scientific reports

OPEN

The complete chloroplast genome sequences of five pinnate-leaved *Primula* species and phylogenetic analyses

Wenbin Xu, Boshun Xia & Xinwei Li✉

The six pinnate-leaved species are a very particular group in the genus *Primula*. In the present paper, we sequenced, assembled and annotated the chloroplast genomes of five of them (*P. cicutarrifolia*, *P. hubeiensis*, *P. jiugongshanensis*, *P. merrilliana*, *P. ranunculoides*). The five chloroplast genomes ranged from ~150 to 152 kb, containing 113 genes (four ribosomal RNA genes, 29 tRNA genes and 80 protein-coding genes). The six pinnate-leaved species exhibited synteny of gene order and possessed similar IR boundary regions in chloroplast genomes. The gene *accD* was pseudogenized in *P. filchnerae*. In the chloroplast genomes of the six pinnate-leaved *Primula* species, SSRs, repeating sequences and divergence hotspots were identified; *ycf1* and *trnH-psbA* were the most variable markers among CDSs and noncoding sequences, respectively. Phylogenetic analyses showed that the six *Primula* species were separated into two distant clades: one was formed by *P. filchnerae* and *P. sinensis* and the other clade was consisting of two subclades, one formed by *P. hubeiensis* and *P. ranunculoides*, the other by *P. merrilliana*, *P. cicutarrifolia* and *P. jiugongshanensis*. *P. hubeiensis* was closely related with *P. ranunculoides* and therefore it should be placed into Sect. *Ranunculoides*. *P. cicutarrifolia* did not group first with *P. ranunculoides* but with *P. merrilliana*, although the former two were once united in one species, our results supported the separation of *P. ranunculoides* from *P. cicutarrifolia* as one distinct species.

Primula L. consists of about 430 species (seven subgenera, 38 sections) in the world[1], and there are 300 species (24 sections) in China[2]. Altogether six species have leaves pinnately compound or pinnately lobed to the midvein: *P. cicutarrifolia*, *P. filchnerae*, *P. hubeiensis*, *P. jiugongshanensis*, *P. merrilliana* and *P. ranunculoides*[3–5]. These species are all endangered[3,4,6–9]. The ITS (internal transcribed spacer) phylogeny trees showed that *P. filchnerae* should be placed in Sect. *Auganthus*, and *P. cicutarrifolia*, *P. merrilliana* and *P. jiugongshanensis* belonged in Sect. *Ranunculoides*[4,10,11]. *P. hubeiensis* might attribute to Sect. *Auganthus*[5]. Chloroplast fragments *matK*, *rps16*, and *trnL-F* data also supported *P. filchnerae* to be included in Sect. *Auganthus*[12]. Based on the ITS phylogeny that revealed *P. fichnerae*, *P. cicutarrifolia* and *P. merrilliana* not to cluster into a monophyly Hao et al. (2002)[10] suggested that the character of pinnately lobed or divided leaves had evolved in parallel. However, the phylogenetic relationships among the six pinnate-leaved species were not explored yet.

The chloroplast (cp) genomes possess conserved structure including two copies of an inverted repeat regions (IRs) linking large and small single-copy regions (LSC and SSC)[13]. Due to moderate substitution rate[14], molecular markers derived from cp genomes are widely used in plant population genetics, molecular phylogenetics, evolutionary biology and species identification. The complete cp genomes could provide higher phylogenetic resolution than ITS or selected chloroplast DNA data[15–17], chloroplast genomic data provided strong support for resolution of controversial phylogenetic relationships[18,19]. A few cp genomes were reported for *Primula* species including *P. filchnerae*[20,21]. In the present study, we will release five complete cp genomes of pinnate-leaved *Primula* species, comparing their genome contents and structure, exploring SSRs and repeats, identifying variable regions, in order to facilitate conservation and systematics of the genus *Primula* .

Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China. ✉email: forfortomorrow@163.com

Materials and methods

Genome sequencing, assembly and annotation. We collected six species in China: *P. cicutarrifolia* in Hangzhou, Zhejiang, *P. merrilliana* in Mt. Huang, Anhui, *P. hubeiensis*, *P. jiugongshanensis* and *P. ranunculoides* in Tongshan, Hubei, and *P. filchnerae* in Zhuxi, Hubei.

DNA was isolated from fresh leaves using CTAB method[22]. Paired-end libraries with about 350 bp DNA insertion were prepared using Illumina TruSeq Library preparation kits (Illumina, San Diego, CA, USA) according to the manufacturer's protocol. The libraries were sequenced on the Illumina Hiseq 2500 platform (Illumina Inc.), generating raw data of 150 bp paired-end reads.

The raw data were subjected to quality control using NGS QC Toolkit[23] (cut-off value for PHRED quality score = 30), then the filtered data were imported into CLC Genomics Workbench v. 11.0.1 (https://www.qiagenbioinformatics.com) to generate contigs with the word value of 60. The relative order and orientation of the contigs of the cp genomes of five species (*P. cicutarrifolia*, *P. merrilliana*, *P. hubeiensis*, *P. jiugongshanensis* and *P. ranunculoides*) were determined by BLAST search against the cp genome of *P. sinensis* (NC_030609). The hit contigs were then concatenated into complete sequences with minimum overlap of 31 bp in Geneious 9 (Biomatters, Auckland, New Zealand); gaps between contigs were closed by comparison with the contigs produced by IOGA[24] with *P. sinensis* (NC_030609) as the reference. The filtered data were mapped back onto the newly assembled cp genomes to confirm no assembly errors by the Geneious plugin in Geneious 9 (Biomatters, Auckland, New Zealand). The cp genome of *P. filchnerae*[21] was downloaded from NCBI because we collected the sample of *P. filchnerae* from same place as Sun et al.[21].

The ITS (Internal Transcribed Spacer) sequences of the six species were generated from the consensus of the reads of the quality controlled data mapped onto that of *P. sinensis* (JF978052) by the Geneious plugin in Geneious 9 (Biomatters, Auckland, New Zealand).

The cp genomes in this study were annotated using GeSeq[25], choosing the MPI-MP chloroplast references as the references. The annotations were modified manually by comparing with Primulaceae cp genomes available in the Genbank. The cp genome maps were drawn using OGDRAW[26]. The sequences of the cp genomes were visualized in Geneious 9 (Biomatters, Auckland, New Zealand). The synteny of the cp genomes of six *Primula* species with pinnatisect leaves was estimated with MAUVE 20150226[27].

All alignment was done with MAFFT[28] for further analyses.

Repeat element analysis. Simple sequence repeats (SSR) were detected for the cp genomes using MISA[29]. The minimum numbers for the SSR motifs were 10, 5, 4, 3, 3 and 3 for mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats, respectively. REPuter[30] was used to identify the repeating sequences (forward, reverse, complement and palindrome) with three for Hamming distance, 30 for Minimal Repeat Size.

Sequence divergence analysis. We calculated the nucleotide variability (Pi) values of the protein coding sequences, introns and intergenic spacers of the cp genomes of the six species with pinnatisect leaves and 16 other *Primula* species available in Genbank (Table S1) using DnaSP 6.12[31].

Phylogenetic analysis. We constructed phylogenetic trees by Maximum likelihood (ML) and neighbor-joining (NJ) methods. *Androsace paxiana* and *Lysimachia congestiflora* were treated as outgroups. The ML phylogenetic trees (1000 bootstrap replicates) were inferred with RAxML 8.2.10[32] based on whole cp genomes (Tables S1), *ycf1*, and concatenation of ITS, *matK* and *rbcL* (Tables S2), respectively. *ycf1* was extracted from the cp genomes (Tables S1), because the gene *ycf1* did not exist in *P. tsiangii*, we used its homologous part in the cp genome. NJ analysis of 71 *Primula* species including the six species with pinnatisect leaves based on concatenation of ITS, *matK* and *rbcL* (Tables S2) was carried out using MEGA-X[33] with 1000 bootstrap replicates. ITS of the six pinnate-leaved *Primula* species was newly sequenced in this study; *matK* and *rbcL* were derived from the chloroplast sequences in this study and from the cp genome of *P. filchnerae*[21]. The accessions of the cp genomes and DNA fragments were listed in Tables S1, S2, respectively.

Results

Basic characters of the six chloroplast genomes. The cp genomes of *P. cicutarrifolia*, *P. hubeiensis*, *P. jiugongshanensis*, *P. merrilliana* and *P. ranunculoides* (GenBank accessions: MT268974, MT268976, MT937162, MT268977, MT268978) were reported for the first time here, and that of *P. filchnerae* was downloaded from NCBI (MK888698[21]).

The sequencing coverage of our five newly assembled cp genomes was from 923 to 6237 (Figure S1). The six cp genomes possessed typical quadripartite structure: IRa, IRb, LSC and SSC (Table 1), and they exhibited the same gene order, no gene rearrangement or inversion occurred (Figure S2). The physical map of the cp genome of *P. hubeiensis* was shown in Fig. 1. The GC content was ~ 37%. The newly sequenced genomes ranged from 150,187 bp to 151,972 bp, harboring 113 genes: four ribosomal RNA genes, 29 tRNA genes and 80 protein-coding genes, and among them 14 genes was duplicated in IRa and IRb (Table 1). Due to presence of multiple stop codons, the gene *infA* was pseudogenized in the five newly sequenced species. The open reading frame (ORF) in *accD* of *P. filchnerae* (MK888698) was truncated to be only 1305 bp compared with 1455 or 1464 bp ORF of other five species. Lee et al.[34] identified five conserved amino acid sequence motifs in *accD* gene. Conserved amino acid sequence motifs IV and V were absent in *accD* of *P. filchnerae*. Therefore, *accD* was nonfunctional in *P. filchnerae*.

SSRs and repeats. Five categories of SSRs were identified for the six species (Table 2). The least number of SSRs was 41 for *P. ranunculoides* and the most 59 for *P. merrilliana*. Three types of SSRs were detected for

Species	Pc	Pf	Ph	Pj	Pm	Pr
Total length	151,972 bp	151,547 bp	151,759 bp	151,696 bp	151,843 bp	150,187 bp
GC%	36.8%	37.2%	36.8%	36.8%	36.8%	36.8%
LSC	83,945 bp	82,662 bp	83,523 bp	83,797 bp	83,847 bp	82,031 bp
SSC	17,839 bp	17,749 bp	17,632 bp	17,521 bp	17,554 bp	17,572 bp
IR	25,094 bp	25,568 bp	25,302 bp	25,189 bp	25,221 bp	25,292 bp
Total genes	113	112	113	113	113	113
Protein genes	80	79	80	80	80	80
rRNA genes	4	4	4	4	4	4
tRNA genes	29	29	29	29	29	29

Table 1. Basic characteristics of cp genomes of the six *Primula* species (Pc: *P. cicutarrifolia*; Pf: *P. filchnerae*; Ph: *P. hubeiensis*; Pj: *P. jiugongshanensis*; Pm: *P. merrilliana*; Pr: *P. ranunculoides*).

P. filchnerae, and in the rest species four types could be found. While mono-, di- and tetra-nucleotide repeats existed across all the six species, tri- and penta-inucleotide repeats resided in three and two species respectively. Mono- and dinucleotide repeats accounted for the vast majority of SSRs (65.1% for *P. cicutariifolia*, 87.5% for *P. filchnerae*, 69.0% for *P. hubeiensis*, 62.8% for *P. jiugongshanensis*, 72.9% for *P. merrilliana*, 73.2% for *P. ranunculoides*). Most or all mono- repeats were A/T repeats including 10 to 16 nucleotides. The number of repeat units ranged from five to eight for dinucleotide repeats. The tri- and penta-nucleotide SSRs consisted of four motifs, and tetra-nucleotide SSRs of four to five repeat units.

Except the largest repeat for each genome (i.e. IRs), a total of 183 repeat pairs (three types: forward (F), reverse (R), and palindromic repeats (P)) were detected in the six genomes (Fig. 2), which ranged from 30 to 137 bp in length. Palindromic repeats were the most common, accounting for 55.2% (101 of 183), followed by forward repeats (44.3%, 81 of 183). No complement repeats were identified in all species and one pair of reverse repeats existed specifically in *P. ranunculoides*. In the six species, 96.7% (177 of 183 repeat pairs) repeats were 30–59 bp in length, consistent with the length reported in other *Primula* species[20]. The longest repeat (137 bp) was found in *P. cicutariifolia*, and this species contained the most repeats (44 pairs), while *P. jiugongshanensis* had the least (24 pairs).

IR/SC boundary. The IR/SC boundary regions of the six *Primula* cp genomes were compared, and the IR/SC junction regions showed slight differences in the length of organization genes flanking the junctions or the distance between the junctions and the organization genes (Fig. 3). The genes spanning or flanking the junction of LSC/IRb, IRb/SSC, SSC/IRa and IRa/LSC were *rps19/rpl2*, *ndhF*, *ycf1*, *rpl2/trnH*, respectively. IR expansion and contraction was observed. *P. cicutarrifolia* had the smallest size of IR but largest size of both LSC and SSC; though largest size of IR was detected in *P. filchnerae*, the LSC or SSC was not the smallest in this species. The gene *trnH* was located in LSC, 0–24 bp away from the IRa/LSC border. The largest extensions of *ycf1* into both SSC and IRa occurred in *P. filchnerae* (4566 bp and 1023 bp, respectively) and *ycf1* of *P. filchnerae* were the longest among the six species. The gene *ndhF* was utterly situated in SSC and 108 bp distant from the IRb/SSC junction in *P. cicutarrifolia*; in the rest five species the fragment size of *ndhF* in SSC was largest in *P. hubeiensis* (2194 bp). In *P. cicutarrifolia*, *P. jiugongshanensis* and *P. merrilliana*, *rps19* and *rpl2* were located in the upstream and downstream of the LSC/IRb junction, respectively; *rps19* ran across the LSC/IRb junction in *P. filchnerae*, *P. hubeiensis*, *P. ranunculoides* with 161, 62, 56 bp extension in IRb, respectively.

Divergent hotspots in the *Primula* chloroplast genome. As indicated by the value of Pi, the nucleotide variability of the 22 *Primula* species (Table S1) was evaluated by DnaSP 6.12[31] using noncoding sequences (intron and intergenic spacer) or protein coding sequences (CDS) at least 200 bp long. The variation level of DNA polymorphism was 0.00444–0.11369 for noncoding sequences or 0.00094–0.05036 for CDSs. For the CDSs, the highest Pi value were detected for *ycf1* (0.05036), followed by *matK* (0.04878), *rpl22* (0.04364), *ndhF* (0.03975), *rps8* (0.03658), *ndhD* (0.03455), *ccsA* (0.03292), *rpl33* (0.0303), *rps15* (0.03022), and *rpoC2* (0.02954). These markers had higher Pi than *rbcL* (0.02149). Obviously, the gene *ycf1* exhibited the greatest diversity and harbored the most abundant variation. The ten most divergent regions among noncoding regons included *trnH* (GUG)-*psbA* (0.11369), *trnW* (CCA)-*trnP* (UGG) (0.09463), *rpl32*-*trnL* (UAG) (0.09337), *ndhC*-*trnV* (UAC) (0.09148), *ccsA*-*ndhD* (0.08745), *ndhG*-*ndhI* (0.08363), *trnK* (UUU)-*rps16* (0.08334), *trnM* (CAU)-*atpE* (0.08273), *trnS* (GGA)-*rps4* (0.08028), and *trnC* (GCA)-*petN* (0.07971). No intron ranked among the top ten variable noncoding regions.

Phylogenetic analysis. The ML tree of 22 *Primula* species was constructed with RAxML[32] (Fig. 4), based on the whole cp genomes. The six pinnate-leaved *Primula* species did not form a monophyly, but separated into two distant clades. *P. filchnerae* grouped with *P. sinensis*, the other five species clustered together and constituted the clade Sect. *Ranunculoides* with 100% bootstrap. In the ML tree, Sect. *Proliferae* exhibited monophyly, while species of Sect. *Crystallophlomis* separated into different clades.

The topology of the ML tree based on *ycf1* (Figure S3) was consistent with that based on whole cp genomes (Fig. 4), except that the clade formed by *P. veris* and *P. knuthiana* were sister to the clade consisting of Sects.

Figure 1. Physical map of the *P. hubeiensis* chloroplast genome.

Auganthus, Obconicolisteri, Carolinella and *Monocarpicae* instead of being sister to the clade of Sects. *Proliferae, Ranunculoides* and *Crystallophlomis*.

We also constructed both ML and NJ tree of 71 *Primula* species based on the concatenation of three common barcoding markers (ITS, *matK* and *rbcL*). Only the results of NJ analysis (Fig. 5) showed consistency with those of Yan et al.[12], Liu et al.[35], and ML analysis based on whole cp genomes (Fig. 4). The six pinnate-leaved *Primula* species were separated into two distantly related groups. The clade consisting of *P. filchnerae* and *P. sinensis* (Sect. *Auganthus*) was sister to the clade formed by Sects. *Carolinella, Obconicolisteri, Monocarpicae, Cortusoides, Malvacea, Pycnoloba*. The five pinnatisect-leaved species *P. cicutarrifolia, P. hubeiensis, P. jiugonshanensis, P. merrilliana* and *P. ranunculoides* (Sect. *Ranunculoides*) comprised a 100% supported clade, which was sister to the group containing Sects. *Crystallophlomis, Petiolares, Proliferae, Amethystina*. Sect. *Carolinella* and Sect. *Crystallophlomis,* and Sect. *Malvacea* were polyphyletic.

Type	Repeat unit	Pc (43)	Pf (56)	Ph (42)	Pj (43)	Pm (59)	Pr (41)
Mono	A/T	27	49	29	27	42	29
	C/G	1	–	–	–	1	1
Di	AT/AT	8	5	9	9	8	8
Tri	AAG/CTT	2	–	–	2	2	–
Tetra	AAAT/ATTT	2	2	2	2	4	1
	AACG/CGTT	2	–	–	2	2	–
	AATT/AATT	1	–	–	–	–	–
	AGAT/ATCT	–	–	–	1	–	–
	AAAG/CTTT	–	–	–	–	–	1
Penta	AATGT/ACATT	–	–	1	–	–	1
	AAAAT/ATTTT	–	–	1	–	–	–

Table 2. Types and numbers of SSRs in the cp genomes of six *Primula* species, the numbers in the bracket indicating total number of SSRs (Pc: *P. cicutarrifolia*; Pf: *P. filchnerae*; Ph: *P. hubeiensis*; Pj: *P. jiugongshanensis*; Pm: *P. merrilliana*; Pr: *P. ranunculoides*).
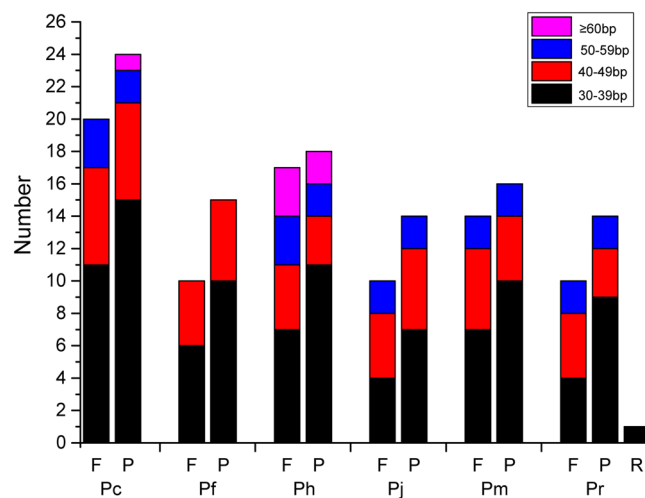
Figure 2. Types and numbers of repeat pairs in the cp genomes of six *Primula* species (Pc: *P. cicutarrifolia*; Pf: *P. filchnerae*; Ph: *P. hubeiensis*; Pj: *P. jiugongshanensis*; Pm: *P. merrilliana*; Pr: *P. ranunculoides*).

Discussion

The six cp genomes of pinnate-leaved species were ~ 150–152 kb with similar GC content. The gene content and organization were similar and a high degree of synteny in gene order was observed across all the genomes. The gene *accD* was normal in five species but perhaps pseudogenized duo to lack of two conserved amino acid sequence motifs in *P. filchnerae*. In *P. sinensis*, this gene was pseudogenized and another copy of *accD* were detected in the nucleus[36]. Whether *accD* was functionally transferred to the nucleus in *P. filchnerae* needs further confirmation. Interestingly, *P. filchnerae* and *P. sinensis* always grouped together on the phylogenetic trees (Figs. 4, 5).

In the six *Primula* species, the IR/SC boundary regions exhibited similar feature, with slight differences observed in the length of organization genes flanking the junctions or the distance between the junctions and the organization genes (Fig. 3), and the situation is similar to ten other *Primula* species[20], which indicates the structural conservation of *Primula*. Expansion of IR regions may cause size increase in chloroplast genomes[37], however, it seems that the size of whole cp genomes did not always increase with expansion of IR in *Primula*. For example, among the six pinnate-leaved species, *P. cicutarrifolia* possessed the smallest IR (25,094 bp) but the largest whole genome size (151,972 bp); in *P. filchnerae*, the IR was the longest (25,568 bp), and the whole genome size (151,547 bp) was only bigger than *P. ranunculoides* (150,187 bp). In *P. kwangtungensis*[20], both IR (25,855 bp) and the whole genome size (153,757 bp) exceeded all other species (Table S1) including the six pinnate-leaved species.
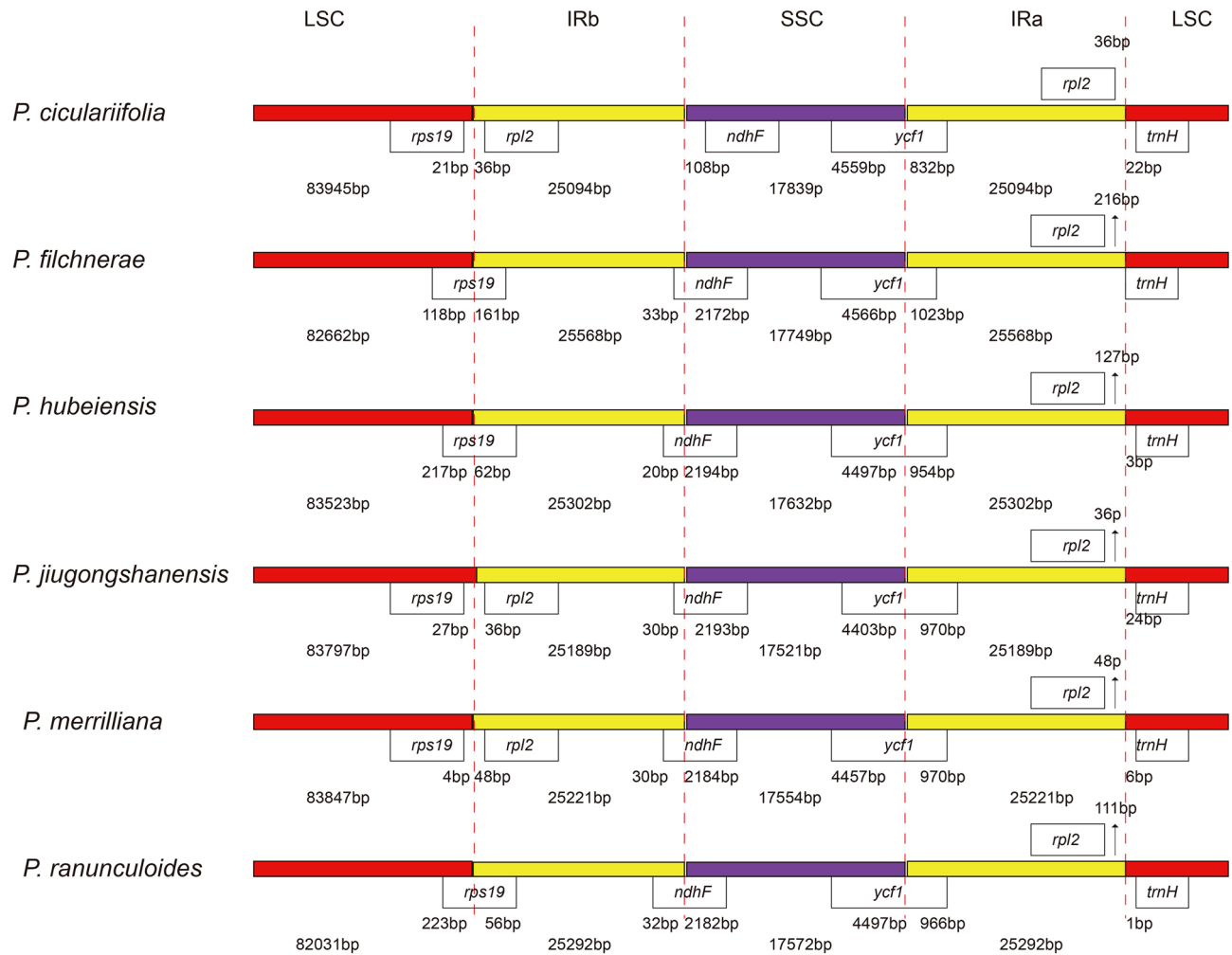
**Figure 3.** LSC/IR, and SSC/IR border regions of the six *Primula* cp genomes.

Except the IRs, 183 pairs of repeats were detected in the six cp genomes, only one which were longer than 70 bp (137 bp), which is similar in ten other *Primula* species, most of repeats ranged in size from 14 to 62 bp and all (except one pair of 111 bp repeat) were not large repeats (> 100 bp)[20]. No rearrangement was found in our six species, the reason may be lack of large complex repeating sequences (> 100 bp) just as suggested by Ren et al.[20]. The SSR marker analyses have been proven to be powerful to assess the genetic diversity and population structure of *P. cicutarrifolia*, *P. merrilliana* and *P. sikkimensis*[7,38,39]. The usefulness of the SSRs located in the six chloroplast genomes may be tried in future studies on population genetics of *Primula* species.

Using the six pinnate-leaved species cp genomes and 16 other *Primula* cp genomes available in NCBI, the divergence hotspots were identified among CDSs and noncoding regions. The nucleotide diversity (Pi) of *ycf1* and *matK* reached 0.05036 and 0.04878, respectively, much higher than *rbcL* (0.02149), which was a common barcode for species identification. The gene *ycf1* was considered to be the most promising barcode to identify plant species[40]. Two chloroplast genes, *ycf1* and *psbM-psbD*, had much better discriminatory power (both 87.5%) than did other chloroplast barcodes for identifying *Fritillaria* species[41]. The ML species tree based on *ycf1* (Figure S3) showed similar topology to that based on whole cp genome. Except *matk* CDS, other hotspots regions identified here were not tested for species identification or phylogeny reconstruction[42,43]. Among the noncoding sequences, *trnH* (GUG) *-psbA* was the most variable one, which showed better discriminatory power than *matK* and *rbcL*[43]. These highly variable regions have the potential to be used for *Primula* species discrimination or phylogeny investigation in future study.
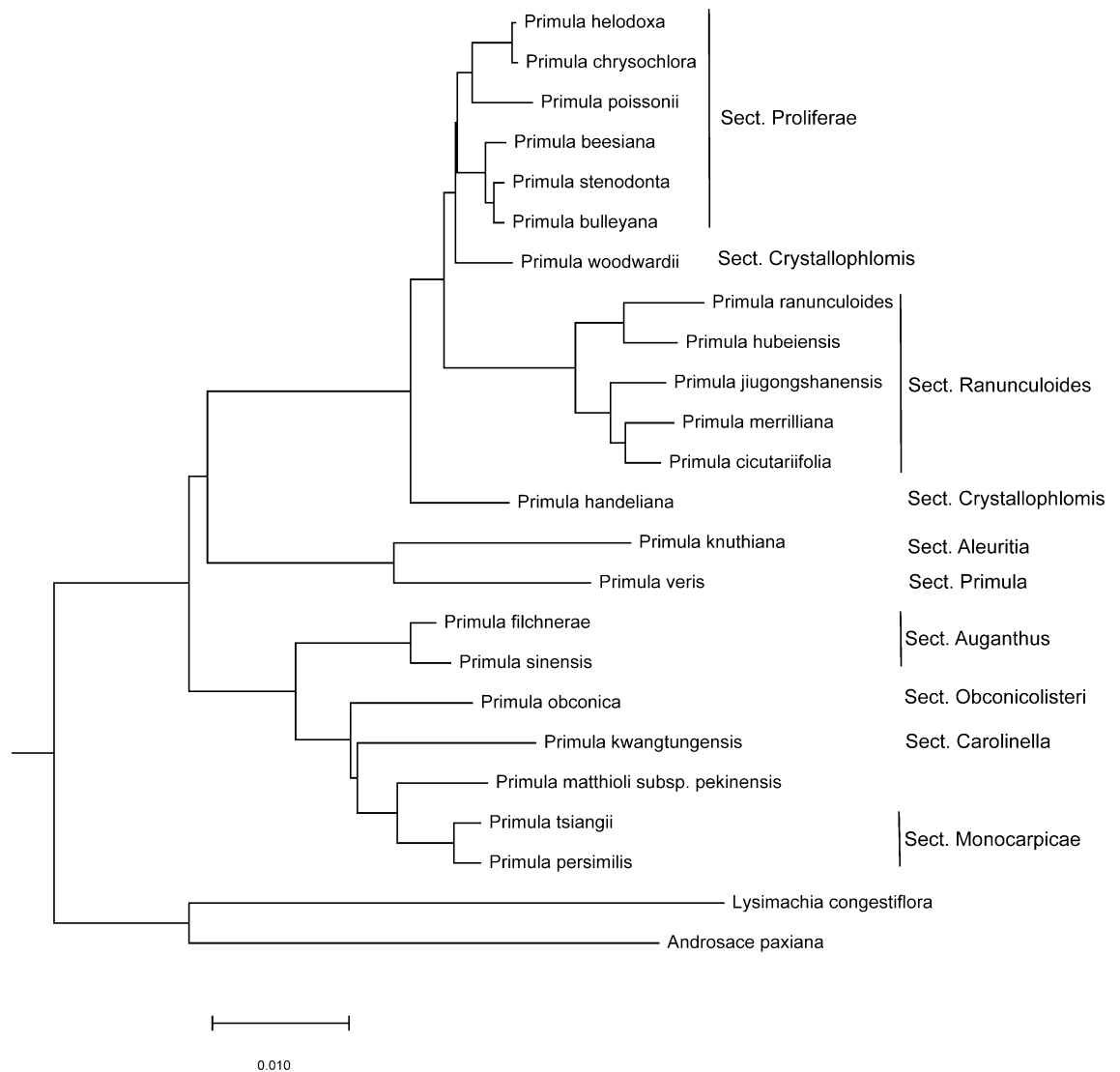
Figure 4. ML phylogenetic tree of *Primula* species based on cp genomes. Bootstrap support at nodes are all 100%.

Both ML and NJ phylogenetic analyses revealed that the six pinnate-leaved *Primula* species did not form a monophyletic group, probably due to parallel evolution of pinnately lobed or divided leaves[10]. In the ML and NJ trees, the phylogenetic placement of the clade consisting of *P. filchnerae* and *P. sinensis* was near to Sect. *Carolinella* and Sect. *Obconicolisteri*, which is similar to the results of Yan et al.[12]. Liu et al.[35] proposed Subgen. *Auganthus* (Sect. *Auganthus*, *Bullatae*, *Cortusoides*, *Dryadifolia*, *Malvacea*, *Monocarpicae*, *Obconicolisteri*, *Pycnoloba*) to include Subgen. *Carolinella* (Sects. *Carolinella*) and exclude *P. aromatica*, *P. filchnerae* and *P. sinensis* thus were in the basal clade of Subgen. *Auganthus*. The close relatedness of *P. filchnerae* and *P. sinensis* was also indicated by the pseudogenization of the gene *accD*. And our study showed that Sect. *Ranunculoides* (*P. cicutarrifolia*, *P. hubeiensis*, *P. jiugongshanensis*, *P. merrilliana*, *P. ranunculoides*) was closely related to Sects. *Crystallophlomis*. Li et al.[5] conjectured that *P. hubeiensis* resembled *P. filchnerae*, and might belong in Sect. *Auganthus*. However, the present study clearly indicated that *P. hubeiensis* grouped with *P. ranunculoides* first, *P. cicutarrifolia* grouped with *P. merrilliana* first, and therefore *P. hubeiensis* should be placed in Sect. *Ranunculoides*. He et al.[4] also demonstrated that *P. cicutarrifolia* was closely related with *P. merrilliana*. *P. ranunculoides* and *P. cicutarrifolia* were united into one species[2] but later separated as two species[3], and our ML and NJ analyses strongly supported the taxonomic treatment of Shao et al.[3].
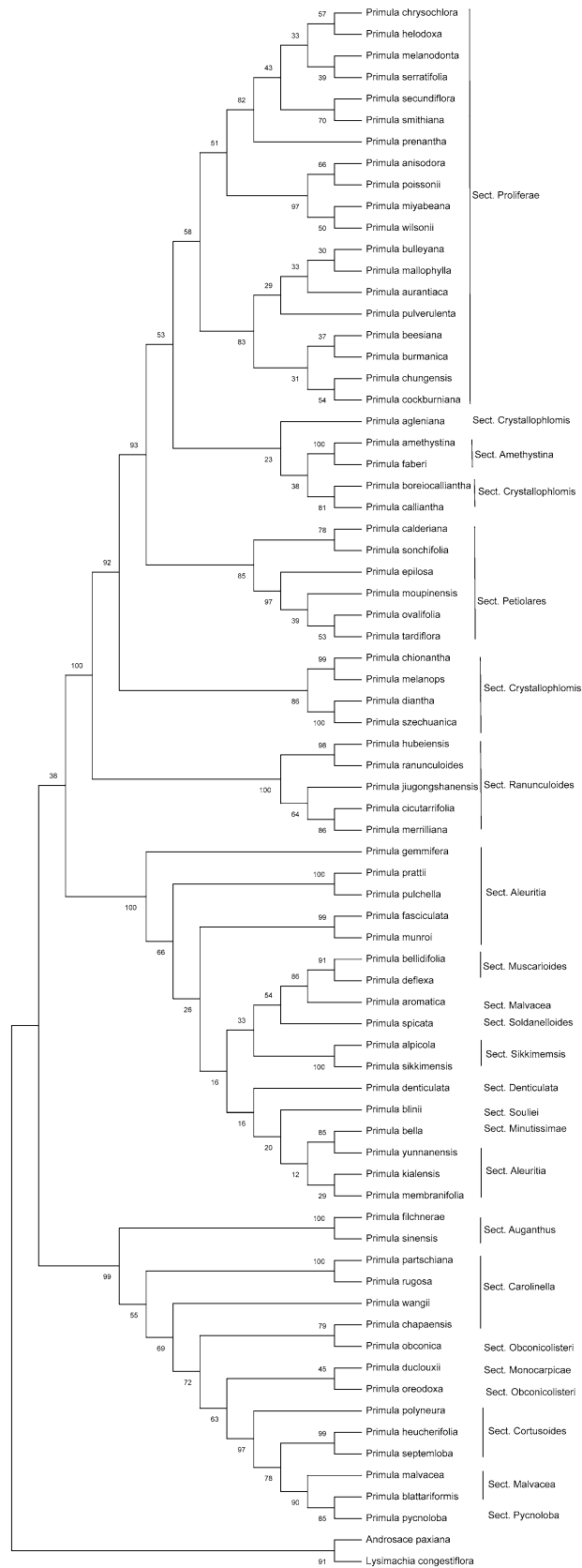
Figure 5. NJ bootstrap consensus tree of *Primula* based on concatenation of ITS, *matK* and *rbcL*.

Data availability

References

1. Richards, J. *Primula* 2nd edn. (Timber Press, Portland, 2003).
2. Hu, C.M. & Kelso, S. Primulaceae. 39–189. in *Flora of China* (Wu, Z.Y. & Raven, P.H. (Ed.) Vol. 15 (Myrsinaceae through Loganiaceae). (Science Press, Beijing/Missouri Botanical Garden Press, St. Louis, 1996).
3. Shao, J. W. *et al.* Reappraisal of *Primula ranunculoides* (Primulaceae), an endangered species endemic to China, based on morphological, molecular genetic and reproductive characters. *Bot. J. Linn. Soc.* **169**, 338–349. https://doi.org/10.1111/j.1095-8339.2012.01228.x (2012).
4. He, X. *et al. Primula jiugongshanensis* sp. nov. (Primulaceae) from China, based on morphological and molecular evidence. *Nord. J. Bot.* **35**, 328–333, https://doi.org/10.1111/njb.01471 (2017).
5. Li, X. W., Bao, D. C., Huang, H. D. & Xie, J. F. *Primula hubeiensis* (Primulaceae), a new species from Central China. *Novon* **25**, 162–165. https://doi.org/10.3417/2016032 (2017).
6. Shao, J. W., Zhang, X. P., Zhang, Z. X. & Zhu, G. P. Effects of population size on reproductive success of the endangered and endemic species *Primula merrilliana*. *J. Integr. Plant Biol.* **50**, 1151–1160. https://doi.org/10.1111/j.1744-7909.2008.00655.x (2008).
7. Wang, D. Y. *et al.* Highly differentiated populations of the narrow endemic and endangered species *Primula cicutariifolia* in China, revealed by ISSR and SSR. *Biochem. Syst. Ecol.* **53**, 59–68. https://doi.org/10.1016/j.bse.2013.12.025 (2014).
8. Gan, Q. L. & Li, X. W. Neotypification of *Primula filchnerae* (Primulaceae). *Novon.* **24**, 155–158. https://doi.org/10.3417/2012075 (2015).
9. Jiang, M. X. *Rare and Endangered Plants in Hubei* (Hubei Science and Technology Press, Wuhan, 2019).
10. Hao, G., Hu, C.M. & Lee, N.S. Circumscriptions and phylogenetic relationships of *Primula* sects. *Auganthus* and *Ranunculoides*: Evidence from nrDNA ITS sequences. *Acta Bot. Sin.* **44**, 72–75 (2002).
11. Xu, C. Y. *et al.* Pollen morphological variation of *Primula merrilliana* and its systematic significance. *Flora* **253**, 43–48. https://doi.org/10.1016/j.flora.2019.03.010 (2019).
12. Yan, H. F. *et al.* Circumscription of *Primula* subgenus *Auganthus* (Primulaceae) based on chloroplast DNA sequences. *J. Syst. Evol.* **48**, 123–132. https://doi.org/10.1111/j.1759-6831.2010.00068.x (2010).
13. Daniell, H., Lin, C. S., Yu, M. & Chang, W. J. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol.* **17**, 134. https://doi.org/10.1186/s13059-016-1004-2 (2016).
14. Drouin, G., Daoud, H. & Xia, J. N. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol. Phylogenet. Evol.* **49**, 827–831. https://doi.org/10.1016/j.ympev.2008.09.009 (2008).
15. Sanz, M. *et al.* Molecular phylogeny and evolution of floral characters of Artemisia and allies (anthemideae, Asteraceae): Evidence from nrDNA ETS and ITS sequences. *Taxon* **57**, 66–78 (2008).
16. Riggins, C. & Seigler, D. The genus *Artemisia* (Asteraceae: Anthemideae) at a continental crossroads: Molecular insights into migrations, disjunctions, and reticulations among old and new world species from a Beringian perspective. *Mol. Phylogenet. Evol.* **64**, 471–490. https://doi.org/10.1016/j.ympev.2012.05.003 (2012).
17. Nikiforova, S. V., Cavalieri, D., Velasco, R. & Goremykin, V. Phylogenetic analysis of 47 chloroplast genomes clarifies the contribution of wild species to the domesticated apple maternal line. *Mol. Biol. Evol.* **30**, 1751–1760. https://doi.org/10.1093/molbev/mst092 (2013).
18. De Abreu, N. L. *et al.* The use of chloroplast genome sequences to solve phylogenetic incongruences in Polystachya Hook (Orchidaceae Juss). *PeerJ* **6**, e4916. https://doi.org/10.7717/peerj.4916 (2018).
19. Zhang, R. *et al.* Exploration of plastid phylogenomic conflict yields new insights into the deep relationships of Leguminosae. *Syst. Biol.* **69**, 613–622. https://doi.org/10.1093/sysbio/syaa013 (2020).
20. Ren, T., Yang, Y. C., Zhou, T. & Liu, Z. L. Comparative plastid genomes of *Primula* species: Sequence divergence and phylogenetic relationships. *Int. J. Mol. Sci.* **19**, 1050. https://doi.org/10.3390/ijms19041050 (2018).
21. Sun, H. Y. *et al.* The complete chloroplast genome of an endangered endemic herb species in China, *Primula filchnerae* (Primulaceae). *Mitochondrial DNA Part B* **4**, 2746–2747. https://doi.org/10.1080/23802359.2019.1644221 (2019).
22. Doyle, J.J. & Doyle, J.L.A rapid DNA isolation procedure for small quantities of leaf tissue. *Phytochem. Bull.* **19**, 11–15 (1987).
23. Patel, R. K. & Jain, M. NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. *PLoS ONE* **7**, e30619. https://doi.org/10.1371/journal.pone.0030619((2012).
24. Bakker, F. T. *et al.* Herbarium genomics: Plastome sequence assembly from a range of herbarium specimens using an iterative organelle genome assembly pipeline. *Biol. J. Lin. Soc.* **117**, 33–43. https://doi.org/10.1111/bij.12642 (2016).
25. Tillich, M. *et al.* GeSeq – Versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **45**, W6–W11. https://doi.org/10.1093/nar/gkx391 (2017).
26. Greiner, S., Lehwark, P. & Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **47**, W59–W64, https://doi.org/10.1093/nar/gkz238 (2019).
27. Darling, A. C. E., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394–1403. https://doi.org/10.1101/gr.2289704 (2004).
28. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. https://doi.org/10.1093/molbev/mst010 (2013).
29. Tiel, T., Michalek, W., Varshney, R. & Graner, A. Exploiting EST databases for the development and characterization of genederived SSR-markers in barley (*Hordeum vulgare* L.). *Theoret. Appl. Genet.* **106**, 411–422, https://doi.org/10.1007/s00122-002-1031-0 (2003).
30. Kurtz, S. *et al.* REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **29**, 4633–4642. https://doi.org/10.1093/nar/29.22.4633 (2001).
31. Rozas, J. *et al.* DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **34**, 3299–3302. https://doi.org/10.1093/molbev/msx248 (2017).
32. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033 (2014).
33. Kumar, S. *et al.* MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evolut.* **35**, 1547–1549, https://doi.org/10.1093/molbev/msy096 (2018)
34. Lee, S. S. *et al.* Characterization of the plastidencoded carboxyltransferase subunit (accD) gene of potato. *Mol. Cells* **17**, 422–429 (2004).
35. Liu, Y. J., Liu, J., Hu, C. M. & Hao, G. Non-monophyly of *Primula* subgenera *Auganthus* and *Carolinella* (Primlaceae) as confirmed by the nuclear DNA sequence variation. *Plant Syst. Evol.* **301**, 2057–2071 (2015).

36. Liu, T. J. *et al.* Complete plastid genome sequence of *Primula sinensis* (Primulaceae): structure comparison, sequence variation and evidence for *accD* transfer to nucleus. *PeerJ* **4**, e2101. https://doi.org/10.7717/peerj.2101 (2016).
37. Asaf, S. *et al.* Expanded inverted repeat region with large scale inversion in the first complete plastid genome sequence of *Plantago ovata*. *Sci. Rep.* **10**, 3881. https://doi.org/10.1038/s41598-020-60803-y (2020).
38. Wang, F. Y. *et al.* Strong genetic differentiation of *Primula sikkimensis* in the east Himalaya-Hengduan mountains. *Biochem. Genet.* **46**, 75–87. https://doi.org/10.1007/s10528-007-9131-9 (2008).
39. Peng, Y. Q. *et al.* Isolation and characterization of fifteen polymorphic microsatellite loci from *Primula merrilliana* (Primulaceae), an endemic from China. *Conserv. Genet.* **10**, 1441–1443. https://doi.org/10.1007/s10592-008-9756-1 (2009).
40. Dong, W. P. *et al.* ycf1, the most promising plastid DNA barcode of land plants. *Sci. Rep.* **5**, 8348. https://doi.org/10.1038/srep08348 (2015).
41. Chen, Q., Wu, X. & Zhang, D. Comparison of the abilities of universal, super, and specific DNA barcodes to discriminate among the original species of Fritillariae cirrhosae bulbus and its adulterants. *PLoS ONE* **15**, e0229181. https://doi.org/10.1371/journal.pone.0229181 (2020).
42. Yan, H.F., Hao, G., Hu, C.M. & Ge, X.J. DNA barcoding in closely related species: A case study of *Primula* L. sect. *Proliferae* Pax (Primulaceae) in China. *J. Syst. Evolut.* **49**, 225–236, https://doi.org/10.1111/j.1759-6831.2011.00115.x (2011).
43. Yan, H.F. *et al.* DNA barcoding evaluation and its taxonomic implications in the species-rich genus *Primula* L. in China. *PLoS One* **10**, e0122903. https://doi.org/10.1371/journal.pone.0122903 (2015).

Author contributions

X.-W.B. and X.-B.S. designed the experiment, analyzed the data and drafted and revised the manuscript. L.-X.W. conceived, designed research and revised the manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-77661-3.

Correspondence and requests for materials should be addressed to X.L.

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

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