

Article

Assembly of the Mitochondrial Genome in the Campanulaceae Family Using Illumina Low-Coverage Sequencing

Hyun-Oh Lee ^{1,2,†}, Ji-Weon Choi ^{3,†}, Jeong-Ho Baek ⁴, Jae-Hyeon Oh ⁵ , Sang-Choon Lee ¹ and Chang-Kug Kim ^{5,*} 

¹ Phyzen Genomics Institute, Seongnam 13558, Korea; dlgusdh88@phyzen.com (H.-O.L.); sclee0923@phyzen.com (S.-C.L.)

² Department of Plant Science, Seoul National University, Seoul 08826, Korea

³ Postharvest Technology Division, National Institute of Horticultural and Herbal Science, Wanju 55365, Korea; jwcnpri@rda.go.kr

⁴ Gene Engineering Division, National Institute of Agricultural Sciences, RDA, Jeonju 54874, Korea; firstleon@korea.kr

⁵ Genomics Division, National Institute of Agricultural Sciences, RDA, Jeonju 54874, Korea; jhoh8288@korea.kr

* Correspondence: chang@korea.kr; Tel.: +82-63-238-4555

† These authors contributed equally to this work.

Received: 28 June 2018; Accepted: 25 July 2018; Published: 30 July 2018



Abstract: *Platycodon grandiflorus* (balloon flower) and *Codonopsis lanceolata* (bonnet bellflower) are important herbs used in Asian traditional medicine, and both belong to the botanical family Campanulaceae. In this study, we designed and implemented a de novo DNA sequencing and assembly strategy to map the complete mitochondrial genomes of the first two members of the Campanulaceae using low-coverage Illumina DNA sequencing data. We produced a total of 28.9 Gb of paired-end sequencing data from the genomic DNA of *P. grandiflorus* (20.9 Gb) and *C. lanceolata* (8.0 Gb). The assembled mitochondrial genome of *P. grandiflorus* was found to consist of two circular chromosomes; the master circle contains 56 genes, and the minor circle contains 42 genes. The *C. lanceolata* mitochondrial genome consists of a single circle harboring 54 genes. Using a comparative genome structure and a pattern of repeated sequences, we show that the *P. grandiflorus* minor circle resulted from a recombination event involving the direct repeats of the master circle. Our dataset will be useful for comparative genomics and for evolutionary studies, and will facilitate further biological and phylogenetic characterization of species in the Campanulaceae.

Keywords: Campanulaceae; *Platycodon grandiflorus*; *Codonopsis lanceolata*; mitochondrial genome

1. Introduction

The large angiosperm Campanulaceae family comprises five subfamilies and over 2300 species. Members of this family—predominantly herbaceous perennials with capsular fruits—are found in almost all habitats [1]. Several Campanulaceae species have been used as medicinal plants in Asia. *Platycodon grandiflorus* (balloon flower) has distinct pharmacological effects in the treatment of coughs, excessive phlegm, and sore throats. Its root, called doraji in Korea, jiegeng in China, and kikyō in Japan, has been widely used as a traditional medicine in these countries [2]. In addition, *P. grandiflorus*, which is also used as a vegetable and an ornamental plant, shows strong ecological adaptability facilitated by resistance to drought, cold, and disease [3,4]. *Codonopsis lanceolata* (bonnet bellflower) exhibits various health benefits including antioxidant, antimicrobial, anti-inflammatory,

and immune-modulating properties. The root of *C. lanceolata* has been used in traditional medicine to treat various conditions and symptoms, including bronchitis, coughs, and lung injury [5].

In *P. grandiflorus* and *C. lanceolata*, most studies have reported on their pharmacological effects and chloroplast genome [2,4]. However, the mitochondrial genome has not been reported for the Campanulaceae family including *P. grandiflorus* and *C. lanceolata*. The plant mitochondrial genome is an essential organelle that ranges from 200 to 700 kb in size, with some exceeding 1 Mb [6], and typically contain abundant repeated sequences [7]. These dynamic structures are useful for molecular ecology, evolutionary biology, and phylogenetic studies. Recently, next-generation sequencing (NGS) and related bioinformatics tools have been providing improved strategies for assembling genomes in plant species [8,9]. However, the complexed mitochondrial genomes of plants are more difficult to analyze than animal species [10,11]. In this study, first in Campanulaceae family, we generated the complete mitochondrial genomes of *P. grandiflorus* and *C. lanceolata* using a de novo NGS strategy without reference sequence and characterized the features of the genomes.

2. Materials and Methods

2.1. Plant Materials and Whole-Genome Sequencing

Platycodon grandiflorus (voucher # IT105900) and *C. lanceolata* (voucher # IT239919) were obtained from the National Institute of Horticultural and Herbal Science (NIHHS, Wanju, Korea). Total genomic DNAs were extracted from fresh leaves using a modified cetyltrimethylammonium bromide (CTAB) method [12]. DNA quality and concentration were examined using a spectrophotometer and a 2100 Expert Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Paired-end (PE) libraries with an insert size of 270 to 700 bp were constructed according to the standard Illumina PE protocol and sequenced using an Illumina MiSeq platform (Illumina, San Diego, CA, USA) by Macrogen Biotechnology Center (Macrogen Inc., Seoul, Korea).

2.2. Assembly of Mitochondrial Genomes

In the assembling for organelle genome, high-quality read sequences can be obtained using low coverage data (0.5 to 5× based on the nuclear genome) [13]. In the schematic pipeline for de novo assembling (Figure 1), first, low-coverage PE data were extracted (Supplementary Table S1).

Second, PE data were filtered to obtain high-quality read sequences with Phred scores > 20 using the quality trim tool in the CLC Assembly Cell package ver. 4.2 (CLC Inc., Aarhus, Denmark). Third, four de novo assemblers were tested: Celera ver. 8.2 [14], SOAPdenovo ver. 2.02 [15], and SPAdes ver. 3.8 [16], CLC assembly Cell package (Supplementary Table S2). Fourth, high-quality read sequences were de novo assembled using the CLC assembly Cell package with the following parameters: distance between forward start and reverse end reads ranging from 300 to 700 bp; similarity fraction = 0.8; length fraction = 0.5. Fifth, mitochondrial contigs were selected using the mitochondrial genome sequence of *Ipomoea nil* [17] and the NUCmer tool [18]. Sixth, the contigs were extended by read mapping using the CLC reference mapper tool (CLC Inc, Denmark) with the same parameters as used for the de novo assembly. Seventh, mismatched assembly events were removed using BLASTN searches against the National Center for Biotechnology Information (NCBI) non-redundant nucleotide database. Finally, our processes generated a circular mitochondrial genome sequences whose gaps were completely filled using nucleotide sequences with the highest depth after read mapping.

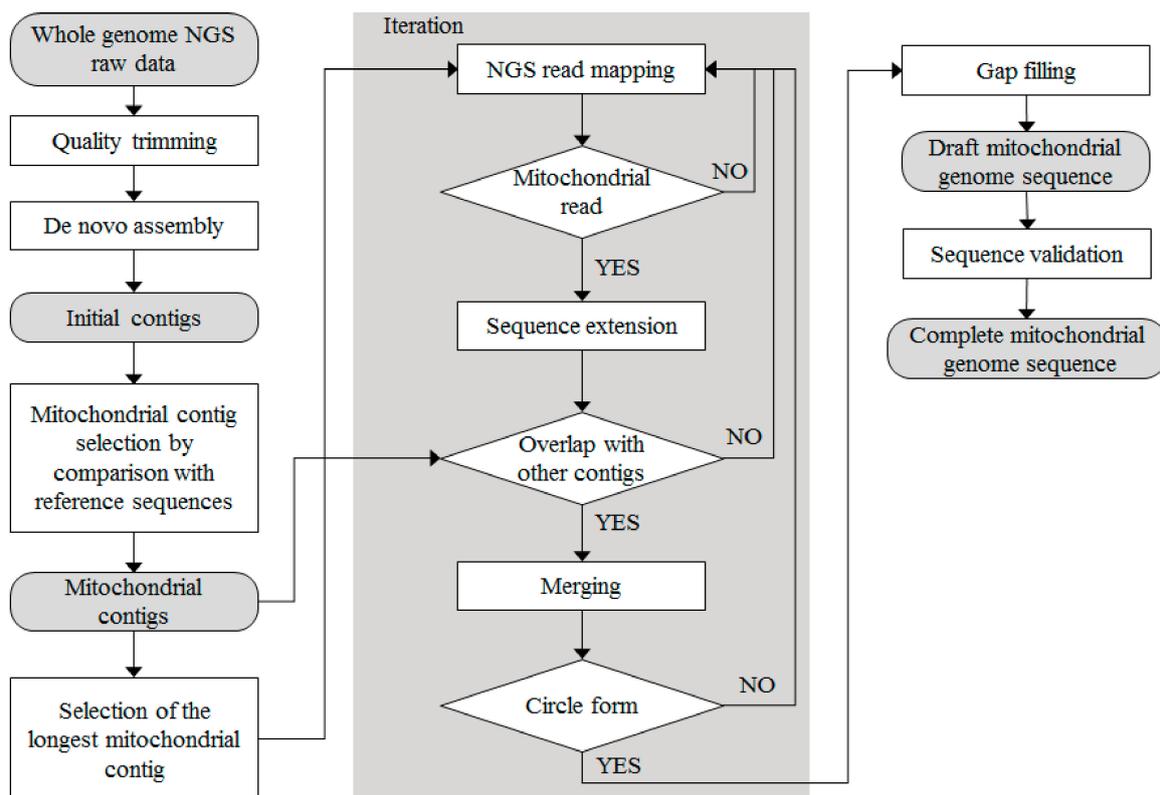


Figure 1. Pipeline for de novo assembly of mitochondrial genomes in the Campanulaceae family using Illumina, low-coverage, whole-genome next-generation sequencing (NGS) data.

2.3. Validation of Mitochondrial Genomes

The draft mitochondrial genome sequences were validated by PE read mapping and PCR amplification. For validation using PE read mapping, PE reads used for assembly were mapped on the mitochondrial genome sequences. Depth of mapped reads was investigated using the CLC Assembly Cell package and visualized using Microsoft Excel. Consistency and connectivity of mapped reads were confirmed. In addition, junctions between seed contigs were confirmed. For validation using PCR amplification, 12 primers were designed in the *P. grandiflorus* (Supplementary Table S3). The six primers were designed within contig sequences of >1 kb, and another six primers were designed at junction regions between contigs or within contig sequences of <1 kb. All PCRs were performed using the DNA Engine Tetrad 2 Peltier Thermal Cycler (BIO-RAD, Hercules, CA, USA). The following PCR conditions were used: 95 °C for 5 min; 35 cycles of 95 °C for 30 sec, 58 °C for 30 sec, and 72 °C for 1 min; and 72 °C for 10 min.

2.4. Annotation of Mitochondrial Genome

Mitochondrial genomes were annotated using the GeSeq program (Max Planck Institute, Golm, Germany). Three species were used as references for gene prediction: *Helianthus annuus* (KF815390) [19], *Daucus carota* (JQ248574) [8], and *Vaccinium macrocarpon* (KF386162) [20]. Ambiguous gene positions were manually corrected using NCBI BLASTN searches and the Artemis annotation tool (<http://www.sanger.ac.uk/science/tools/artemis>). Circular maps of mitochondrial genomes with the annotation information were drawn using the OGDRAW program [21]. The complete mitochondrial genome sequences assembled in this study were deposited into the GenBank database under KX887331, MG775429, and MG775430.

2.5. Analysis of Repetitive Sequences

Repetitive sequences in the mitochondrial genomes were searched using the REPuter program [22], with a minimum repeat size of 20 bp and four selected repeat types (forward, reverse, complement, and reverse complement). The detected repetitive sequences were divided into two groups based on a size cutoff of 100 bp. Repetitive sequences of >100 bp were depicted on the circular genome maps using the ClicO FS (Codon Genomics Inc., Selangor, Malaysia) program. Tandem repeat sequences were detected using the Tandem Repeats Finder software (BU-Bioinformatics, MA, USA) with the following parameters: alignment parameters of 2 bp match, 2 bp mismatch, and 7-bp insertion and deletions (InDels); minimum alignment score of 50; maximum period size of 500; and maximum tandem repeat array size of 2 bp.

2.6. Bayesian Phylogenetic Analysis

Phylogenetic analysis based on Bayesian inference was performed using 14 protein-coding sequences (*atp1*, *ccmB*, *ccmC*, *ccmFc*, *ccmFn*, *cox1*, *cox3*, *matR*, *nad3*, *nad6*, *nad7*, *rps4*, *rps12*, and *rps13*) conserved in the mitochondrial genomes of 16 species in the asterids group including *Ajuga reptans* (KF709392), *Asclepias syriaca* (KF541337), *Boea hygrometrica* (JN107812), *Capsicum annuum* cultivar Jeju (KJ865410), *Castilleja paramensis* (KT959112), *C. lanceolata* (MG775430), *D. carota* (JQ2485740), *H. annuus* (KF815390), *Hyoscyamus niger* (KM207685), *I. nil* (AP017303), *Mimulus guttatus* (JN098455), *Nicotiana tabacum* (BA000042), *P. grandiflorus* (KX887331), *Rhazya stricta* (KJ485850), *Salvia miltiorrhiza* (KF177345), and *V. macrocarpon* (KF386162). In order to obtain a phylogenetic tree, BEAST ver. 2.4 program was employed and Markov chain Monte Carlo runs were performed for 1,000,000 chain length and sampling frequency set to every 10,000 steps [23]. The phylogenetic tree was annotated as a maximum clade credibility tree using the TreeAnnotator ver. 2.4 [23].

3. Results

3.1. Pipeline for the Assembly

We generated the complete mitochondrial genomes of *P. grandiflorus* and *C. lanceolata*, via de novo assembly using NGS data without reference genomes. The raw PE data were produced from *P. grandiflorus* (20.9 Gb, 79.1× coverage) and *C. lanceolata* (8.0 Gb, 161.4× coverage), respectively. The high quality 23,971,262 (*P. grandiflorus*, 95.9% of all sequences) reads and 23,424,520 (*C. lanceolata*, 87.7% of all sequences) reads were obtained after quality trimming (Supplementary Table S1). When four de novo assemblers were compared with two classified contigs group, SOAPdenovo assembler generated the smallest number of 58 contigs and SPAdes assembler resulted in the longest 1.8 Mb contig length in the *P. grandiflorus*. However, N50 and L50 statistics show that SOAPdenovo and CLC assembler have better performance than other assemblers. Therefore, we performed the assembly using the CLC assembler (Supplementary Table S2).

3.2. Complete Mitochondrial Genomes

In the *P. grandiflorus*, mitochondrial genome consists of two circular chromosomes, master circle of 1,249,593 bp and minor circle of 1,070,431 bp (Supplementary Table S4). The master circle harbored a pair of large identical 21.8 kb repeats (R1 and R2), which divided the genome into two regions (Figure 2A,C and Figure 3A). The minor circle contained region A and a single copy of the large repeat, both of which were identical to the ones found in the master circle (Figures 2C and 3B).

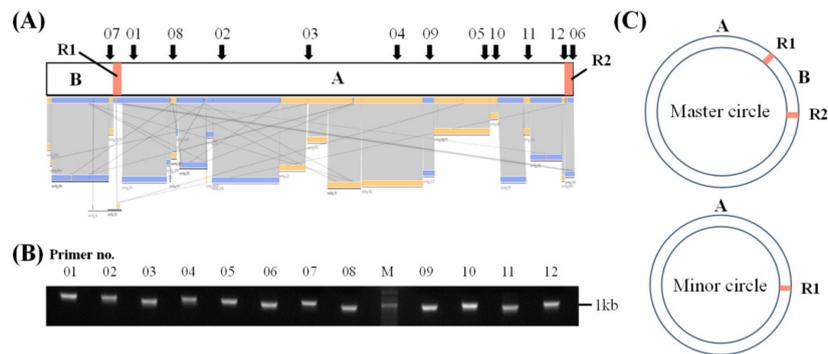


Figure 2. Assembly of the mitochondrial genome sequences of *Platycodon grandiflorus*. (A) Mitochondrial contigs that were utilized in the assembly of the complete mitochondrial genome. Blue and yellow bars indicate mitochondrial contigs used for extension, gap-filling, and merging. Black arrows indicate target locations of PCR primers used for validation. Primers 01 through 06 annealed within >1 kb contigs, and primers 07 through 12 annealed at junction regions between contigs or within <1 kb contigs; (B) PCR validation of the assembled master circle mitochondrial sequence. The 12 primer sets were used for genomic DNA PCR analysis. The sequences of PCR products of expected size were confirmed to be the same as the assembled sequences. M indicates the DNA ladder; (C) Schematic diagrams of the master circle and the minor circle of the *P. grandiflorus* mitochondrial genome, showing mitochondrial heteroplasmy in *P. grandiflorus*. The master circle consists of two regions, A and B, separated by the 21.8 kb large repeats R1 and R2. The minor circle contains only region A.

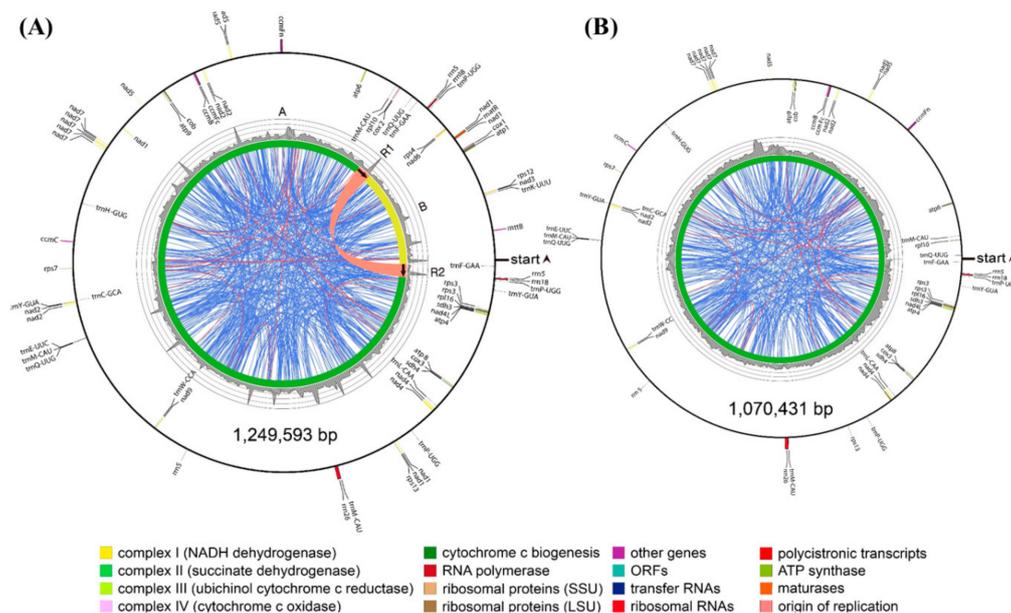


Figure 3. Maps of the *P. grandiflorus* mitochondrial genome. (A) Master circle; (B) Minor circle. Direction of gene transcription is represented by the location of the genes on the inside (clockwise) or the outside (counterclockwise) of the outermost circle. Inner histogram indicates average coverage depth of reads mapping to the mitochondrial genome sequence (each grey circle represents 50×, with a maximum of 200×). Innermost circle represents the structural scheme. Repeats of >100 bp are indicated by connecting orange lines and repeats of <100 bp are indicated by connecting blue lines. The thick orange line connects the largest 21.8 kb R1 and R2 repeat regions in the master circle. Tandem repeats are not shown.

Thus, due to the presence of region B and the R2 repeat in the master circle, the two *P. grandiflorus* mitochondrial genome structures shared 85.7% sequence similarity (Supplementary Figure S1).

In the *C. lanceolata*, mitochondrial genome was revealed as a 403,704 bp circle assembled using 11 mitochondrial contigs, with no large repeats detected (Figure 4).

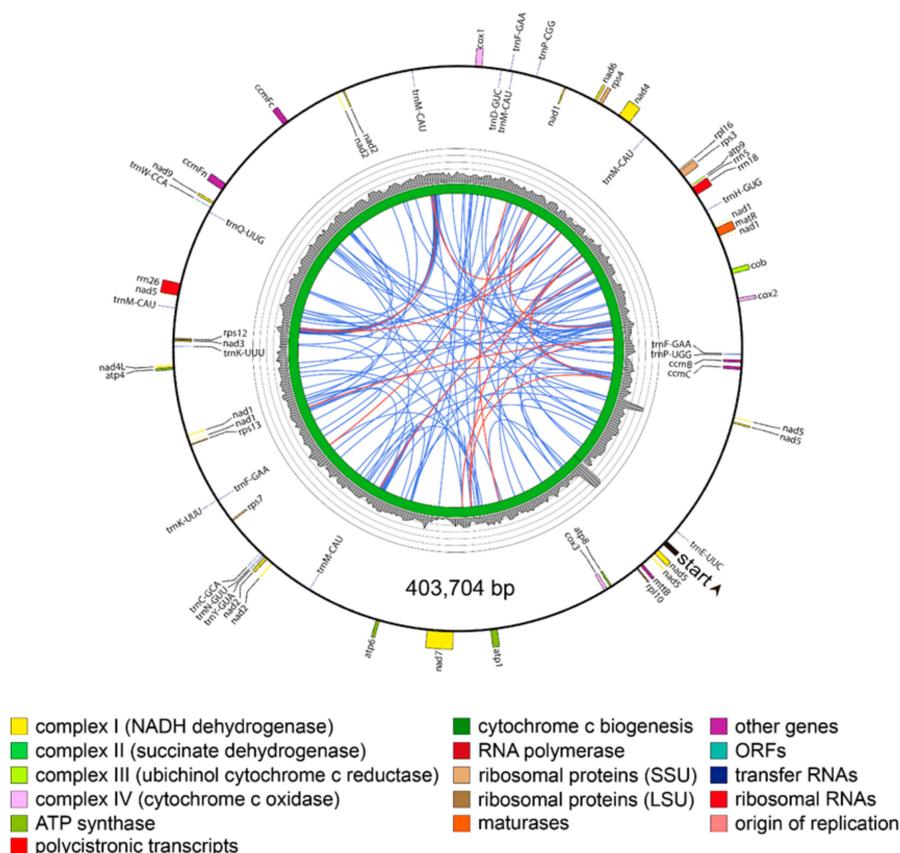


Figure 4. Mitochondrial genome map of *Codonopsis lanceolata*. Direction of transcription is represented by the location of the respective genes on the inside (clockwise) or the outside (counterclockwise) of the outermost circle. Inner histogram indicates average coverage depth of reads mapped to the mitochondrial genome sequence (each grey circle represents 60×, with a maximum of 300×). Innermost circle indicates the structural scheme. Repeats of >100 bp are indicated by connecting orange lines and repeats of <100 bp are indicated by connecting blue lines. Tandem repeats are not shown.

In order to validate the draft mitochondrial genome, first, sequence of the *P. grandiflorus* master circle was validated by PCR amplification and nucleotide sequencing using 12 primer pairs designed to anneal within junction regions and contigs (Figure 2B). Finally, the sequences of the two circles were validated by NGS read mapping, demonstrating consistency and connectivity of mapped reads at the whole-genome level and junctions between contigs (Supplementary Figure S1). In the *C. lanceolata*, mitochondrial genome was validated by consistency and connectivity of the mapped reads (Supplementary Figures S2 and S3). Although the *P. grandiflorus* and *C. lanceolata* shared <21.9% overall sequence similarity, the coding regions showed substantially higher similarity relative to the intergenic regions. These multi-validation steps confirm the reliability of our assembly pipeline, as well as the reliability of the complete mitochondrial genome sequence.

3.3. Annotated Genes in the Mitochondrial Genomes

In the *P. grandiflorus*, we identified 56 and 42 putative genes within the master and minor circles of the mitochondrial genome, respectively. In addition, the master and minor circles contained 46 and 36 non-redundant genes, respectively. In the *C. lanceolata*, mitochondrial genome contained a total of 54 genes, of which 47 were identified as non-redundant (Supplementary Table S4). Total

of 23 redundant genes were detected by only dispersed repeats in *P. grandiflorus* and *C. lanceolata*. These redundant genes were found in the master circle of *P. grandiflorus* (*rrn18*, *rrn5_2 copy*, *trnF-GAA*, *trnM-CAU_2 copy*, *trnP-UGG_2 copy*, *trnQ-UUG*, *trnW-CCA*), minor circle of *P. grandiflorus* (*rrn5*, *trnM-CAU_2 copy*, *trnP-UGG*, *trnQ-UUG*, *trnY-GUA*), and *C. lanceolata* (*trnF-GAA_2 copy*, *trnK-UUU*, *trnM-CAU_4 copy*), respectively.

In the *P. grandiflorus* master circle, total 46 non-redundant genes accounted for 3.23% of the sequence, comprised 32 protein-coding genes, 11 transfer RNAs (tRNAs), and 3 ribosomal RNAs (rRNAs). However, the minor circle lacked 10 genes (i.e., *nad1*, *nad3*, *nad6*, *cox1*, *atp1*, *rps4*, *rps12*, *matR*, *mttB*, and *trnK-UUU*) compared to the master circle. In the *C. lanceolata*, non-redundant genes accounted for 9.91% of the sequence, comprising 31 protein-coding genes, 13 tRNAs, and three rRNAs (Table 1). In addition, we compared the reported mitochondrial genome of *H. annuus* (sunflower), which is a closely related species belonging to the asterids group [19].

Table 1. Annotated genes in the mitochondrial genomes of *P. grandiflorus* (including master and minor circle) and *C. lanceolata*, and comparison with *Helianthus annuus*.

Category ¹	Common Genes	Unique Genes			
		<i>P. Grandiflorus</i>		<i>C. Lanceolata</i>	<i>H. Annuus</i>
		(Master)	Minor		
Complex I	<i>nad4L</i> , <i>nad9</i>	<i>nad1–4</i> , <i>nad6</i> , <i>nad7</i>	<i>nad2</i> , <i>nad4</i> , <i>nad7</i>	<i>nad1–7</i>	<i>nad3</i> , <i>nad5–6</i>
Complex II		<i>sdh3</i> , <i>sdh4</i>	<i>sdh3</i> , <i>sdh4</i>		
Complex III	<i>cob</i>				
Complex IV	<i>cox3</i>	<i>cox1</i> , <i>cox2</i>	<i>cox2</i>	<i>cox1</i> , <i>cox2</i>	<i>cox1</i>
Complex V	<i>atp4</i> , <i>atp6</i> , <i>atp8</i> , <i>atp9</i>	<i>atp1</i>		<i>atp1</i>	<i>atp1</i>
Cytochrome	<i>ccmB</i> , <i>ccmC</i> , <i>ccmFc</i> , <i>ccmFn</i>				
Large subunit	<i>rpl10</i> , <i>rpl16</i>				<i>rpl5</i>
Small subunit	<i>rps13</i>	<i>rps3</i> , <i>rps4</i> , <i>rps7</i> , <i>rps12</i>	<i>rps3</i> , <i>rps7</i>	<i>rps3</i> , <i>rps4</i> , <i>rps7</i> , <i>rps12</i>	<i>rps4</i> , <i>rps12</i> , <i>rps13</i>
Maturase		<i>matR</i>		<i>matR</i>	<i>matR</i>
Transport protein		<i>mttB</i>		<i>mttB</i>	
Ribosomal RNAs	<i>rrn5</i> , <i>rrn18</i> , <i>rrn26</i>				
Transfer RNAs	<i>trnC-GCA</i> , <i>trnE-UUC</i> , <i>trnF-GAA</i> , <i>trnH-GUG</i> , <i>trnM-CAU</i> , <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnW-CCA</i> , <i>trnY-GUA</i>	<i>trnK-UUU</i> , <i>trnL-CAA</i>	<i>trnL-CAA</i>	<i>trnD-GUC</i> , <i>trnK-UUU</i> , <i>trnN-GUU</i> , <i>trnP-CGG</i>	<i>trnD-GUC</i> , <i>trnG-GCC</i> , <i>trnN-GUU</i> , <i>trnS-GCT</i>
Total genes	27	19	9	20	13

¹ Complex I (NADH dehydrogenase), complex II (succinate dehydrogenase), complex III (ubiquinol cytochrome c reductase), complex IV (cytochrome c oxidase), complex V (ATP synthase), cytochrome c biogenesis, large subunit ribosomal proteins, and small subunit ribosomal proteins.

The 27 common genes are conserved in the three species (i.e., *P. grandiflorus*, *C. lanceolata*, and *H. annuus*). Most of the genes were associated with the electron transport chain, cytochrome c biogenesis, and protein synthesis. The unique genes (i.e., *sdh3* and *sdh4*) of *P. grandiflorus* were found in the encoding subunits of the succinate dehydrogenase complex (Complex II) category. The trans-spliced protein-coding genes were found in the *P. grandiflorus* master circle (*nad1* and *nad2*), minor circle (*nad2*), and *C. lanceolata* (*nad1* and *nad2*). The *mttB* gene, which catalyzes the transfer of a methyl group from trimethylamine, was found in the master circle of *P. grandiflorus* and *C. lanceolata* (Table 1).

3.4. Repetitive Sequences in the Mitochondrial Genomes

Repetitive sequences consist mostly of dispersed repeats and tandem repeats, and they can vary in size in plant mitochondrial genomes [7]. The dispersed repeats are segments of DNA that

4. Discussion

In this study, we first designed and implemented a de novo assembly pipeline to map the complete mitochondrial genomes of Campanulaceae family using NGS data. In order to find the optimal method of assembling, we compared the efficiency of four de novo assemblers. Ultimately, our method employed the CLC Assembly Cell assembler, which demonstrated superior performance with respect to N50 and L50 relative statistics. However, other plants have been generated using different assemblers, such as MITObim, Newbler, and PLATANUS [26–29], combination of assemblers [30]. Thus, we suggest that the assembly method is different depending on the NGS data type and complexity characteristics of each plant.

The *P. grandiflorus* and *C. lanceolata* mitochondrial genomes were within the 1.5 to 2000 kb range observed for reported mitochondrial genomes of 223 land plant species (Supplementary Figure S5). Although two species are in the same Campanulaceae family, mitochondrial genome sizes of *P. grandiflorus* were more than two times larger than the *C. lanceolata*. Substantial within-family differences have been observed between the members of other plant families. In the Cucurbitaceae family, cucumber and watermelon are 1556 kb and 379 kb, respectively [31,32]. Despite differences in overall size, the mitochondrial genomes of *P. grandiflorus* and *C. lanceolata* harbored similar numbers of coding genes, as reported in other plant mitochondrial genomes (Supplementary Figure S6). We revealed the presence of species-specific genes, as well as potential gene loss in *P. grandiflorus* and *C. lanceolata*. Interestingly, genes encoding subunits of the succinate dehydrogenase complex (Complex II), *sdh3* and *sdh4*, were found only in the master and minor circles of *P. grandiflorus*, whereas *mttB* gene was found only in the master circle of *P. grandiflorus* and in the *C. lanceolata* genome. These *sdh3* and *sdh4* genes were transferred from the mitochondrial genome to the nuclear genome during angiosperm evolution, and diverse angiosperm species have retained both genes in their mitochondrial genomes [33]. Therefore, *sdh3* and *sdh4* genes in *C. lanceolata* and *H. annuus* are likely present in the nuclear genomes of these two species. Overall, our findings regarding mitochondrial gene distribution underscored the dynamic nature of mitochondrial genomes.

Mitochondrial genomes of *P. grandiflorus* and *C. lanceolata* were characterized by high-repeat content and low gene density, properties that are typical of plant mitochondrial genomes. The differences of plant mitochondrial genomes can be shown by repetitive sequences and frequent recombination events [34]. The dispersed and tandem repeats were proportional to the size of the mitochondrial genome, and intergenic, region-specific repeats have likely contributed to increases in the complexity of mitochondrial genome [7,35]. The *H. annuus* has a similar distribution of repeats to mitochondrial genomes of *P. grandiflorus* and *C. lanceolata*, despite the differences in dispersed repeats. However, the localization of the dispersed repeats was similar in all mitochondrial genomes, with >90% of repeats associating with the intergenic regions in three species. Although there was a difference between the mitochondrial genome length and the dispersed repeats, three plant species (i.e., *P. grandiflorus*, *C. lanceolata*, and *H. annuus*) were similar in GC content, total gene number, and average gene length (Supplementary Table S4).

Interestingly, we identified a multipartite (i.e., master and minor circles) mitochondrial genome of *P. grandiflorus*. Multipartite mitochondrial genomes have been reported in some plants, animals, and fungi [36]. Generally, multipartite genome structure was generated using high-frequency recombination via repeated sequences, intramolecular homologous recombination, random genetic drift, and loss of mitochondrial single-strand binding protein [37,38]. In the heteroplasmy, in which divergent mitochondrial genotypes co-exist in a plant cell [39], *P. grandiflorus* is potentially present in additional types of mitochondrial genomes. However, in the multipartite mitochondrial genomes of *P. grandiflorus*, substantial differences exist in both sequence and structural levels with repeat sequences (Figures 2 and 3). The comparison of sequence and linear genome map of *P. grandiflorus* indicates that the master and minor circle are almost identical, except for deletion of 179 kb, one direct repeat (R2 region), and related region (B region) in minor circle (Figure 2, Supplementary Figures S1 and S2). Therefore, we assume that the largest repeats of the master circle contributed to creation of

the minor circle of *P. grandiflorus*. In additional validation, we are going to verify the heteroplasmy of *P. grandiflorus* using higher coverage and longer read sequencing data, under assembly conditions that can attenuate the issues associated with high-depth repeats.

Despite many phylogenetic analyses using chloroplast and nuclear sequences, such as *rbcL*, the internal transcribed spacer (ITS), and pentatricopeptide repeat (PPR) genes [40–43], the Campanulaceae family's phylogenetic relationships have remained poorly characterized [44]. To date, mitochondrial sequences have not been utilized for phylogenetic analysis of the Campanulaceae family species. Although there are limited mitochondrial sequences, we sought to perform a mitochondrial, genome-based analysis using our newly assembled data. Unlike sequence and structural conversation of chloroplast genomes among plant species, mitochondrial genomes show much difference both at sequence and structural level due to abundance of repetitive sequences and frequent recombination (Figures 2–4, Supplementary Figures S1–S3). Therefore, instead of whole mitochondrial genome sequence or total protein-coding sequences, we use the 14 conserved protein-coding sequences commonly present among 16 mitochondrial genomes of asterids group (Figure 5). The phylogenetic tree divided 16 species into six groups. The *P. grandiflorus* and *C. lanceolata* clustered the Asterales order group with *H. annuus*. The *D. carota* (carrot) demonstrated a sister relationship in the Apiales order. The phylogenetic relationship of *P. grandiflorus*, *C. lanceolata*, and other species was in agreement with established taxonomical classifications [40,45].

Overall, we first assembled the complete mitochondrial genomes in the Campanulaceae family; annotated genes will be useful for biology and plant evolution. In addition, our newly designed assembly pipeline will facilitate further analyses of plant mitochondrial genomes.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4425/9/8/383/s1>, Figure S1: Sequence comparison between the master and the minor circles of the *P. grandiflorus* mitochondrial genome. Figure S2: Linear genome map and read depth of the mitochondrial genomes of *P. grandiflorus* and *C. lanceolata*. Figure S3: Mitochondrial genome sequence comparison between *P. grandiflorus* and *C. lanceolata*. Figure S4: Repetitive sequences in the mitochondrial genomes of *P. grandiflorus* (including master and minor circle), *C. lanceolata*, and *H. annuus*. Figure S5: Size distribution of mitochondrial genomes of land plant species, based on the NCBI Organelle Genome database (<https://www.ncbi.nlm.nih.gov/genome/browse/?report=5#!/organelles/>). The 223 mitochondrial genomes of land plant species have been deposited into the database. Figure S6. Numbers of annotated genes in assembled mitochondrial genomes of land plants, as revealed by the NCBI Organelle Genome database. Table S1: Overview of raw and trimmed data. Table S2: Comparison of four assemblers with two contig group. Table S3: Primers used for assembly validation. Table S4: Features of assembled mitochondrial genomes of *P. grandiflorus* (including master and minor circle) and *C. lanceolata* with *H. annuus*. Table S5: Repeat regions identified in the *P. grandiflorus* (including master and minor circle), *C. lanceolata* and *H. annuus*. Table S6: Tandem repeat sequences identified in the *P. grandiflorus* (including master and minor circle), *C. lanceolata*, and *H. annuus*.

Author Contributions: C.-K.K. and J.-W.C. planned and designed the research. H.-O.L. contributed genome assembly and comparative analysis. J.-H.O. conducted genome sequencing, and J.-W.C. contributed materials. J.-H.B. and S.-C.L. analyzed data and interpreted the results. C.-K.K. and H.-O.L. wrote the manuscript, and C.-K.K. edited the manuscript. All authors read and approved the final manuscript.

Funding: This study was support by the Research Program for Agricultural Science & Technology Development (Fund No. PJ013485) and the Post-genome project (Fund No. PJ010351), Rural Development Administration.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Lammers, T.G. *The Families and Genera of Vascular Plants*; Kadereit, J.W., Bittrich, V., Eds.; Springer: New York, NY, USA, 2007; Volume 8, pp. 26–56.
2. Hong, D.; Song, G.; Lammers, T.G.; Klein, L. *Flora of China*; Zhengyi, W., Raven, P.H., Hong, D.Y., Eds.; Science Press: Beijing, China; Missouri Botanical Garden: St. Louis, MO, USA, 2011; Volume 19.
3. Halevy, A.H.; Shlomo, E.; Ziv, O. Improving cut flower production of balloon flower. *HortScience* **2002**, *37*, 759–761.

4. Zhang, L.; Wang, Y.; Yang, D.; Zhang, C.; Zhang, N.; Li, M.; Liu, Y. *Platycodon grandiflorus*—An ethnopharmacological, phytochemical and pharmacological review. *J. Ethnopharmacol.* **2015**, *164*, 147–161. [[CrossRef](#)] [[PubMed](#)]
5. Hossen, M.J.; Kim, M.Y.; Kim, J.H.; Cho, J.Y. *Codonopsis lanceolata*: A review of its therapeutic potentials. *Phytother. Res.* **2016**, *30*, 347–356. [[CrossRef](#)] [[PubMed](#)]
6. Buchanan, B.B.; Gruissem, W.; Jones, R.L. *Biochemistry & Molecular Biology of Plants*; Jones, R.L., Buchanan, B.B., Gruissem, W., Eds.; American Society of Plant Physiologists: Rockville, MD, USA, 2000; Volume 40.
7. Gualberto, J.M.; Newton, K.J. Plant mitochondrial genomes: Dynamics and mechanisms of mutation. *Annu. Rev. Plant Biol.* **2017**, *68*, 225–252. [[CrossRef](#)] [[PubMed](#)]
8. Iorizzo, M.; Senalik, D.; Szklarczyk, M.; Grzebelus, D.; Spooner, D.; Simon, P. De novo assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome. *BMC Plant Biol.* **2012**, *12*, 61. [[CrossRef](#)] [[PubMed](#)]
9. Li, F.; Yang, A.; Lv, J.; Gong, D.; Sun, Y. The complete mitochondrial genome sequence of Sua-type cytoplasmic male sterility of tobacco (*Nicotiana tabacum*). *Mitochondrial DNA Part A* **2016**, *27*, 2929–2930. [[CrossRef](#)] [[PubMed](#)]
10. Hajibabaei, M.; Singer, G.A.; Hebert, P.D.; Hickey, D.A. DNA barcoding: How it complements taxonomy, molecular phylogenetics and population genetics. *TRENDS Genet.* **2007**, *23*, 167–172. [[CrossRef](#)] [[PubMed](#)]
11. Caballero, J.; Smit, A.F.; Hood, L.; Glusman, G. Realistic artificial DNA sequences as negative controls for computational genomics. *Nucleic Acids Res.* **2014**, *42*, e99. [[CrossRef](#)] [[PubMed](#)]
12. Allen, G.; Flores-Vergara, M.; Krasynanski, S.; Kumar, S.; Thompson, W. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* **2006**, *1*, 2320. [[CrossRef](#)] [[PubMed](#)]
13. Kim, K.; Lee, S.-C.; Lee, J.; Yu, Y.; Yang, K.; Choi, B.-S.; Koh, H.-J.; Waminal, N.E.; Choi, H.-I.; Kim, N.-H. Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of *Oryza* Aa genome species. *Sci. Rep.* **2015**, *5*, 15655. [[CrossRef](#)] [[PubMed](#)]
14. Myers, E.W.; Sutton, G.G.; Delcher, A.L.; Dew, I.M.; Fasulo, D.P.; Flanigan, M.J.; Kravitz, S.A.; Mobarry, C.M.; Reinert, K.H.; Remington, K.A. A whole-genome assembly of *Drosophila*. *Science* **2000**, *287*, 2196–2204. [[CrossRef](#)] [[PubMed](#)]
15. Luo, R.; Liu, B.; Xie, Y.; Li, Z.; Huang, W.; Yuan, J.; He, G.; Chen, Y.; Pan, Q.; Liu, Y. Soapdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience* **2012**, *1*, 18. [[CrossRef](#)] [[PubMed](#)]
16. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Pribelski, A.D. Spades: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [[CrossRef](#)] [[PubMed](#)]
17. Hoshino, A.; Jayakumar, V.; Nitasaka, E.; Toyoda, A.; Noguchi, H.; Itoh, T.; Shin, T.; Minakuchi, Y.; Koda, Y.; Nagano, A.J. Genome sequence and analysis of the Japanese morning glory *Ipomoea Nil*. *Nat. Commun.* **2016**, *7*, 13295. [[CrossRef](#)] [[PubMed](#)]
18. Delcher, A.L.; Salzberg, S.L.; Phillippy, A.M. Using MUMmer to identify similar regions in large sequence sets. *Curr. Protoc. Bioinform.* **2003**, 10.3.11–10.3.18. [[CrossRef](#)] [[PubMed](#)]
19. Grassa, C.J.; Ebert, D.P.; Kane, N.C.; Rieseberg, L.H. Complete mitochondrial genome sequence of sunflower (*Helianthus Annuus* L.). *Genome Announc.* **2016**, *4*, e00981-16. [[CrossRef](#)] [[PubMed](#)]
20. Fajardo, D.; Schlautman, B.; Steffan, S.; Polashock, J.; Vorsa, N.; Zalapa, J. The american cranberry mitochondrial genome reveals the presence of selenocysteine (tRNA-Sec and SECIS) insertion machinery in land plants. *Gene* **2014**, *536*, 336–343. [[CrossRef](#)] [[PubMed](#)]
21. Lohse, M.; Drechsel, O.; Bock, R. Organellergenomedraw (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](#)] [[PubMed](#)]
22. 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](#)]

23. Bouckaert, R.; Heled, J.; Kühnert, D.; Vaughan, T.; Wu, C.-H.; Xie, D.; Suchard, M.A.; Rambaut, A.; Drummond, A.J. Beast 2: A software platform for bayesian evolutionary analysis. *PLoS Comput. Biol.* **2014**, *10*, e1003537. [[CrossRef](#)] [[PubMed](#)]
24. Richard, G.-F.; Kerrest, A.; Dujon, B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol. Mol. Biol. Rev.* **2008**, *72*, 686–727. [[CrossRef](#)] [[PubMed](#)]
25. 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. [[CrossRef](#)] [[PubMed](#)]
26. Lin, C.-P.; Lo, H.-F.; Chen, C.-Y.; Chen, L.-F.O. The complete mitochondrial genome of mungbean *Vigna radiata* Var. *radiata* NM92 and a phylogenetic analysis of crops in angiosperms. *Mitochondrial DNA Part A* **2016**, *27*, 3731–3732.
27. Petersen, G.; Cuenca, A.; Zervas, A.; Ross, G.T.; Graham, S.W.; Barrett, C.F.; Davis, J.I.; Seberg, O. Mitochondrial genome evolution in Alismatales: Size reduction and extensive loss of ribosomal protein genes. *PLoS ONE* **2017**, *12*, e0177606. [[CrossRef](#)] [[PubMed](#)]
28. Xu, Y.; Zhang, F.; Hu, K. The complete mitochondrial genome sequence of an annual wild tobacco *Nicotiana attenuata*. *Mitochondrial DNA Part B* **2017**, *2*, 924–925. [[CrossRef](#)]
29. Ye, N.; Wang, X.; Li, J.; Bi, C.; Xu, Y.; Wu, D.; Ye, Q. Assembly and comparative analysis of complete mitochondrial genome sequence of an economic plant *Salix suchowensis*. *PeerJ* **2017**, *5*, e3148. [[CrossRef](#)] [[PubMed](#)]
30. Silva, S.R.; Alvarenga, D.O.; Aranguren, Y.; Penha, H.A.; Fernandes, C.C.; Pinheiro, D.G.; Oliveira, M.T.; Michael, T.P.; Miranda, V.F.; Varani, A.M. The mitochondrial genome of the terrestrial carnivorous plant *Utricularia reniformis* (Lentibulariaceae): Structure, comparative analysis and evolutionary landmarks. *PLoS ONE* **2017**, *12*, e0180484. [[CrossRef](#)] [[PubMed](#)]
31. Alverson, A.J.; Wei, X.; Rice, D.W.; Stern, D.B.; Barry, K.; Palmer, J.D. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Mol. Biol. Evol.* **2010**, *27*, 1436–1448. [[CrossRef](#)] [[PubMed](#)]
32. Alverson, A.J.; Rice, D.W.; Dickinson, S.; Barry, K.; Palmer, J.D. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. *Plant Cell* **2011**, *23*, 2499–2513. [[CrossRef](#)] [[PubMed](#)]
33. Adams, K.L.; Rosenblueth, M.; Qiu, Y.-L.; Palmer, J.D. Multiple losses and transfers to the nucleus of two mitochondrial succinate dehydrogenase genes during angiosperm evolution. *Genetics* **2001**, *158*, 1289–1300. [[PubMed](#)]
34. Woloszyńska, M. Heteroplasmy and stoichiometric complexity of plant mitochondrial genomes—Though this be madness, yet there’s method in’t. *J. Exp. Bot.* **2010**, *61*, 657–671. [[CrossRef](#)] [[PubMed](#)]
35. Treangen, T.J.; Salzberg, S.L. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* **2012**, *13*, 36. [[CrossRef](#)] [[PubMed](#)]
36. Wei, D.-D.; Shao, R.; Yuan, M.-L.; Dou, W.; Barker, S.C.; Wang, J.-J. The multipartite mitochondrial genome of *Liposcelis bostrychophila*: Insights into the evolution of mitochondrial genomes in bilateral animals. *PLoS ONE* **2012**, *7*, e33973. [[CrossRef](#)] [[PubMed](#)]
37. Sugiyama, Y.; Watase, Y.; Nagase, M.; Makita, N.; Yagura, S.; Hirai, A.; Sugiura, M. The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: Comparative analysis of mitochondrial genomes in higher plants. *Mol. Genet. Genom.* **2005**, *272*, 603–615. [[CrossRef](#)] [[PubMed](#)]
38. Lin, L.; Ni, B.; Lin, H.; Zhang, M.; Li, X.; Yin, X.; Qu, C.; Ni, J. Traditional usages, botany, phytochemistry, pharmacology and toxicology of *Polygonum multiflorum* thunb.: A review. *J. Ethnopharmacol.* **2015**, *159*, 158–183. [[CrossRef](#)] [[PubMed](#)]
39. Sloan, D.B.; Alverson, A.J.; Chuckalovcak, J.P.; Wu, M.; McCauley, D.E.; Palmer, J.D.; Taylor, D.R. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biol.* **2012**, *10*, e1001241. [[CrossRef](#)] [[PubMed](#)]
40. Eddie, W.; Shulkin, T.; Gaskin, J.; Haberle, R.; Jansen, R. Phylogeny of Campanulaceae s. Str. inferred from ITS sequences of nuclear ribosomal DNA. *Ann. Mo. Bot. Gard.* **2003**, *90*, 554–575. [[CrossRef](#)]
41. Wendling, B.M.; Galbreath, K.E.; DeChaine, E.G. Resolving the evolutionary history of *Campanula* (Campanulaceae) in Western North America. *PLoS ONE* **2011**, *6*, e23559. [[CrossRef](#)] [[PubMed](#)]

42. Crawl, A.A.; Mavrodiev, E.; Mansion, G.; Haberle, R.; Pistarino, A.; Kamari, G.; Phitos, D.; Borsch, T.; Cellinese, N. Phylogeny of Campanuloideae (Campanulaceae) with emphasis on the utility of nuclear pentatricopeptide repeat (PPR) genes. *PLoS ONE* **2014**, *9*, e94199. [[CrossRef](#)] [[PubMed](#)]
43. Hong, C.P.; Park, J.; Lee, Y.; Lee, M.; Park, S.G.; Uhm, Y.; Lee, J.; Kim, C.-K. accD nuclear transfer of *Platycodon grandiflorum* and the plastid of early Campanulaceae. *BMC Genom.* **2017**, *18*, 607. [[CrossRef](#)] [[PubMed](#)]
44. Haberle, R.C.; Dang, A.; Lee, T.; Peñaflor, C.; Cortes-Burns, H.; Oestreich, A.; Raubeson, L.; Cellinese, N.; Edwards, E.J.; Kim, S.-T. Taxonomic and biogeographic implications of a phylogenetic analysis of the Campanulaceae based on three chloroplast genes. *Taxon* **2009**, *58*, 715–734.
45. Byng, J.W.; Chase, M.W.; Christenhusz, M.J.; Fay, M.F.; Judd, W.S.; Mabberley, D.J.; Sennikov, A.N.; Soltis, D.E.; Soltis, P.S.; Stevens, P.F. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* **2016**, *181*, 1–20.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).