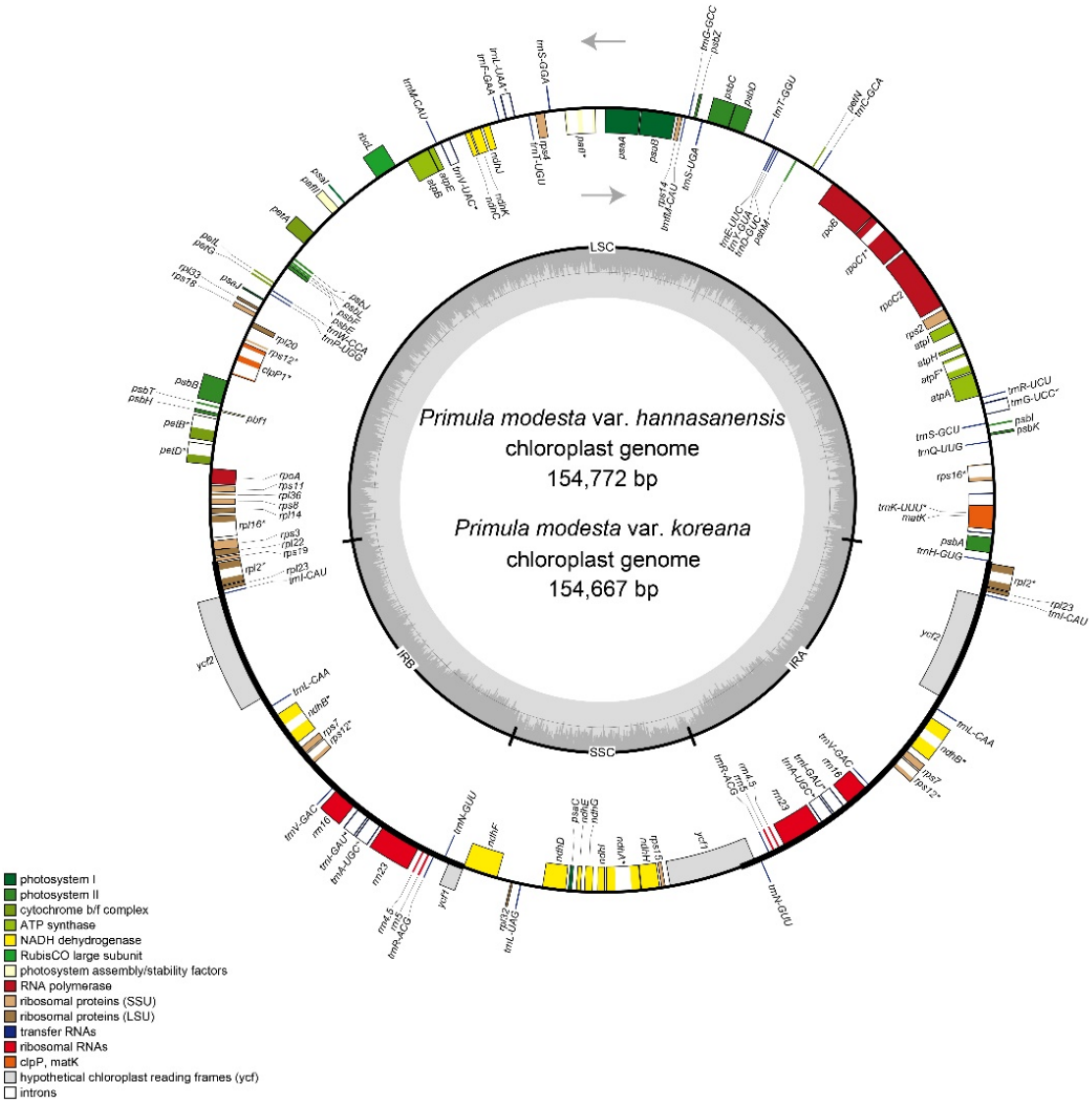

Figure 1. Complete chloroplast (CP) genome map of *Primula modesta* var. *koreana* and *P. modesta* var. *hannasanensis*. Genes drawn outside and inside the circle are transcribed counterclockwise and clockwise, respectively. Functional categories of various genes are marked in color. Light gray corresponds to AT content and dark gray in the inner circle corresponds to GC content.

3.3. Divergent Hotspots in the Primula CP genomes

Overall, 536 SNPs were found in the CP genomes of the two *Primula* varieties. In total, 205 (38.2%) SNPs were located in the CDS regions, and 331 (61.8%) were located in the IGS regions and introns. In addition, 501 indels were identified. The level of sequence divergence was determined by calculating the *Pi* values for the CP genomes of the two varieties (Figure 4 and Table S2). The *Pi* values for SNPs in the CDSs ranged from 0.00045 (*psaB*) to 0.01465 (*rps15*), with an average of 0.0023. The *Pi* values for SNPs in the IGSs ranged from 0.00146 (*ndhB* intron) to 0.03318 (*psbA~trnK-UUU*), with an average of 0.005. SNPs were identified in the *trnY-GUA*, 16S rRNA, and 23S rRNA genes. CDSs with a *Pi* value higher than 0.008 were identified as *cemA*, *rpl33*, *rps11*, and *rps15*. The regions with the highest *Pi* values in the IGSs were identified as the *psbA~trnK-UUU*, *psbZ~trnG-GCC*, *ndhD~psaC*, and *rps15~ycf1* regions (average *Pi* value: 0.0259).

Table 2. CP genome gene content and functional classification in *Primula modesta* var. *koreana* and *P. modesta* var. *hannasanensis*.

Gene Category	Gene Group	Gene Names
Self-replication	Large subunit ribosomal proteins	*rpl2*(×2) *, *rpl14*, *rpl16* *, *rpl20*, *rpl22*, *rpl23*(×2), *rpl32*, *rpl33*, *rpl36*
	DNA dependent RNA polymerase	*rpoA*, *rpoB*, *rpoC1* *, *rpoC2*
	Small subunit ribosomal proteins	*rps2*, *rps3*, *rps4*, *rps7*(×2), *rps8*, *rps11*, *rps12*(×2) *, *rps14*, *rps15*, *rps16* *, *rps18*, *rps19*
	rRNAs	*rrn4.5S*(×2), *rrn5S*(×2), *rrn16S*(×2), *rrn23S*(×2)
	tRNAs	*trnA-UGC*(×2) *, *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnfM-CAU*, *trnG-GCC*, *trnG-UCC* *, *trnH-GUG*, *trnI-GAU*(×2) *, *trnI-CAU*(×2), *trnK-UUU* *, *trnL-CAA*(×2), *trnL-UAA* *, *trnL-UAG*, *trnM-CAU*, *trnN-GUU*(×2), *trnP-UGG*, *trnQ-UUG*, *trnR-ACG*(×2), *trnR-UCU*, *trnS-GCU*, *trnS-GGA*, *trnS-UGA*, *trnT-GGU*, *trnT-UGU*, *trnV-GAC*(×2), *trnV-UAC* *, *trnW-CCA*, *trnY-GUA*
Photosynthesis	Subunits of ATP synthase	*atpA*, *atpB*, *atpE*, *atpF* *, *atpH*, *atpI*
	Subunits of NADH-dehydrogenase	*ndhA* *, *ndhB*(×2) *, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*
	Subunits of cytochrome b/f complex	*petA*, *petB* *, *petD*, *petG*, *petL*, *petN*
	Subunits of photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*
	Subunits of photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbT*, *psbZ*
	Subunit of rubisco	*rbcL*
	Photosystem assembly factors	*pafI* **, *pafII*
	Photosystem biogenesis factor	*pbf1*
Other genes	Subunit of acetyl-CoA-carboxylase	*accD*
	C-type cytochrome synthesis gene	*ccsA*
	Envelop membrane protein	*cemA*
	ATP-dependent protease subunit P	*clpP1* **
	Maturase	*matK*
Unknown function	Conserved open reading frames	*ycf1*, *ycf2*(×2)

Note: (×2), two gene copies in IRs; *, gene containing a single intron; **, gene containing two introns.

3.4. Relative Synonymous Codon Usage Analysis

A total of 78 CDSs from the CP genomes of the two *Primula* varieties were used to estimate the frequency of relative synonymous codon usage (except for the stop codons UAA, UAG, and UGA). In total, 22,696 codons were detected in *P. modesta* var. *koreana* and 22,691 in *P. modesta* var. *hannasanensis*. The two varieties showed similar results. Leucine was the most abundant amino acid (10.65%), whereas cysteine was the least abundant (1.06%, *P. modesta* var. *koreana*; 1.07%, *P. modesta* var. *hannasanensis*). The most used codon was AUU (998), encoding isoleucine, and the least used codon was UGC (57), which encoded cysteine in both varieties. The RSCU frequency analyses of the two CP genomes revealed a bias in codon usage; 29 amino acids had RSCU > 1. Two amino acids, methionine (AUG) and tryptophan (UGG), did not display codon usage bias (RSCU = 1.00). In both varieties, the highest RSCU value was recorded for GCU (1.879, *P. modesta* var. *koreana*; 1.872, *P. modesta* var. *hannasanensis*), which encodes alanine, and the lowest for UAC (0.359, *P. modesta* var. *koreana*; 0.355, *P. modesta* var. *hannasanensis*), which encodes tyrosine (Figure 5).

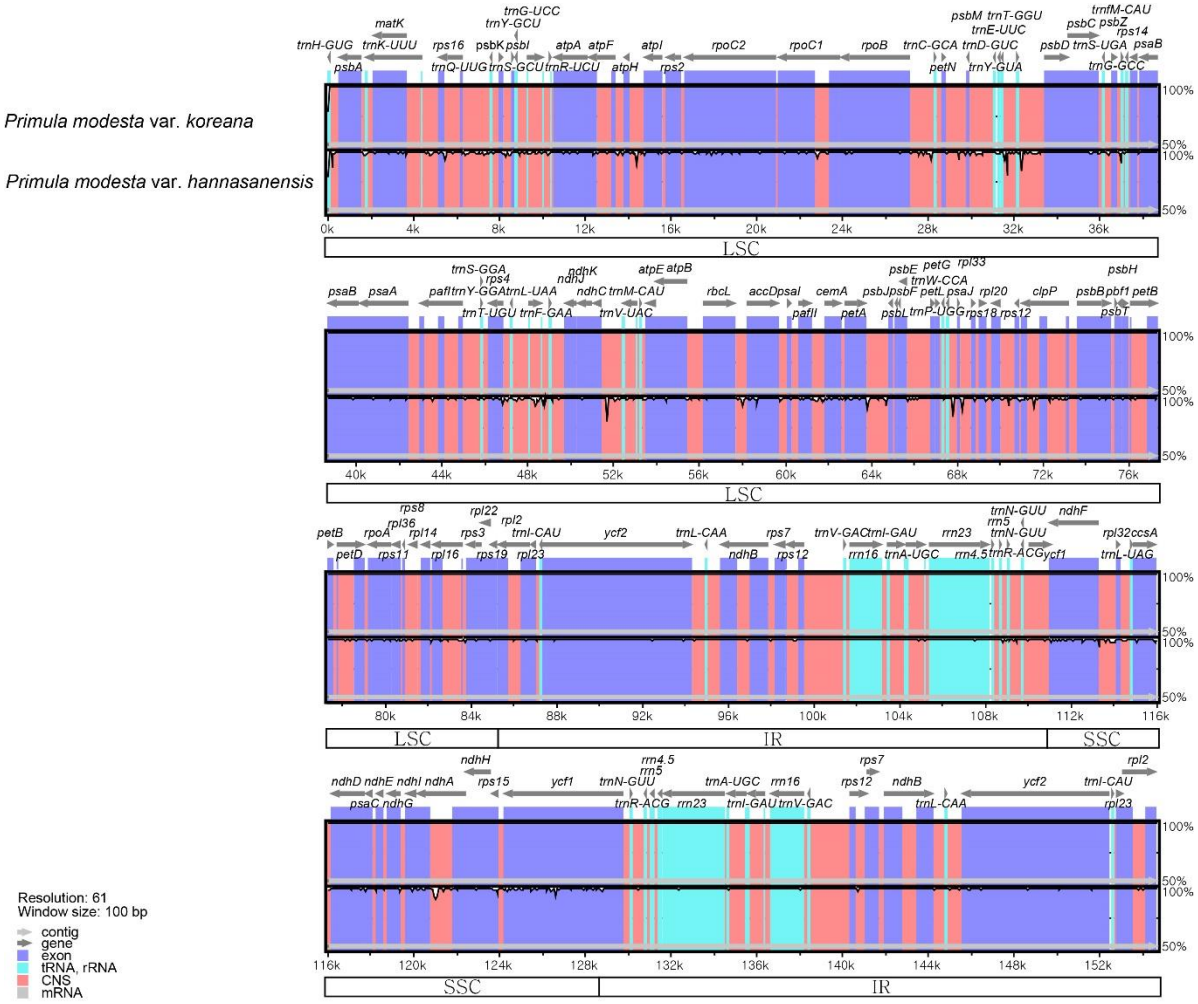

Figure 2. Comparison of the CP genomes of the two Korean endemic *Primula* varieties using m-VISTA. Gray arrows above the alignments indicate gene orientation, purple bars represent coding sequences (CDSs) and exons, and blue bars represent RNAs. The y-axis shows the identity from 50% to 100%.

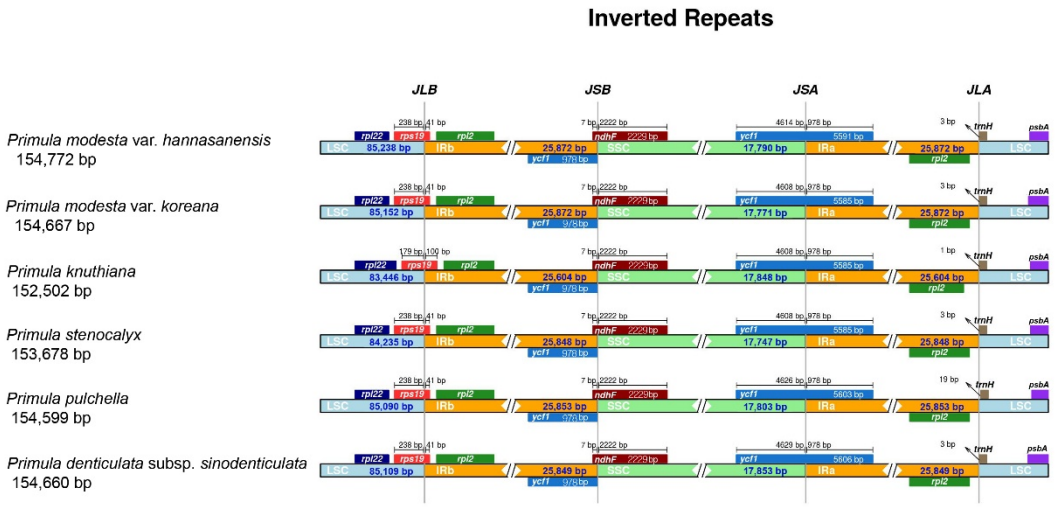

Figure 3. Comparison of the boundary distances between the adjacent genes and junctions of large single-copy (LSC), small single-copy (SSC), and two inverted repeat (IR) regions among the CP genomes of six *Primula* species. Genes are displayed as colored boxes. The figure is not representative of sequence length but shows the relative differences in the IR/SC borders.

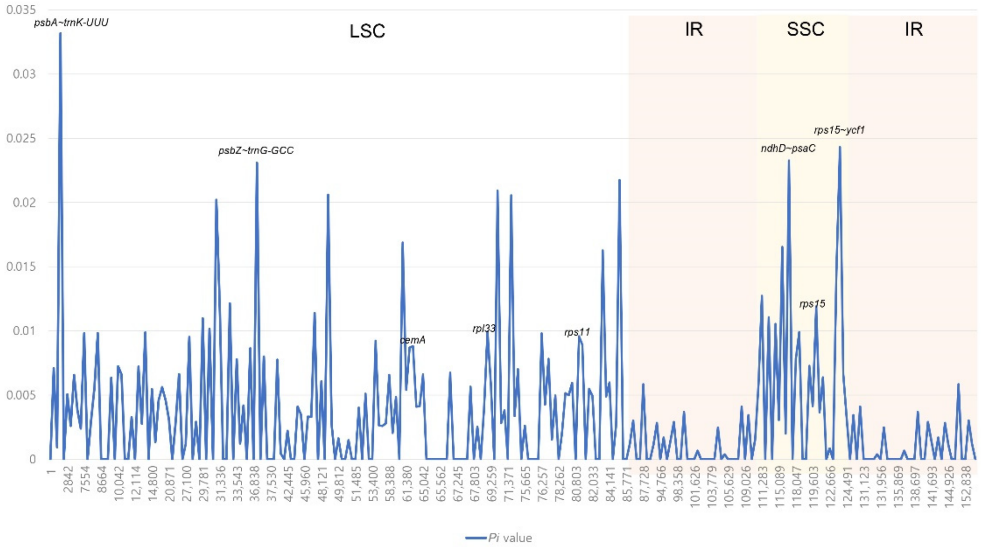

Figure 4. Sliding window analysis of the complete CP genome nucleotide diversity (*Pi*) between the Korean *Primula* varieties.

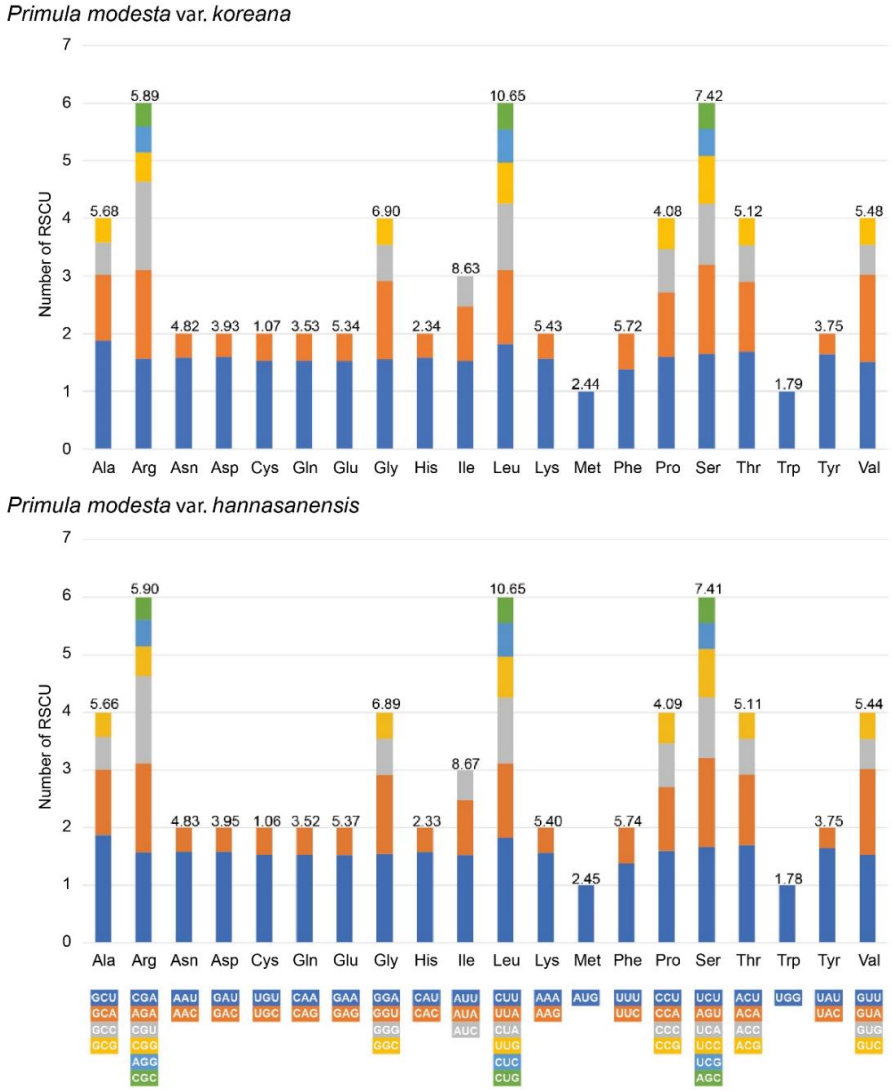

Figure 5. Relative synonymous codon usage (RSCU) analysis of 20 amino acids in all CDSs of the complete CP genomes of the two Korean *Primula* varieties.

3.5. SSR and Long Repeat Analyses

In total, 58 SSRs were identified in *P. modesta* var. *hannasanensis* and 53 in *P. modesta* var. *koreana*. Both varieties had a high number of mononucleotide repeats. There were 4 (*P. modesta* var. *koreana*) and 6 (*P. modesta* var. *hannasanensis*) dinucleotide repeats and 2 identical tetranucleotide and hexanucleotide repeats; a single pentanucleotide repeat was identified in *P. modesta* var. *koreana*. Most SSRs consisted of the A/T motif rather than the G/C motif (Table 3 and Table S3).

Table 3. Information on the types and numbers of SSRs in the CP genomes of the two Korean *Primula* varieties.

SSR Type	Repeat Unit	*Primula modesta* var. *koreana*	*Primula modesta* var. *hannasanensis*	Total
Mononucleotide	A/T	43	48	92
	C/G	1	0	
Dinucleotide	AT/AT	4	6	10
Tetranucleotide	AGAT/ATCT	2	2	4
Pentanucleotide	AAAGT/ACTTT	1	0	1
Hexanucleotide	AAATAG/ATTTCT	1	1	4
	AAGATG/ATCTTC	1	1	
Total		53	58	111

The long repeat analysis identified more forward and palindromic repeats than reverse and complementary repeats in the two *Primula* varieties. A total of 30 long repeats were identified in *P. modesta* var. *koreana* and 32 in *P. modesta* var. *hannasanensis*. Only one reverse repeat was found in *P. modesta* var. *hannasanensis*. The length of most repeats ranged from 30 to 39 bp, whereas the largest repeat was 50 bp long (*P. modesta* var. *koreana*). The locations and numbers of iterations of long repeats are listed in Table 4 and Table S4.

Table 4. Types and numbers of repeats in the CP genomes of the two *Primula* varieties, as identified using REPuter.

Type of Repeat	*Primula modesta* var. *koreana*	*Primula modesta* var. *hannasanensis*
Forward	12	13
Reverse	0	1
Palindromic	18	18
Total	30	32
Length of repeat (bp)		
30–39	22	24
40–49	7	8
50–59	1	

3.6. Phylogenetic Analysis

The phylogenetic analysis was performed using the ML method and 77 genes from 85 Ericales CP genomes (Figure 6 and Table S1). The best-fit model according to ModelFinder was GTR + F + R3. The resulting phylogeny showed that the monophyly of the Primulaceae clade was highly bootstrap supported (BS = 100). Within the family, *Maesa montana* A. DC. branched first to form the basal group (BS = 100). *Lysimachia* L., *Aegiceras* Gaertn., *Myrsine* L., *Embelia* Burm. f., *Elingamita* G.T.S. Baylis, *Parathesis* (A. DC.) Hook. f., *Tapeinosperma* Hook. f., and *Ardisia* Sw. showed close relationships and formed a monophyletic group (BS = 100). The genus *Androsace* formed a sister group with the *Primula* and *Bryocarpum* Hook clades. f. and Thomson (BS = 100). The genus *Primula* was divided into two clades (BS = 100). The clade containing *P. modesta* var. *koreana* and *P. modesta* var. *hannasanensis* was further divided into two clades (BS = 100). In this clade, *P. veris* branched first, followed by *P. denticulata* subsp. *sinodenticulata* and *P. pulchella*. Two groups were then formed: one comprising *P. modesta* var. *koreana* and *P. modesta* var. *hannasanensis* and the other comprising *P. stenocalyx* and *P. knuthiana* (BS = 100).

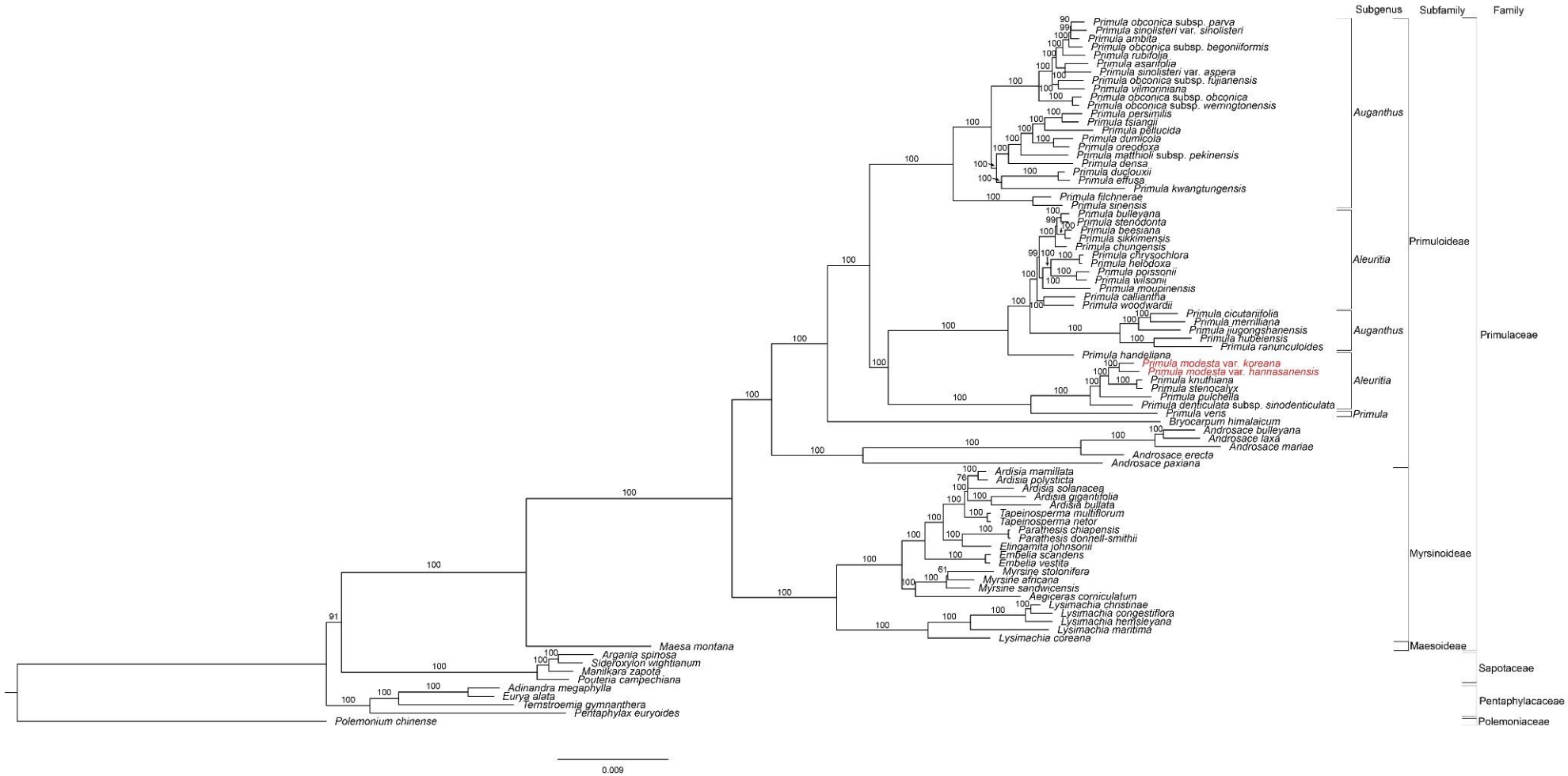

Figure 6. Maximum likelihood (ML) phylogenetic tree based on 77 protein-coding genes from 85 Ericales species. Bootstrap support values are shown at the nodes.

4. Discussion

Most *Primula* species are distributed in highlands. Some species are threatened by anthropogenic disturbances, such as collection for horticultural purposes, and the wild population is rapidly declining [3,27]. Through genome analyses, it is possible to provide basic information to evaluate the genetic diversity of species and develop identification markers. Although Yamazaki distinguished two varieties endemic to Korea (*P. modesta* var. *koreana* and *P. modesta* var. *hannasanensis*), genetic analyses of these varieties have not yet progressed. Therefore, we compared the CP genomes of the two varieties using next-generation sequencing. Herein, we reported the CP genome structures of the two Korean *Primula* varieties and presented the results of a comparative genome study.

The CP genomes of the two *Primula* varieties were well conserved and had equal gene numbers, gene orders, and typical quadripartite molecular structures. Our findings are consistent with those reported for the CP genomes of other *Primula* species [28–42]. The genome size of *P. modesta* var. *koreana* and *P. modesta* var. *hannasanesis* is 154,667 and 154,772 bp, respectively (Figure 1). It was found that *infA*, which encodes translation initiation factor 1 and aids in the assembly of the translation initiation complex, was not present in these varieties, but it has been identified in other *Primula* CP genomes [43]. We found that the varieties were located in the same clade (Figure 6). In recent studies, *ycf3* and *ycf4*—encoding the photosystem assembly factors that act in the photosystem I complex—were renamed *pafI* and *pafII*, respectively [44], and photosystem biogenesis factor 1 (*psbN*) was renamed *pbf1* [45]. Here, we used the new names. An analysis using MAFFT revealed a 99.3% sequence identity between the varieties. In addition, 536 SNPs and 501 indels were identified. Our results revealed a significant number of structural variations (SNPs and indels) in the cp genome. These results were less than the structural variations in *Commiphora gileadensis* [46]. It may be the result of subspecies taxa.

SSRs have codominant and highly polymorphic features, and they are used as markers for phylogenetic and population genetics studies [47–49]. The SSRs discovered in this study had a high A/T content; accordingly, most of the confirmed mononucleotide repeats were composed of A/T (*P. modesta* var. *koreana*: 97.7% and *P. modesta* var. *hannasanensis*: 100%). The CDS region with the highest number of SSRs was found in *ycf1*, and this is consistent with the findings of a previous study [43]. The SSRs identified here may be useful as molecular markers for studies on *Primula* species. In total, 62 long repeat sequences were discovered in both varieties and were identical to those of *Primula* species of the same order. Most repeats were in the 30–50 bp range, which was slightly smaller than previously reported ranges for *Primula* species [43]. This may be additional evidence that *Primula* species have not undergone rearrangement events [50]. Information on structural variation and SSR obtained in this study can be used to select effective molecular markers for the identification of inter- and intra-specific polymorphisms.

The CP genomes have been widely used in species verifications and phylogenetic studies on terrestrial plants [51,52]. In the present study, we performed an ML analysis to construct a phylogenetic tree. Primulaceae formed a monophyletic group that was divided into three subfamilies. Maesoideae branched first within the family, and Myrsinoideae and Primuloideae formed a monophyletic group. *Primula* formed a strong monophyletic group and was closely related to *Bryocarpum* in Primulaceae. The genus *Primula* was divided into two clades. The first clade comprised the subgenus *Auganthus*, except for species included in *Ranunculoides* (*P. cicutariifolia*, *P. merrilliana*, *P. jiugongshanensis*, *P. hubeiensis*, and *P. ranunculoides*) [1,53]. The second clade comprised the subgenus *Aleuritia*, *Primula*, and *Ranunculoides* of the subgenus *Auganthus* and was further divided into two clades. The varieties in this study, *P. modesta* var. *koreana* and *P. modesta* var. *hannasanensis*, formed a sister group with the clades *P. knuthiana* and *P. stenocalyx*. The phylogenetic tree was similar to those in previous molecular studies, but it did not follow traditional subgenus taxonomy [1,54]. Owing to insufficient information on taxa belonging to Theophrastoideae within Primulaceae, a more reliable relationship within the family could not be obtained. However, the genetic differences between the varieties were clearly identified and support

the results of previous studies [6,8]. The identification of species-level differences is essential for the ongoing conservation of vulnerable members of the genus *Primula*.

5. Conclusions

Herein, we reported the complete CP genomic sequences of *P. modesta* var. *koreana* and *P. modesta* var. *hannasanensis* for the first time. The comprehensive genetic information presented herein can form a basis for the identification of *Primula* species and the analysis of genetic differences at the individual level. Phylogenetically, *Primula* was found to be related to *Bryocarpum* belonging to Primulaceae. We also obtained important genetic information on SNPs, SSRs, long repeats, divergent hotspot regions, and phylogeny. The results of our CP genome analysis provide a valuable resource to facilitate phylogenetic, taxonomic, and evolutionary studies of this genus. Our findings will further contribute to the analysis of the genetic diversity of both species and development of identification markers.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/d14060458/s1: Table S1: NCBI data used in the ML Analysis; Table S2: Single-nucleotide polymorphisms (SNPs) identified in the chloroplast genome of the two *Primula* varieties; Table S3: SSR information of the two *Primula* varieties; Table S4: Information of long repeats of the two *Primula* varieties.

Author Contributions: Validation, H.-J.K., D.C.S. and Y.-H.H.; formal analysis, S.-C.K.; investigation, D.-K.K. and Y.-H.H.; writing—original draft preparation, S.-C.K.; writing—review and editing, H.-J.K., D.C.S. and S.-C.K.; visualization, H.-J.K.; supervision, K.C.; funding acquisition, H.-J.K. and K.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Scientific Research Grants (KNA1-1-13, 14–1) from the Korea National Arboretum.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The genome sequence data that support the findings of this study are openly available in the GenBank of NCBI at https://www.ncbi.nlm.nih.gov (Registered Date on 13 August 2021) under accession nos. MZ779112 (*P. modesta* var. *hannasanensis*) and MZ779113 (*P. modesta* var. *koreana*).

Acknowledgments: We thank Sa-Bum Jang, Hee Young Gil, and Eun-Ho Lee for sampling and laboratory assistance throughout the study.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript or in the decision to publish the results.

References

1. Xu, W.; Xia, B.; Li, X. The complete chloroplast genome sequences of five pinnate-leaved *Primula* species and phylogenetic analyses. *Sci. Rep.* **2020**, *10*, 20782. [CrossRef] [PubMed]
2. Yan, H.F.; He, C.H.; Peng, C.I.; Hu, C.M.; Hao, G. Circumscription of *Primula* subgenus *Auganthus* (Primulaceae) based on chloroplast DNA sequences. *J. Syst. Evol.* **2010**, *48*, 123–132. [CrossRef]
3. Mast, A.R.; Kelso, S.; Richards, A.J.; Lang, D.J.; Feller, D.M.; Conti, E. Phylogenetic relationships in *Primula* L. and related genera (Primulaceae) based on noncoding chloroplast DNA. *Int. J. Plant Sci.* **2001**, *162*, 1381–1400. [CrossRef]
4. Singh, S.; Ali, S.; Singh, M. Biological screening of plants extract showing hypoglycaemic and wound healing properties: *Capparis zeylanica* and *Primula denticulata*. *Am. J. Phytomed. Clin. Ther.* **2014**, *12*, 1338–1345.
5. Korea National Arboretum. *Checklist of Vascular Plants in Korea (Native Plants)*; Korea National Arboretum: Pocheon, Korea, 2020.
6. Chung, J.M.; Son, S.W.; Kim, S.Y.; Park, G.W.; Kim, S.S. Genetic diversity and geographic differentiation in the endangered *Primula farinosa* subsp. *modesta*, a subalpine endemic to Korea. *Korean J. Plant Taxon.* **2013**, *43*, 236–243. [CrossRef]
7. Korea National Arboretum. *Rare Plants Data Book in Korea*; Geobook: Seoul, Korea, 2008; p. 148. (In Korean)
8. Chung, G.Y.; Chang, K.S.; Chung, J.M.; Choi, H.J.; Paik, W.K.; Hyun, J.O. A checklist of endemic plants on the Korean Peninsula. *Korean J. Plant Taxon.* **2017**, *47*, 264–288. [CrossRef]
9. Yamazaki, T. Intraspecific taxa in *Primula farinosa* L. subsp. *modesta* (Bisset and Moore) Pax. *J. Jpn. Bot.* **2003**, *78*, 295–299. (In Japanese)

10. Jansen, R.K.; Raubeson, L.A.; Boore, J.L.; de Pamphilis, C.W.; Chumley, T.W.; Haberle, R.C.; Wyman, S.K.; Alverson, A.J.; Peery, R.; Herman, S.J.; et al. Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods Enzymol.* **2005**, *395*, 348–384.
11. Daniell, H.; Lin, C.S.; Yu, M.; Chang, W.J. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol.* **2016**, *17*, 134. [CrossRef]
12. Drouin, G.; Daoud, H.; Xia, J. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol. Phylogenet. Evol.* **2008**, *49*, 827–831. [CrossRef]
13. Jin, J.J.; Yu, W.B.; Yang, J.B.; Song, Y.; Yi, T.S.; Li, D.Z. GetOrganelle: A simple and fast pipeline for de novo assembly of a complete circular chloroplast genome using genome skimming data. *BioRxiv* **2018**, *4*, 256479.
14. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef] [PubMed]
15. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbricht-Jones, E.S.; Fischer, A.; Bock, R.; Greiner, S. GeSeq—Versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **2017**, *45*, W6–W11. [CrossRef] [PubMed]
16. Lowe, T.M.; Chan, P.P. TRNAscan-SE On-Line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* **2016**, *44*, W54–W57. [CrossRef] [PubMed]
17. Lohse, M.; Drechsel, O.; Bock, R. OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* **2007**, *52*, 267–274. [CrossRef]
18. Katoh, K.; Kuma, K.I.; Toh, H.; Miyata, T. MAFFT Version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **2005**, *33*, 511–518. [CrossRef]
19. Mayor, C.; Brudno, M.; Schwartz, J.R.; Poliakov, A.; Rubin, E.M.; Frazer, K.A.; Pachter, L.S.; Dubchak, I. VISTA: Visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* **2000**, *16*, 1046–1047. [CrossRef]
20. Amiryousefi, A.; Hyvönen, J.; Poczai, P. IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **2018**, *34*, 3030–3031. [CrossRef]
21. Librado, P.; Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **2009**, *25*, 1451–1452. [CrossRef]
22. Xia, X.; Xie, Z. DAMBE: Software package for data analysis in molecular biology and evolution. *J. Hered.* **2001**, *92*, 371–373. [CrossRef]
23. Thiel, T.; Michalek, W.; Varshney, R.K.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-Markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **2003**, *106*, 411–422. [CrossRef] [PubMed]
24. Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642. [CrossRef] [PubMed]
25. Zhang, D.; Gao, F.; Jakovlić, I.; Zou, H.; Zhang, J.; Li, W.X.; Wang, G.T. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.* **2020**, *20*, 348–355. [CrossRef] [PubMed]
26. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.; Von Haeseler, A.; Jermiin, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589. [CrossRef]
27. Richards, J. *Primula*; Timber: Portland, OR, USA, 1993.
28. Zhou, T.; Zhao, J.; Chen, C.; Meng, X.; Zhao, G. Characterization of the complete chloroplast genome sequence of *Primula veris* (Ericales: Primulaceae). *Conserv. Genet. Resour.* **2016**, *8*, 455–458. [CrossRef]
29. Sun, H.Y.; Zhong, L.; Gan, Q.L.; Zhang, T.; Wu, Z.K. The complete chloroplast genome of an endangered endemic herb species in China, *Primula filchnerae* (Primulaceae). *Mitochondrial DNA B Resour.* **2019**, *4*, 2746–2747. [CrossRef] [PubMed]
30. Zhang, C.Y.; Liu, T.J.; Yan, H.F.; Ge, X.J.; Hao, G. The complete chloroplast genome of a rare candelabra primrose *Primula stenodonta* (Primulaceae). *Conserv. Genet. Resour.* **2017**, *9*, 123–125. [CrossRef]
31. Zhang, L.; Chen, X.; Huang, Y.; Wu, Z. The complete chloroplast genome of *Primula helodoxa*, a species endemic to China. *Mitochondrial DNA B Resour.* **2020**, *5*, 194–195. [CrossRef]
32. Chen, X.; Zhang, L.; Li, W.; Huang, Y.; Wu, Z. The complete chloroplast genome of *Primula bulleyana*, a popular ornamental species. *Mitochondrial DNA B Resour.* **2019**, *4*, 3673–3674. [CrossRef]
33. Zhang, C.; Yuan, X.; Yang, T.; Yan, H.; Liu, T. The complete chloroplast genome of *Primula obconica* (Primulaceae). *Mitochondrial DNA B Resour.* **2019**, *4*, 2189–2190. [CrossRef]
34. Sun, H.Y.; Zhong, L.; Guo, Y.J.; Zhou, W.; Wu, Z.K. The complete chloroplast genome of a distylous-homostylous species, *Primula homogama* (Primulaceae). *Mitochondrial DNA B Resour.* **2021**, *6*, 393–394. [CrossRef] [PubMed]
35. Wu, Z.; Chen, X.; Zhang, L.; Huang, Y. The complete chloroplast genome of *Primula beesiana*, an ornamental alpine plant from SW China. *Mitochondrial DNA B Resour.* **2020**, *5*, 182–183. [CrossRef] [PubMed]
36. Liu, T.J.; Zhang, C.Y.; Yan, H.F.; Zhang, L.; Ge, X.J.; Hao, G. Complete plastid genome sequence of *Primula sinensis* (Primulaceae): Structure comparison, sequence variation and evidence for *accD* transfer to nucleus. *PeerJ* **2016**, *4*, e2101. [CrossRef] [PubMed]
37. Chen, S.; Yan, X.; Hao, G.; Xu, Y. The complete chloroplast genome of *Primula tsiangii* WW Smith (Primulaceae): A karst endemic primrose in Southwest China. *Mitochondrial DNA B Resour.* **2019**, *4*, 2627–2628. [CrossRef]

38. Lu, Y.; Liu, P.L.; Sun, Y.F.; Li, S.F. The complete chloroplast genome sequence of *Primula filchnerae* Knuth (Primulaceae), an endangered species in China. *Mitochondrial DNA B Resour.* **2020**, *5*, 2047–2048. [CrossRef]
39. Zhang, C.Y.; Liu, T.J.; Yan, H.F.; Xu, Y. The complete chloroplast genome of *Primula persimilis* (Primulaceae). *Conserv. Genet. Resour.* **2017**, *9*, 189–191. [CrossRef]
40. Zhang, C.Y.; Liu, T.J.; Xu, Y.; Yan, H.F.; Hao, G.; Ge, X.J. Characterization of the whole chloroplast genome of an endangered species *Primula kwangtungensis* (Primulaceae). *Conserv. Genet. Resour.* **2017**, *9*, 87–89. [CrossRef]
41. Wang, J.; Zhang, R.; Ren, T.; Han, K.; Zeng, S.; Biffin, E.; Liu, Z.L. Characterization of the complete chloroplast genome of the *Cortusa matthioli* subsp. *pekinensis* (Primulaceae). *Conserv. Genet. Resour.* **2017**, *9*, 603–605. [CrossRef]
42. Zhang, C.Y.; Liu, T.J.; Xu, Y.; Yan, H.F. Characterization of the whole chloroplast genome of a rare candelabra primrose *Primula chrysochlora* (Primulaceae). *Conserv. Genet. Resour.* **2017**, *9*, 361–363. [CrossRef]
43. Ren, T.; Yang, Y.; Zhou, T.; Liu, Z.L. Comparative plastid genomes of *Primula* species: Sequence divergence and phylogenetic relationships. *Int. J. Mol. Sci.* **2018**, *19*, 1050. [CrossRef]
44. 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.* **2011**, *76*, 273–297. [CrossRef] [PubMed]
45. Krech, K.; Fu, H.Y.; Thiele, W.; Ruf, S.; Schöttler, M.A.; Bock, R. Reverse genetics in complex multigene operons by co-transformation of the plastid genome and its application to the open reading frame previously designated *psbN*. *Plant J.* **2013**, *75*, 1062–1074. [CrossRef] [PubMed]
46. Khan, A.; Asaf, S.; Khan, A.L.; Al-Harrasi, A.; Al-Sudairy, O.; AbdulKareem, N.M.; Khan, A.; Shehzad, T.; Alsaady, N.; Al-Lawati, A.; et al. First complete chloroplast genomics and comparative phylogenetic analysis of *Commiphora gileadensis* and *C. foliacea*: Myrrh producing trees. *PLoS ONE* **2019**, *14*, e0208511. [CrossRef] [PubMed]
47. Grassi, F.; Labra, M.; Scienza, A.; Imazio, S. Chloroplast SSR markers to assess DNA diversity in wild and cultivated grapevines. *Vitis* **2002**, *41*, 157–158.
48. He, S.L.; Wang, Y.S.; Volis, S.; Li, D.Z.; Yi, T.S. Genetic diversity and population structure: Implications for conservation of wild soybean (*Glycine soja* Sieb. et Zucc) based on nuclear and chloroplast microsatellite variation. *Int. J. Mol. Sci.* **2012**, *13*, 12608–12628. [CrossRef]
49. Xue, J.; Wang, S.; Zhou, S.L. Polymorphic chloroplast microsatellite loci in *Nelumbo* (Nelumbonaceae). *Am. J. Bot.* **2012**, *99*, e240–e244. [CrossRef]
50. Kim, S.C.; Lee, J.W.; Choi, B.K. Seven complete chloroplast genomes from *Symplocos*: Genome organization and comparative analysis. *Forests* **2021**, *12*, 608. [CrossRef]
51. Särkinen, T.; George, M. Predicting plastid marker variation: Can complete plastid genomes from closely related species help? *PLoS ONE* **2013**, *8*, e82266. [CrossRef]
52. Ma, P.F.; Zhang, Y.X.; Zeng, C.X.; Guo, Z.H.; Li, D.Z. Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (Poaceae). *Syst. Biol.* **2014**, *63*, 933–950. [CrossRef]
53. Shao, J.W.; Wu, Y.F.; Kan, X.Z.; Liang, T.J.; Zhang, X.P. Reappraisal of *Primula ranunculoides* (Primulaceae), an endangered species endemic to China, based on morphological, molecular genetic and reproductive characters. *Bot. J. Linn. Soc.* **2012**, *169*, 338–349. [CrossRef]
54. Gang, H.A.O.; Chi-Ming, H.U.; Nam-Sook, L.E.E. Circumscriptions and phylogenetic relationships of *Primula* Sects. *Auganthus* and *Ranunculoides*: Evidence from nrDNA ITS sequences. *J. Integr. Plant Biol.* **2002**, *44*, 72.