Peer-reviewed version available at Molecules 2017, 22, , 982; doi:10.3390/molecules22060982

Article

The Complete Chloroplast Genome Sequences of *Fritillaria ussuriensis* Maxim. and *Fritillaria cirrhosa* D. Don, and Comparative Analysis with Other *Fritillaria* Species

Inkyu Park[1], Wook-Jin Kim[1], Sang-Min Yeo[1], Goya Choi[1], Young-Min Kang[1], Renzhe Piao[2] and Byeong-Cheol Moon[1]*

[1] K-herb Research Center, Korea Institute of Oriental Medicine, Daejeon 305-811, Republic of Korea; pik6885@kiom.re.kr (I.K.P.); ukgene@kiom.re.kr (W.J.K); yeosm1115@kiom.re.kr (S.M.Y); serparas@kiom.re.kr (G.C.); ymkang@kiom.re.kr (Y.M.K)

[2] Department of Agronomy, Yanbian University Agriculture College, Yanji 133002, China ; rzpiao@ybu.edu.cn (R.P.)

* Correspondence: e-mail@e-mail.com; Tel.: +82-42-868-9530

Abstract: The genus *Fritillaria* belongs to the widely distributed family Liliaceae. The bulbs of *Fritillaria ussuriensis* and *Fritillaria cirrhosa* are valuable herbaceous medicinal ingredients. However, they are still used indiscriminately in herbal medicine. Identification and molecular phylogenic analysis of *Fritillaria* species is therefore required. Here, we report the complete chloroplast (cp) genome sequences of *F. ussuriensis* and *F. cirrhosa*. The two *Fritillaria* cp genomes were 151,524 and 151,083 bp in length, respectively, including a pair of inverted repeat regions (52,678 and 52,156 bp) separated by a large single copy region (81,732 and 81,390 bp) and small single copy region (17,114 and 17,537 bp). A total of 111 genes in *F. ussuriensis* and 112 in *F. cirrhosa* comprised 77 protein-coding genes in *F. ussuriensis* and 78 in *F. cirrhosa*, 30 tRNA genes, and four rRNA genes. The gene order, content, and orientation of the two *Fritillaria* cp genomes exhibited the general structure of flowering plants, and were similar to those of other *Fritillaria* species. Comparison of the six *Fritillaria* species' cp genomes indicated seven highly divergent regions in intergenic spacers and in the *matK*, *rpoC1*, *rpoC2*, *ycf1*, *ycf2*, *ndhD*, and *ndhF* coding regions. We established the position of the six species through phylogenic analysis. The complete chloroplast genome sequences of two *Fritillaria* species will be useful genomics resources for identification of *Fritillaria* species and for studying the phylogenetic relationship among *Fritillaria* species within the Liliaceae family.

Keywords: *Fritillaria ussuriensis*, *Fritillaria cirrhosa*, Chloroplast genome, Comparative analysis, Highly divergent region

1. Introduction

The genus *Fritillaria* belongs to the family Liliaceae, which consists of 15 genera and about 705 known species, and is most closely related to the genus *Lilium* [1]. The bulb of *Fritillaria*, called *'Pae-mo'* in Korean and '*Bei-mu*' in Chinese, is an important ingredient of herbal drugs in oriental medicine, and has great economic value in Asian countries. The dried bulbs of *Fritillaria ussuriensis* Maxim. and *Fritillaria cirrhosa* D. Don are used in different herbal medicines, namely, Fritillariae Ussuriensis Bulbus (*Ping-bei-mu* in Chinese) and Fritillariae Cirrhosae Bulbus (*Chuan-bei-mu* in Chinese), respectively. The bulbs of *Fritillaria* have biologically active compound steroidal alkaloids including peimine, peiminine, fritilline, fritillarin, and verticine [2,3]. They have been widely used to eliminate phlegm called 'Beimu' [4,5]. They also have pharmacological effects in the treatment of tonsillitis,

bronchitis, fever, and high blood pressure [2,3]. Although the bulbs of *Fritillaria* have value in herbal medicine, different *Fritillaria* species are still used indiscriminately because of their morphological similarity and similar names. Therefore, accurate identification of *Fritillaria* species, for example, using molecular markers, is required to identify medicinal plants or drugs derived from them.

Chloroplasts play an important role in photosynthesis and carbon fixation as well as in the biosynthesis of starch, fatty acids, and amino acids [6]. The cp genome ranges from 120 to 180 kb in higher plants and has a quadripartite structure consisting of a large single copy (LSC) and small single copy (SSC) region and two copies of a larger inverted repeat (IR) [7]. The cp genome encodes 110 to 130 genes with up to 80 unique protein-coding genes, four ribosomal RNAs (rRNAs), and approximately 30 transfer RNAs (tRNAs) [8,9]. However, a few parasitic plants have small chloroplast genomes due to unique life cycles [10]. Since the cp genome of *Marchantia polymorpha* [11] was reported in 1986, more than 500 complete chloroplast genome sequences have been deposited in the GenBank database (http://www.ncbi.nlm.nih.gov/genbank/). With the advancement of next-generation sequencing (NGS) technology, chloroplast genome assembly has become cheaper and easier than with the Sanger method [12]. Through comparison of chloroplast genomes, development of molecular markers has also become more cost effective. The cp genome has been widely used for understanding phylogenetic relationships and discovering useful molecular markers, which are used in DNA barcoding to identify plant species and for the authentication and identification of herbal medicines. In particular, *matK* and *rbcL* are used as universal plant DNA barcodes [13].

Here, we report *de novo* assembly of *F. ussuriensis* and *F. cirrhosa* cp genomes using the Illumina platform. This is the first comparative analysis for *Fritillaria* cp genomes in conjunction with previously reported cp genomes. This study aimed to investigate global structural patterns of six *Fritillaria* cp genomes and to discover highly divergent regions among the species. We also analyzed simple sequence repeats (SSRs) and other repeats in order to develop molecular markers and to reconstruct phylogenic relationships among the six *Fritillaria* species. The results will provide genomic resources for the identification of *Fritillaria* species using DNA barcodes or molecular markers, and will enhance understanding of the evolution of *Fritillaria* species within the family Liliaceae.

3. Results

3.1. Chloroplast genome organization of two Fritillaria species

The Illumina sequencing generated 5.0 and 4.5 Gb of trimmed paired-end reads from *F. ussuriensis* and *F. cirrhosa*, respectively. From *de novo* assembly sequence using low-coverage whole-genome sequencing (WGS), we obtained ten and eight contigs covering the whole chloroplast genome sequence for *F. ussuriensis* and *F. cirrhosa*, respectively (Table S1). Single circular sequences were completed after gap filling and manual editing. The complete circular chloroplast genome of *F. ussuriensis* and *F. cirrhosa* was 151,524 and 151,083 bp with approximately 256× and 452× coverage, respectively (Table S2). The paired-end read mapping was conducted to validate the draft genome, which was compared to our draft genomes and previously reported F. hupehensis genome using BLASTZ analysis (Figure S1). Both *F. ussuriensis* and *F. cirrhosa* chloroplast genomes had a quadripartite structure similar to most land plants consisting of a pair of IRs (52,678 and 52,156 bp, respectively) separated by LSC (81,732 and 81,390 bp) and SSC (17,114 and 17,537 bp) regions (Figure 1, Table 1). The *Fritillaria* chloroplast genomes were AT-rich (63% in both species), and LSC (65.3% and 64.2% in *F. ussuriensis* and *F. cirrhosa*) and SSC (69.4% and 69.6% in *F. ussuriensis* and *F. cirrhosa*) regions were more AT-rich than the IR regions (57.6% and 57.4% in *F. ussuriensis* and *F. cirrhosa*), which is similar to other chloroplast genomes [6,7,14-16].

The gene content, order, and orientation were similar in the *Fritillaria* cp genomes. There were 111 and 112 predicted genes in *F. ussuriensis* and *F. cirrhosa*, respectively. Of these, 94 in *F. ussuriensis*

and 95 in *F. cirrhosa* were unique to the LSC and SSC regions and 18 were duplicated in the IR regions (Table 1, Table 2). The 111 and 112 unique genes consisted of 77 and 78 protein-coding genes in *F. ussuriensis* and *F. cirrhosa*, respectively, and 30 tRNAs, with 17 duplicated genes including seven tRNAs (*trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, and *trnV-GAC*), four rRNAs (*rrn16*, *rrn23*, *rrn4.5*, and *rrn5*), and six protein-coding genes (*ndhB*, *rpl2*, *rpl23*, *Rps12*, *ycf1*, and *ycf2*) in both chloroplast genomes. The two *Fritillaria* chloroplast genomes contained 18 intron-containing genes. Among them, 15 genes (nine protein-coding genes and six tRNA genes) had a single intron and two genes (*ycf3* and *clpP*) had two introns (Table S4). Thirteen genes (nine protein-coding and four tRNA genes) were located in the LSC region, one protein-coding gene in the SSC region, and four genes (two protein-coding and two tRNA genes) in the IR regions. The protein-coding genes included five genes (*ndhB*, *rpl2*, *rpl23*, *rps12*, and *ycf2*) that were duplicated in the IR regions. The *rps12* gene was trans-spliced because the 5′ end was located in the LSC region and the 3′ end in the IR region. The *trnK-UUU* gene had the largest intron region (2,613 bp in F. ussuriensis and 2,562 bp in F. cirrhosa) including makK. The genes *psbT*, *rpl2*, and *ndhD* had the alternative start codon ACG, and *rps19* started with GTG. Use of ACG and GTG as start codons is common for a variety of genes in the chloroplast genomes of land plants [17-20]. Finally, the *cemA* gene was identified as a pseudogene with a premature stop codon in the *F. cirrhosa* cp genome.

Approximately 52% of the *Fritillaria* chloroplast genomes consisted of protein-coding genes (78,951 bp in *F. ussuriensis* and 79,835 bp in *F. cirrhosa*), 1.9% of tRNAs (2,876 bp in both species), and 6.0% of rRNAs (9,048 bp in both species). The remaining 40.1% consisted of intergenic regions, non-coding introns, and pseudogenes. The *ycf1* gene located between IRb and the SSC region had premature stop codons in the coding sequence, and has been annotated as a pseudogene in other angiosperm chloroplast genomes [6,21]. The codon usage and anticodon recognition patterns of the cp genomes are summarized in Table S5. Protein-coding genes comprised 26,317 codons in F. ussuriensis and 26,611 in *F. cirrhosa*. Among these codons, leucine and isoleucine were the most common amino acids in both genomes (Figure 2). The 30 tRNA genes included codons for all 20 amino acids required for biosynthesis. Within protein-coding regions (CDSs), the AT content for the first, second, and third codons, respectively, was 55.1%, 62.1%, and 70.4% in F. ussuriensis and 59.3%, 63.1%, and 65.6% in *F. cirrhosa* (Table S3). The bias towards a higher AT content in the third position has been observed in other land plant cp genomes [6,16].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2017 doi:10.20944/preprints201703.0139.v1

Peer-reviewed version available at Molecules 2017, 22, , 982; doi:10.3390/molecules22060982

Table 1. Size comparison of two *Fritillaria* species' chloroplast genomic regions

Species	*Fritillaria ussuriensis*	*Fritillaria cirrhosa*
Total cp genome size (bp)	151,524	151,083
Large single copy (LSC) region (bp)	81,732	81,390
Inverted repeat (IR) region (bp)	52,678	52,156
Small single copy (SSC) region (bp)	17,114	17,537
GC content (%)	36.95	36.96
LSC (%)	34.71	34.79
IR (%)	42.40	42.60
SSC (%)	30.63	30.50
Total number of genes	111	112
Protein coding gene	77	78
rRNA	4	4
tRNA	30	30

Table 2. **Genes present in the two *Fritillaria* chloroplast genomes**

Gene products of two *Fritillaria* species	
Photosystem I	psaA, B, C, I, J
Photosystem II	psbA, B, C, D, E, F, H, I, J, K, L, M, N, T, Z
Cytochrome b6/f	petA, B[1)], D[1)], G, L, N
ATP synthase	atpA, B, E, F[1)], H, I
Rubisco	rbcL
NADH oxidoreductase	ndhA[1)], B[1) 3)], C, D, E, F, G, H, I, J, K
Large subunit ribosomal proteins	rpl2[1) 3)], 14, 16[1)], 20, 22, 23[3)], 32, 33, 36
Small subunit ribosomal proteins	rps2, 3, 4, 73), 8, 11, 12 [2) 3) 4)], 14, 15, 16, 18, 19
RNA polymerase	rpoA, B, C1[1)], C2
Unknown function protein coding gene	ycf1[3)], 2[3)], 3[2)], 4
Other genes	accD, ccsA, cemA[5)], clpP[2)], matK
Ribosomal RNAs	rrn16[3)], 23[3)], 4.5[3)], 5[3)]
Transfer RNAs	trnA-UGC[1)3)], trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-UCC[1)], trnG-GCC, trnH-GUG, trnI-CAU[3)], trnI-GAU[1)3)] trnK-UUU1), trnL-UAA1), trnL-UAG, trnL-CAA[3)], trnM-CAU, trnfM-CAU, trnN-GUU[3)], trnP-UGG, trnQ-UUG, trnR-ACG[3)], trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-UAC[1)], trnV-GAC[3)], trnW-CCA, trnY-GUA

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2017 doi:10.20944/preprints201703.0139.v1

Peer-reviewed version available at *Molecules* **2017**, *22*, , 982; doi:10.3390/molecules22060982

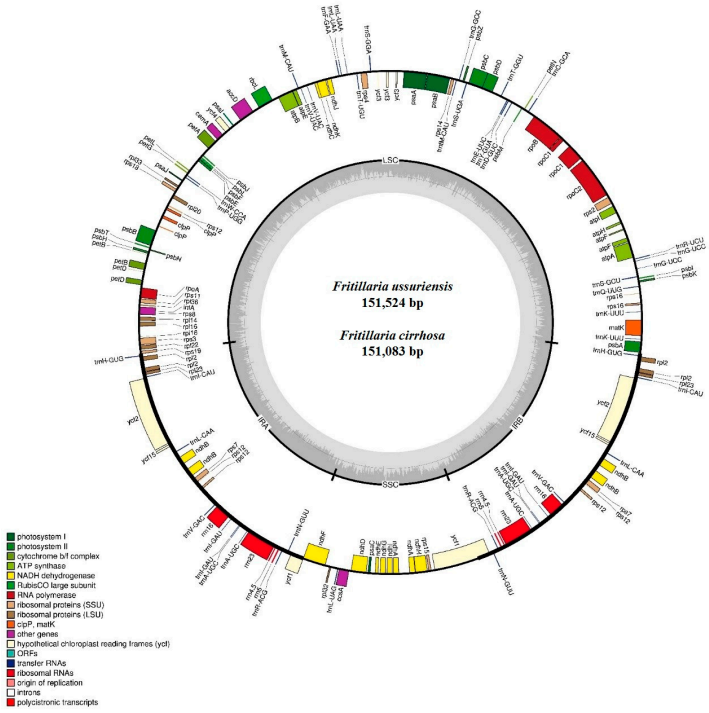

Figure 1. Circular gene map of two *Fritillaria* species' chloroplast genomes. Genes drawn inside the circle are transcribed clockwise, and those outside the circle are transcribed counterclockwise. The darker gray in the inner circle corresponds to GC content

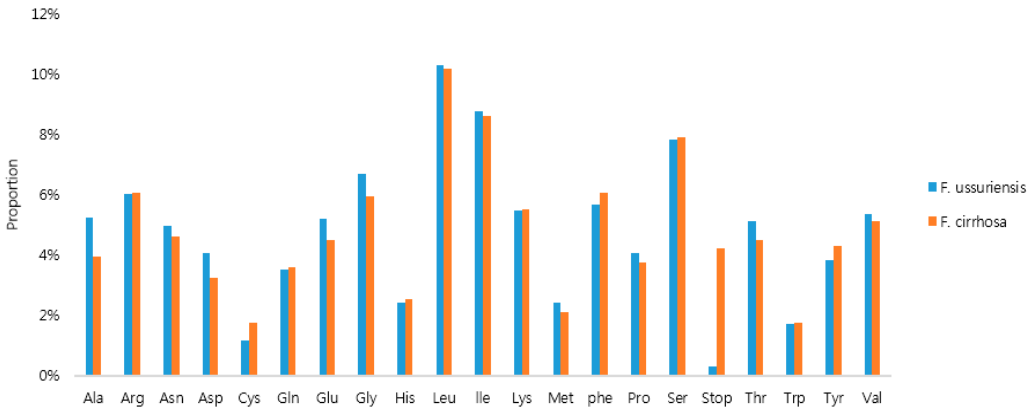

Figure 2. The amino acid frequencies in *F. ussuriensis* and *F. cirrhosa* protein-coding sequences.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2017 doi:10.20944/preprints201703.0139.v1

Peer-reviewed version available at Molecules 2017, 22, , 982; doi:10.3390/molecules22060982

3.2. Repeat analysis in two Fritillaria chloroplast genomes

SSRs or microsatellites are tandem repeat sequences consisting of 1–6 nt sequence motifs in prokaryotic and eukaryotic genomes [22,23]. We detected 183 and 178 SSRs in *F. ussuriensis* and *F. cirrhosa* cp genomes, respectively. Mononucleotide motifs were the most abundant type of repeat and di-nucleotides the second most abundant in both *Fritillaria* genomes (Figure 3). Almost all SSR loci were composed of A or T. Thus, *Fritillaria* species have AT-rich cp genomes. SSRs are more abundant in non-coding regions than in coding regions, explaining 63% of all SSRs in both genomes. Furthermore, most SSRs were located in the LSC region. We also identified 15 tandem repeats in F. ussuriensis and 13 tandem repeats in *F. cirrhosa* of more than 20 bp (Table S6). Of these, most were located in IGS, LSC, and IR regions. The longest tandem repeats were 108 bp in F. ussuriensis (located in the *trnT-UGU/trnL-UAA* IGS) and 94 bp in F. cirrhosa (located in the *trnG-UCC/trnR-UCU* IGS). Four (two in IGS, two in CDS) tandem repeats represented the same region in both cp genomes. Four and six palindromic repeats were also detected in F. ussuriensis and F. cirrhosa, respectively (Table S7). In *F. ussuriensis*, two of these were located in the LSC region and two in the IR regions. Both species had palindromic repeats at four locations, namely, the IGS of *accD/psaI*, *petD/rpoA*, *ccsA/ndhD*, and *rps15/ycf1* regions.

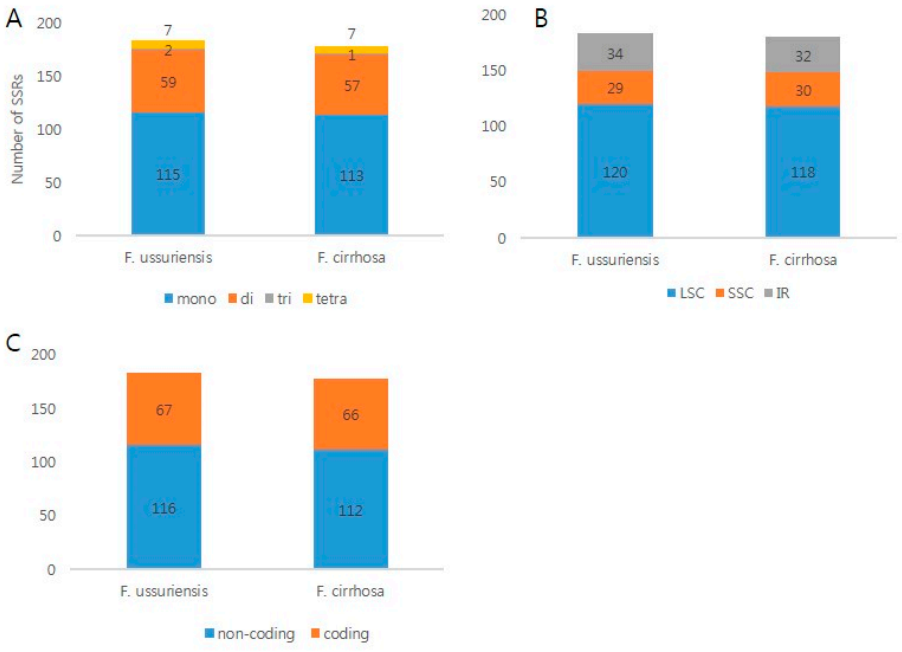

Figure 3. Distribution of SSRs in the *F. ussuriensis* and *F. cirrhosa* cp genomes. (A) SSR type distribution in the two *Fritillaria* cp genomes. (B) The proportion of SSRs in different genomic regions of *Fritillaria* cp genomes. (C) SSR distribution between coding and non-coding regions.

3.3. Comparison of chloroplast genomes with those of other Fritillaria species

The two *Fritillaria* chloroplast genomes had approximately 98% sequence identity and contained similar genes. The genome of F. ussuriensis was approximately 441 bp longer than that of *F. cirrhosa* (Table 1). The LSC and IR regions of *F. ussuriensis* were 342 and 522 bp longer, respectively, than those of *F. cirrhosa*. The SSC region of *F. ussuriensis* was 423 bp shorter than that of *F. cirrhosa*. IR contraction and expansion are common evolutionary events and contribute to genome size variation [6,7,24]. We analyzed the border structure of the two cp genomes. Detailed comparison of the LSC,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2017 doi:10.20944/preprints201703.0139.v1

Peer-reviewed version available at Molecules 2017, 22, , 982; doi:10.3390/molecules22060982

SSC, and IR regions is shown in Figure 4. The rps19 gene located in the LSC region extended into the IRb region by 27–46 bp. The border between IRb/SSC and SSC/IRa extended into the ycf1 genes in all Fritillaria species except *F. taipaiensis*. Overlaps of 17–33 bp were observed between the ycf1 pseudogene and *ndhF* gene, except in *F. taipaiensis*. The *trnH* genes were all located in the LSC region, 149–174 bp away from the IRa/LSC boundary. The locations of most other genes (e.g., *ndhF* and *trnH*) were similar in both cp genomes (Figure 4).

We performed multiple sequence alignment between six *Fritillaria* chloroplast genomes using mVISTA by sequence identity (Figure 5). The non-coding regions were more divergent than the coding regions. The most divergent regions were found in IGSs such as *matK*/*trnK-UUU*, *trnK-UUU*/*rps16*, *rps16*/*trnQ-UUG*, *psbK*/*psbI*, *atpH*/*atpI*, *psbM*/*trnD-GUC*, and *ycf4*/*petD*. For the coding regions, the most divergent regions included *matK*, *rpoC1*, *rpoC2*, *ycf1*, *ycf2*, *ndhD*, and *ndhF*. Previous studies reported similar divergent regions [6]. These regions are conserved regions of clustered variation called hotspots, containing single-nucleotide polymorphisms (SNPs) and indels [25,26]. The nucleotide variability (Pi) was calculated to show divergence at the sequence level of Fritillaria cp genomes. As expected, the IR regions were more conserved than the LSC and SSC regions. Between *F. ussuriensis* and *F. cirrhosa* cp genomes, Pi values ranged from 0 to 0.158 with a mean of 0.009. Ten highly divergent loci included *matK*, *atpI*, *trnY-GUA*, *trnE-UUC*, *trnT*, *ycf3*, *rps4*, *ycf4*, *petA*, *rpl16*, *rps3*, *rps19*, *ccsA*, *ndhD*, and *ycf1* (Figure 6). All of these loci had a much higher divergence value than other regions (Pi >0.027). Among the six Fritillaria cp genomes, the Pi values varied from 0 to 0.021 with a mean of 0.004 (Figure 6). The nine loci including *matK*, *atpI*, *trnY*, *trnE*, *trnT*, *rps19*, *ccsA*, *ndhD*, and *ycf1* were highly divergent among *Fritillaria* species.

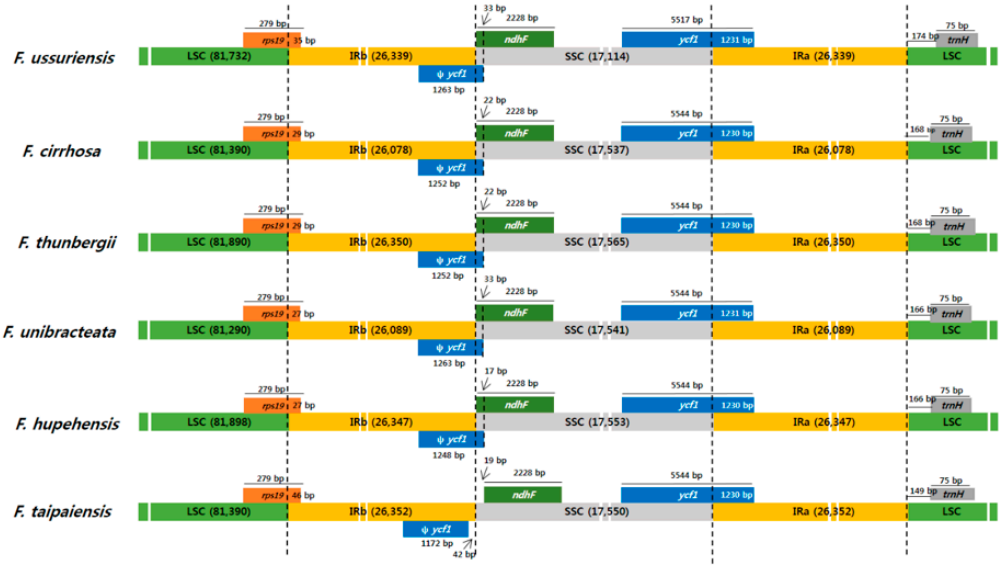

Figure 4. Comparison of LSC, SSC, and IR border regions among the six Fritillaria species' chloroplast genomes. Colored boxes for genes represent the gene position.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2017 doi:10.20944/preprints201703.0139.v1

Peer-reviewed version available at Molecules 2017, 22, , 982; doi:10.3390/molecules22060982

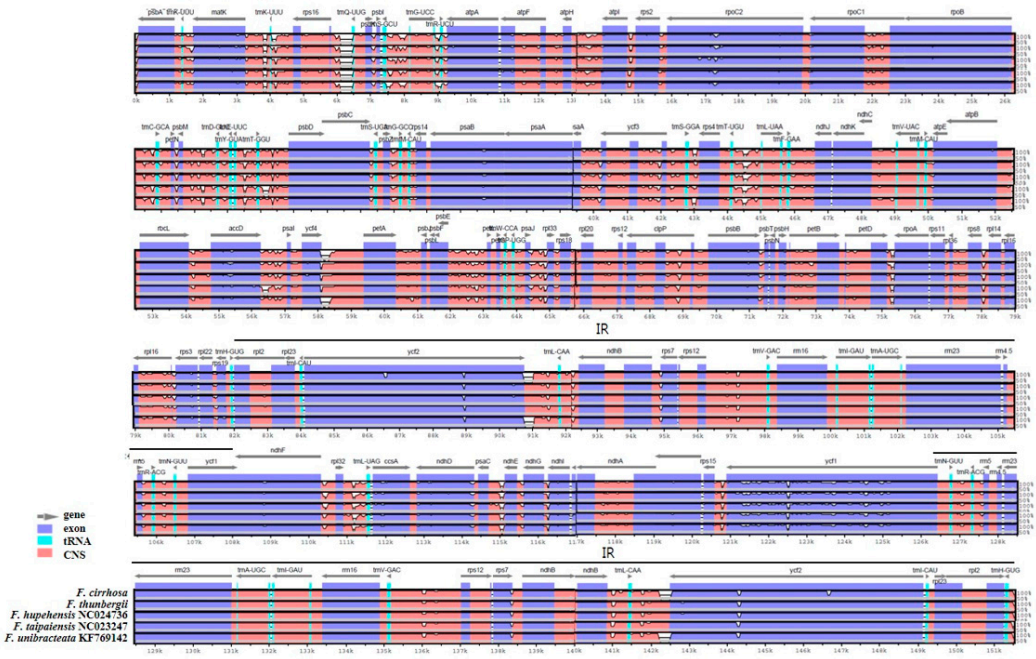

Figure 5. Comparison of six Fritillaria chloroplast genomes using mVISTA. Complete cp genomes of six Fritillaria species were used for comparison within published cp genomes. Blue block, conserved gene; sky-blue block, tRNA and rRNA; red block, intergenic region. White peaks indicate regions with sequence variation among Fritillaria species.

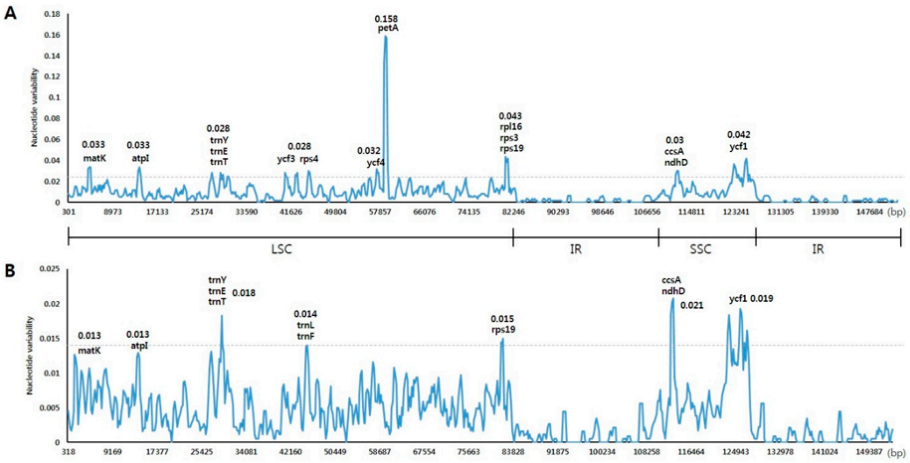

Figure 6. Sliding window analysis of the cp genomes. (A) Comparison of the nucleotide variability (Pi) between F. ussuriensis and F. cirrhosa. (B) Comparison of the nucleotide variability (Pi) among six Fritillaria species' cp genomes.

3.4. Phylogenic analysis

Chloroplast genomes have been successfully used in numerous phylogenetic studies of angiosperms [27,28]. To identify the phylogenetic position of the six *Fritillaria* species within the Liliaceae family, 74 protein-coding sequences shared by 11 cp genomes were aligned (Figure 7). Two species, *Acorus gramineus* and *Cocos nucifera*, were set as outgroups. The alignment covered 80,532 bp. ML and MP analysis revealed six out of eight nodes with 100% bootstrap values. Both the ML and MP phylogenetic results strongly indicated that the six *Fritillaria* species were closely related within

Liliaceae. *F. cirrhosa* and *F. unibracteata* formed a cluster and then a monophyletic clade with *F. taipaiensis*. *F. hupehensis* and *F. thunbergii* formed a cluster and were related to *F. ussuriensis* as a monophyletic branch.

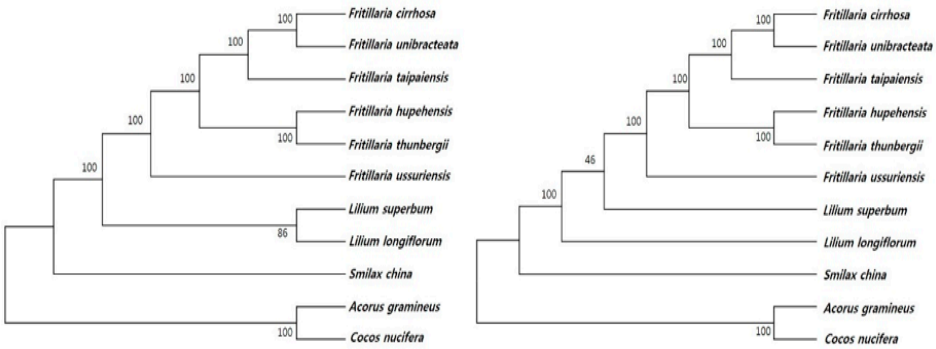

Figure 7. Phylogenetic trees constructed using 74 protein-coding genes of 11 species with ML and MP. Numbers above are bootstrap support values (>50%).

4. Discussion

Advances in NGS technologies make it possible to complete entire chloroplast genomes with the discovery of molecular markers [29]. We used low-coverage WGS data to obtain chloroplast genomes, an approach that has been successfully used in several studies [30,31]. This approach requires less time and has a lower cost than the previously used method. Here, we obtained two *Fritillaria* chloroplast genomes and applied comparative analysis to six *Fritillaria* species cp genomes. The two *Fritillaria* cp genomes contained a pair of IRs, and LSC and SSC regions. The two *Fritillaria* genomes have similar genome structure, gene order, and gene contents, including introns and base compositions. They show characteristics typical of land plant cp genomes [6,32].

SSRs are suitable molecular markers because they are distributed throughout the whole genome and display high polymorphism between species, locus-specific co-dominance, and high transferability [33]. SSRs play an important role in cp genome rearrangement during evolution [21,34]. A and T repeat units occupy the highest portion of SSRs, which therefore contribute to the AT-richness of the cp genome [35-37]. In this study, we identified 183 and 178 SSR loci in *F. ussuriensis* and *F. cirrhosa* cp genomes, respectively. Most of the SSRs were located in the LSC and non-coding regions. In general, SSRs of *Fritillaria* chloroplast genomes represented abundant variation. These will be helpful genomic resources for discovery of markers as well as for population genetic studies of *Fritillaria* species.

The IR regions are the most conserved regions in the chloroplast genome [6]. However, the contraction and expansion at the borders of IR regions are common evolutionary events, and are major causes of rearrangements and size variation [7,15,38]. In this study, we compared the SC/IR boundaries among six *Fritillaria* genomes. The SC/IR boundaries showed only slight differences, for example, at *ycf1* and *ψycf1* (Figure 4). This phenomenon is relatively common in other cp genomes [8,16,21]. Several genes, including *ndhF*, *rps19*, and *trnH*, had almost identical locations and sizes among *Fritillaria*. Multiple sequence alignment between six *Fritillaria* genomes indicated that the IR regions were more conserved than the LSC and SSC regions due to copy correction by gene conversion in IR regions [39]. The most divergent regions were found in IGSs, which have been used in phylogenetic studies [9,28]. For the coding regions, the most divergent regions included *matK*, *rpoC1*, *rpoC2*, *ycf1*, *ycf2*, *ndhD*, and *ndhF*. Previous studies reported similar divergent regions [6]. These

regions are conserved regions of clustered variation called hotspots, containing SNPs and indels [25,26]. The *rpoC2*, *rpoC1*, and *ycf1* genes are known as hotspots for variation[9,25,26,40]. Therefore, *Fritillaria* cp genomes contain general hotspot regions for genetic variation as in other land plants. Chloroplast genomes provide genomic resources for phylogenetic analysis, and many studies have used protein-coding sequences or whole chloroplast genome sequences [6,28,41]. We used ML and MP analysis to construct a phylogenetic tree focusing on *Fritillaria* species. Six *Fritillaria* species was clustered. Both ML and MP analysis of the position of *Fritillaria* species within the Liliaceae family corresponded with previous studies [42,43]. Thus, our results established the relationship among six *Fritillaria* species as well as their position within the Liliaceae family. This result will be helpful to the evolutionary relationship among *Fritillaria* species.

2. Materials and Methods

2.1. Genome sequencing and assembly

Fresh leaves of *F. ussuriensis* (KY646166) and *F. cirrhosa* (KY646167) were collected from medicinal plant plantations, and the samples were used for cp genome sequencing. *F. ussuriensis* and *F. cirrhosa* were given identification numbers, and specimens were registered in the Korean Herbarium of Standard Herbal Resources (Index Herbariorum code KIOM) at the Korea Institute of Oriental Medicine (KIOM). DNA was extracted using DNeasy Plant Maxi kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. Illumina short-insert paired-end sequencing libraries were constructed and generated using the NextSeq platform (Illumina, San Diego, Valencia, CA, USA) by LabGenomics, Korea. The cp genomes were obtained by *de novo* assembly from low-coverage whole-genome sequence derived from the Phyzen pipeline (http://phyzen.com). Trimmed paired-end reads (Phred scores ≥20) were assembled using CLC genome assembler (ver. 4.06 beta, CLC Inc, Rarhus, Denmark) with default parameters. SOAPdenovo gap closer was performed to fill gaps based on alignment paired-end reads [44]. The principal contigs representing the cp genome were retrieved from the total contigs using Nucmer [45], and aligned contigs were ordered with the cp genome sequence of *Fritillaria hupehensis* (NC024736) as reference [43].

2.2. Genome annotation and comparative analysis

Gene annotation of the *F. ussuriensis* and *F. cirrhosa* cp genomes was performed using DOGMA annotation [46], and manually corrected for codons and gene boundaries using BLAST searches. The tRNAs were confirmed with tRNAscan-SE 1.21 [47]. The circular maps of the two *Fritillaria* cp genomes were obtained using OGDRAW [48]. GC content and codons were analyzed using the MEGA6 software [49]. The mVISTA program was used to compare the seven *Fritillaria* cp genomes using the *F. ussuriensis* cp genome as reference [50]. Five *Fritillaria* cp genomes were downloaded from GenBank (*F. hupenesis*: NC024736, *F. tapaiensis*: NC023247, *F. unibracteata*: KF769142, and *F. thunbergii*: KY646165).

2.3. Repeat analysis

SSRs in *F. ussuriensis* and *F. cirrhosa* cp genomes were detected using MISA (http://pgrc.ipk-gatersleben.de/misa/misa) with the parameters set to minimum number of repeats to 10, 5, 4, 3, 3, and 3 for mono-, di-, tri-, tetra-, penta- and hexa-nucleotides, respectively. The tandem repeats were 20 bp or more with minimum alignment score and maximum period size set at 50 and 500, respectively, and the identity of repeats was set to ≥90% [51]. IRs were detected using Inverted Repeat Finder with default parameters. The IRs were 20 bp or more with 90% similarity [52].

2.4. Phylogenic and divergence analysis

A molecular phylogenetic tree was constructed using 74 protein-coding genes from 11 species. Among these 11 taxa, 9 completed cp genomes were downloaded from NCBI: *Fritillaria unibracteata* (KF769142), *Fritillaria taipaiensis* (NC023247), *Fritillaria hupehensis* (NC024736), *Fritillaria thunbergii* (KY646165), *Lilium superbum* (NC026787), *Lilium longiflorum* (KC968977), *Smilax china* (HM536959), *Acorus gramineus* (NC026299), and *Cocos nucifera* (KX028884). A total of 71 protein-coding genes were aligned with MAFFT (http://mafft.cbrc.jp/alignment/server/). Maximum likelihood (ML) and maximum parsimony (MP) analysis was performed using MEGA6 with 1,000 bootstrap replicates [49]. Six *Fritillaria* species cp genomes were aligned using MAFFT, and the sequences were manually adjusted using Bioedit [53]. To calculate nucleotide variability (Pi) between cp genomes, sliding window analysis was performed using DnaSP version 5.1 software [54]. Window length was set to 600 bp, and the step size was 200 bp.

Supplementary Materials: The following are available online.

References

1. Christrnhusz, M.J.M.; Byng, J.W. The number of known plants species in the world and its annual increase. *2016* **2016**, *261*, 17.
2. Park, C.H.; Ryu, J.H.; Han, K.S.; Doo, H.S.; Choi, S.Y. Effect of plant growth regulators on bulbiet formation and plant regeneration in fritillaria thunberg miq. *Korean journal of Medicinal Crop Sci* **1996**, *4*, 119-125.
3. Xue, Y.; Wang, F. [simultaneous determination of peimine, peiminine and zhebeinine in fritillaria thunbergii from different habitat by hplc-elsd]. *Zhongguo Zhong yao za zhi = Zhongguo zhongyao zazhi = China journal of Chinese materia medica* **2007**, *32*, 1628-1630.
4. Lin, G.; Li, P.; Li, S.L.; Chan, S.W. Chromatographic analysis of fritillaria isosteroidal alkaloids, the active ingredients of beimu, the antitussive traditional chinese medicinal herb. *Journal of chromatography. A* **2001**, *935*, 321-338.
5. Paek, K.Y.; Yu, K.J.; Seong, N.S.; Cho, J.T. Micropropagation through stem, node-bud shoot tip and bulblet scale culture in fritillaria thunbergii miq. . *Korean journal of Medicinal Crop Sci* **1994**, *2*, 154-161.
6. Qian, J.; Song, J.; Gao, H.; Zhu, Y.; Xu, J.; Pang, X.; Yao, H.; Sun, C.; Li, X.; Li, C., *et al.* The complete chloroplast genome sequence of the medicinal plant salvia miltiorrhiza. *PloS one* **2013**, *8*, e57607.
7. Yang, M.; Zhang, X.; Liu, G.; Yin, Y.; Chen, K.; Yun, Q.; Zhao, D.; Al-Mssallem, I.S.; Yu, J. The complete chloroplast genome sequence of date palm (phoenix dactylifera l.). *PloS one* **2010**, *5*, e12762.
8. Yang, J.B.; Tang, M.; Li, H.T.; Zhang, Z.R.; Li, D.Z. Complete chloroplast genome of the genus cymbidium: Lights into the species identification, phylogenetic implications and population genetic analyses. *BMC evolutionary biology* **2013**, *13*, 84.
9. Shaw, J.; Lickey, E.B.; Schilling, E.E.; Small, R.L. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare iii. *American journal of botany* **2007**, *94*, 275-288.

10. Wicke, S.; Muller, K.F.; de Pamphilis, C.W.; Quandt, D.; Wickett, N.J.; Zhang, Y.; Renner, S.S.; Schneeweiss, G.M. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *The Plant cell* **2013**, *25*, 3711-3725.
11. Umesono, K.; Inokuchi, H.; Ohyama, K.; Ozeki, H. Nucleotide sequence of marchantia polymorpha chloroplast DNA: A region possibly encoding three trnas and three proteins including a homologue of e. Coli ribosomal protein s14. *Nucleic acids research* **1984**, *12*, 9551-9565.
12. Varshney, R.K.; Nayak, S.N.; May, G.D.; Jackson, S.A. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in biotechnology* **2009**, *27*, 522-530.
13. Dong, W.; Liu, J.; Yu, J.; Wang, L.; Zhou, S. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PloS one* **2012**, *7*, e35071.
14. Huang, Y.Y.; Matzke, A.J.; Matzke, M. Complete sequence and comparative analysis of the chloroplast genome of coconut palm (cocos nucifera). *PloS one* **2013**, *8*, e74736.
15. Raubeson, L.A.; Peery, R.; Chumley, T.W.; Dziubek, C.; Fourcade, H.M.; Boore, J.L.; Jansen, R.K. Comparative chloroplast genomics: Analyses including new sequences from the angiosperms nuphar advena and ranunculus macranthus. *BMC genomics* **2007**, *8*, 174.
16. Ye, C.Y.; Lin, Z.; Li, G.; Wang, Y.Y.; Qiu, J.; Fu, F.; Zhang, H.; Chen, L.; Ye, S.; Song, W., *et al.* Echinochloa chloroplast genomes: Insights into the evolution and taxonomic identification of two weedy species. *PloS one* **2014**, *9*, e113657.
17. Sanchez-Puerta, M.V.; Abbona, C.C. The chloroplast genome of hyoscyamus niger and a phylogenetic study of the tribe hyoscyameae (solanaceae). *PloS one* **2014**, *9*, e98353.
18. Kahlau, S.; Aspinall, S.; Gray, J.C.; Bock, R. Sequence of the tomato chloroplast DNA and evolutionary comparison of solanaceous plastid genomes. *Journal of molecular evolution* **2006**, *63*, 194-207.
19. Sasaki, T.; Yukawa, Y.; Miyamoto, T.; Obokata, J.; Sugiura, M. Identification of rna editing sites in chloroplast transcripts from the maternal and paternal progenitors of tobacco (nicotiana tabacum): Comparative analysis shows the involvement of distinct trans-factors for ndhb editing. *Molecular biology and evolution* **2003**, *20*, 1028-1035.
20. Gao, L.; Yi, X.; Yang, Y.X.; Su, Y.J.; Wang, T. Complete chloroplast genome sequence of a tree fern alsophila spinulosa: Insights into evolutionary changes in fern chloroplast genomes. *BMC evolutionary biology* **2009**, *9*, 130.
21. Curci, P.L.; De Paola, D.; Danzi, D.; Vendramin, G.G.; Sonnante, G. Complete chloroplast genome of the multifunctional crop globe artichoke and comparison with other asteraceae. *PloS one* **2015**, *10*, e0120589.
22. Park, I.; Kim, J.; Lee, J.; Kim, S.; Cho, O.; Yang, K.; Ahn, J.; Nahm, S.; Kim, H. Development of ssr markers by next-generation sequencing of korean landraces of chamoe (cucumis melo var. Makuwa). *Molecular biology reports* **2013**, *40*, 6855-6862.

23. Zalapa, J.E.; Cuevas, H.; Zhu, H.; Steffan, S.; Senalik, D.; Zeldin, E.; McCown, B.; Harbut, R.; Simon, P. Using next-generation sequencing approaches to isolate simple sequence repeat (ssr) loci in the plant sciences. *American journal of botany* **2012**, *99*, 193-208.
24. Wang, R.J.; Cheng, C.L.; Chang, C.C.; Wu, C.L.; Su, T.M.; Chaw, S.M. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *Bmc Evol Biol* **2008**, *8*, 36.
25. Redwan, R.M.; Saidin, A.; Kumar, S.V. Complete chloroplast genome sequence of md-2 pineapple and its comparative analysis among nine other plants from the subclass commelinidae. *BMC Plant Biol* **2015**, *15*, 196.
26. Song, Y.; Dong, W.; Liu, B.; Xu, C.; Yao, X.; Gao, J.; Corlett, R.T. Comparative analysis of complete chloroplast genome sequences of two tropical trees machilus yunnanensis and machilus balansae in the family lauraceae. *Frontiers in plant science* **2015**, *6*, 662.
27. Wu, F.H.; Chan, M.T.; Liao, D.C.; Hsu, C.T.; Lee, Y.W.; Daniell, H.; Duvall, M.R.; Lin, C.S. Complete chloroplast genome of oncidium gower ramsey and evaluation of molecular markers for identification and breeding in oncidiinae. *BMC plant biology* **2010**, *10*, 68.
28. Nie, X.; Lv, S.; Zhang, Y.; Du, X.; Wang, L.; Biradar, S.S.; Tan, X.; Wan, F.; Weining, S. Complete chloroplast genome sequence of a major invasive species, crofton weed (ageratina adenophora). *PloS one* **2012**, *7*, e36869.
29. Metzker, M.L. Sequencing technologies - the next generation. *Nature reviews. Genetics* **2010**, *11*, 31-46.
30. Cho, K.S.; Yun, B.K.; Yoon, Y.H.; Hong, S.Y.; Mekapogu, M.; Kim, K.H.; Yang, T.J. Complete chloroplast genome sequence of tartary buckwheat (fagopyrum tataricum) and comparative analysis with common buckwheat (f. Esculentum). *PloS one* **2015**, *10*, e0125332.
31. Kim, K.; Lee, S.C.; Lee, J.; Lee, H.O.; Joh, H.J.; Kim, N.H.; Park, H.S.; Yang, T.J. Comprehensive survey of genetic diversity in chloroplast genomes and 45s nrdnas within panax ginseng species. *PloS one* **2015**, *10*, e0117159.
32. Wicke, S.; Schneeweiss, G.M.; dePamphilis, C.W.; Muller, K.F.; Quandt, D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant molecular biology* **2011**, *76*, 273-297.
33. Saha, M.C.; Cooper, J.D.; Mian, M.A.; Chekhovskiy, K.; May, G.D. Tall fescue genomic ssr markers: Development and transferability across multiple grass species. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **2006**, *113*, 1449-1458.
34. Yap, J.Y.; Rohner, T.; Greenfield, A.; Van Der Merwe, M.; McPherson, H.; Glenn, W.; Kornfeld, G.; Marendy, E.; Pan, A.Y.; Wilton, A.*, et al.* Complete chloroplast genome of the wollemi pine (wollemia nobilis): Structure and evolution. *PloS one* **2015**, *10*, e0128126.
35. Ni, L.; Zhao, Z.; Dorje, G.; Ma, M. The complete chloroplast genome of ye-xing-ba (scrophularia dentata; scrophulariaceae), an alpine tibetan herb. *PloS one* **2016**, *11*, e0158488.

36. Kong, W.; Yang, J. The complete chloroplast genome sequence of morus mongolica and a comparative analysis within the fabidae clade. *Current genetics* **2016**, *62*, 165-172.
37. Zhang, Y.; Du, L.; Liu, A.; Chen, J.; Wu, L.; Hu, W.; Zhang, W.; Kim, K.; Lee, S.C.; Yang, T.J., *et al.* The complete chloroplast genome sequences of five epimedium species: Lights into phylogenetic and taxonomic analyses. *Frontiers in plant science* **2016**, *7*, 306.
38. Wang, R.J.; Cheng, C.L.; Chang, C.C.; Wu, C.L.; Su, T.M.; Chaw, S.M. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *Bmc Evol Biol* **2008**, *8*.
39. Khakhlova, O.; Bock, R. Elimination of deleterious mutations in plastid genomes by gene conversion. *The Plant journal : for cell and molecular biology* **2006**, *46*, 85-94.
40. Jheng, C.F.; Chen, T.C.; Lin, J.Y.; Chen, T.C.; Wu, W.L.; Chang, C.C. The comparative chloroplast genomic analysis of photosynthetic orchids and developing DNA markers to distinguish phalaenopsis orchids. *Plant Sci* **2012**, *190*, 62-73.
41. Yang, Y.; Zhou, T.; Duan, D.; Yang, J.; Feng, L.; Zhao, G. Comparative analysis of the complete chloroplast genomes of five quercus species. *Frontiers in plant science* **2016**, *7*, 959.
42. Li, Y.; Li, Q.; Li, X.; Song, J.; Sun, C. Complete chloroplast genome sequence of fritillaria unibracteata var. Wabuensis based on smrt sequencing technology. *Mitochondrial DNA. Part A, DNA mapping, sequencing, and analysis* **2016**, *27*, 3757-3758.
43. Li, Q.; Li, Y.; Song, J.; Xu, H.; Xu, J.; Zhu, Y.; Li, X.; Gao, H.; Dong, L.; Qian, J., *et al.* High-accuracy de novo assembly and snp detection of chloroplast genomes using a smrt circular consensus sequencing strategy. *The New phytologist* **2014**, *204*, 1041-1049.
44. Luo, R.; Liu, B.; Xie, Y.; Li, Z.; Huang, W.; Yuan, J.; He, G.; Chen, Y.; Pan, Q.; Liu, Y., *et al.* Soapdenovo2: An empirically improved memory-efficient short-read de novo assembler. *GigaScience* **2012**, *1*, 18.
45. Delcher, A.L.; Salzberg, S.L.; Phillippy, A.M. Using mummer to identify similar regions in large sequence sets. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* **2003**, *Chapter 10*, Unit 10 13.
46. Wyman, S.K.; Jansen, R.K.; Boore, J.L. Automatic annotation of organellar genomes with dogma. *Bioinformatics* **2004**, *20*, 3252-3255.
47. Lowe, T.M.; Eddy, S.R. Trnascan-se: A program for improved detection of transfer rna genes in genomic sequence. *Nucleic acids research* **1997**, *25*, 955-964.
48. 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. *Current genetics* **2007**, *52*, 267-274.
49. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. Mega6: Molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution* **2013**, *30*, 2725-2729.

50. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. Vista: Computational tools for comparative genomics. *Nucleic acids research* **2004**, *32*, W273-279.
51. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic acids research* **1999**, *27*, 573-580.
52. Warburton, P.E.; Giordano, J.; Cheung, F.; Gelfand, Y.; Benson, G. Inverted repeat structure of the human genome: The x-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. *Genome research* **2004**, *14*, 1861-1869.
53. Hall, T.A. Bioedit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/nt. *Nucleic Acid Symposium Series* **1999**, *41*.
54. Librado, P.; Rozas, J. Dnasp v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **2009**, *25*, 1451-1452.

Sample Availability: Not available.