



OPEN

Phylogenetic analysis based on single-copy orthologous proteins in highly variable chloroplast genomes of *Corydalis*

Xianmei Yin^{1,4}, Feng Huang^{1,4}, Xiaofen Liu¹, Jiachen Guo¹, Ning Cui², Conglian Liang³, Yan Lian¹, Jingjing Deng¹, Hao Wu², Hongxiang Yin¹✉ & Guihua Jiang¹✉

Corydalis is one of the few lineages that have been reported to have extensive large-scale chloroplast genome (cp-genome) rearrangements. In this study, novel cp-genome rearrangements of *Corydalis pinnata*, *C. mucronate*, and *C. shearerii* are described. *C. pinnata* is a narrow endemic species only distributed at Qingcheng Mountain in southwest China. Two independent relocations of the same four genes (*trnM-CAU-rbcL*) were found relocated from the typically posterior part of the large single-copy region to the front of it. A uniform inversion of an 11–14-kb segment (*ndhB-trnR-ACG*) was found in the inverted repeat region; and extensive losses of *accD*, *clpP*, and *trnV-UAC* genes were detected in all cp-genomes of all three species of *Corydalis*. In addition, a phylogenetic tree was reconstructed based on 31 single-copy orthologous proteins in 27 cp-genomes. This study provides insights into the evolution of cp-genomes throughout the genus *Corydalis* and also provides a reference for further studies on the taxonomy, identification, phylogeny, and genetic transformation of other lineages with extensive rearrangements in cp-genomes.

Corydalis DC. is a large and diverse genus, with ~786 species, within the family Papaveraceae (<http://www.worldfloraonline.org/downloadData> [accessed 9 December 2021]). Plants belonging to the genus *Corydalis* are distributed in the Hengduan Mountains and Qinghai–Tibet Plateau and adjacent areas¹. The structures of the some recognized *Corydalis* chloroplast genomes (cp-genomes) have undergone a series of genetic rearrangements, such as pseudogenization or the loss of genes, to adapt to drastic changes in the environment^{1–4}. *Corydalis pinnata* is a narrow endemic species in China and is only distributed along the streams of Qingcheng Mountain in southwest China at altitudes between 1300 m a.s.l. and 1400 m a.s.l. Consequently, this species must also have undergone a unique genetic shift.

Most of the *Corydalis* plants have potential as medicinal agents due to their therapeutic effects against hepatitis, tumors, cardiovascular diseases, and pain^{5,6}, but some species are toxic⁷. As one of the most taxonomically challenging plant taxa, the genus *Corydalis* has extremely complex morphological variations because of typical reticulate evolution and intense differentiation during evolution⁸, which has hampered understanding of the identification, taxonomy, and utilization of members of this genus.

Chloroplasts are common organelles with an essential role in the photosynthesis of green plants⁹. The cp-genome is an ideal research model for studying molecular identification, phylogeny, species conservation, and genome evolution because of its conservative structure^{10,11}. The increasingly wide application of the cp-genome super-barcode in identification make the development of new cp-genome resources urgent and significant^{12,13}. Cp-genome rearrangements can also be useful as a phylogenetic marker because they lack homoplasy and are easily identified^{14–16}. Although some genetic rearrangements of *Corydalis* cp-genomes have been reported^{1,2}, the pattern, origin, evolution, and phylogenetic relationship of cp-genome rearrangements in *Corydalis* remain unclear because of a lack of sufficient genetic resources. In the present study, three species of the genus *Corydalis* from Qingcheng Mountain, including a narrow endemic species, were identified based on their cp-genomes. In addition, 12 *Corydalis* cp-genomes from the National Centre for Biotechnology Information (NCBI) database were included in the rearrangement analysis to represent all five subgenera of *Corydalis* and cover most of the

¹College of Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu 611130, China. ²Central Laboratory, Shandong Academy of Chinese Medicine, Jinan 250014, China. ³College of Pharmacy, Shandong University of Traditional Chinese Medicine, Jinan 250355, China. ⁴These authors contributed equally: Xianmei Yin and Feng Huang. ✉email: hongxiangyin@126.com; 11469413@qq.com

	Total reads	Total base	Q30	GC	Genome size(bp)	LSC	IR	SSC	Coding gene	tRNA	rRNA	Pseudogene	Assession number
<i>C. pinnata</i>	39,260,005	11.72	88.3	40.47%	158,399	87,573	25,209	20,408	83	28	4	4	OK647837
<i>C. mucronata</i> Franch	34,882,414	10.42	88.11	40.44%	159,803	88,925	23,778	23,322	82	29	4	5	OK647838
<i>C. sheareri</i> S. Moore	40,305,430	12.07	86.86	39.60%	161,105	90,438	23,788	23,091	80	30	4	6	OK647839

Table 1. Summary of genome structure, and gene content of the three newly sequenced *Corydalis* cp-genomes.

distribution areas. The structural characteristics, repeat sequences, and cp-genome rearrangements were documented, and phylogenetic trees based on single-copy orthologous proteins were analyzed. The aim of the study was to assess structural variation and provide valuable resources for identification and classification of members of the genus *Corydalis*.

Results

DNA features of three *Corydalis* cp-genomes. Cp-genomes of three species of the genus *Corydalis* were sequenced; the three species were *Corydalis pinnata*, *C. mucronata*, and *C. sheareri*. The sizes of the three newly sequenced *Corydalis* cp-genomes ranged from 158,399 bp (*C. pinnata*) to 161,105 bp (*C. sheareri*) (Table 1). The guanine+cytosine (G+C) contents of the three genomes were 39.6%–40.47%. The three species each had a cp-genome with typical angiosperm quadripartite structure: a large single-copy (LSC) region, a small single-copy (SSC) region, and a pair of inverted repeats (IRs: IRA and IRB). The lengths of the LSC, SSC, and IR regions of the three newly sequenced *Corydalis* cp-genomes were 87,573–90,438, 20,408–23,322 and 23,778–25,209 bp, respectively (Table 1). After annotation, the sequences of the whole cp-genome sequences of the three *Corydalis* plants were submitted to the NCBI database; GenBank accession numbers are supplied in Table 1. The *C. pinnata* cp-genome was taken as an example and a physical map of the cp-genome was created according to the annotation results using OrganellarGenomeDRAW (OGDraw)¹⁷ (Fig. 1). A total of 115–117 unique genes, comprising 80–83 protein-coding genes, 28–30 tRNA genes, 4 rRNA genes, and 4–6 pseudogenes, were present in the three newly sequenced *Corydalis* cp-genomes (Table 1). In total, seven genes were pseudogenized in one or more *Corydalis* cp-genomes, and three genes (*accD*, *clpP*, and *trnV-UAC*) were lost in the three newly sequenced *Corydalis* cp-genomes (Supplementary Table 1).

In contrast to previously reported *Corydalis* cp-genomes¹, the three newly sequenced *Corydalis* cp-genomes in this study had 11 complete *ndh* genes (Supplementary Table 1). In addition, amongst all the anticipated genes of the three *Corydalis* cp-genomes, introns were discovered in 11–13 genes, including 4–5 tRNA genes and 7–8 protein-coding genes (Supplementary Table 2). The tRNA genes with introns were *trnL-UAA*, *trnK-UUU*, *trnI-GAU*, *trnG-UCC*, and *trnA-UGC*. The eight protein-coding genes with introns were *rps12*, *rpoC1*, *rpl2*, *pafl*, *ndhB*, *ndhA*, *atpF*, and *ycf2*. Two of the 13 intron-containing genes had two introns (*rps12* and *pafl*); the remainder of the genes contained only one intron. The *trnH-UUU* gene contained the largest intron (2474–2488 bp), which contained the whole *matK* gene. Similar to other angiosperms, the gene *rpl2* in the three *Corydalis* cp-genomes resulted from trans-splicing activity. The 5' end of *rpl2* lay in the LSC region, and the 3' end was located in the IR region (Supplementary Table 2).

Chloroplast genome structure rearrangement. Seventeen cp-genomes were included in the synteny comparisons by Mauve alignment (Fig. 2), including 15 *Corydalis* cp-genomes, a representative of Papaveroideae cp-genomes (*Macleaya microcarpa*), and a sister in the Ranunculales cp-genomes (*Euptelea pleiosperma*) to represent a typical angiosperm quadripartite cp-genome structure. More than 30 locally collinear blocks (LCBs) were identified in the *Corydalis* cp-genomes, from which 15 rearrangements were deduced (Fig. 2).

A total of 16 relocation blocks were identified in the 15 *Corydalis* cp-genomes. Block 1 (approximately 6 kb) of 10 *Corydalis* cp-genomes contained 4–5 genes (*trnM-CAU*, *atpE*, *atpB*, *rbcl*, and *trnV-UAC*) relocated from the classically posterior part of the LSC region (downstream of the *ndhC* gene) to the front. Of these cp-genomes with block 1, three cp-genomes from subgenus *Sophorocarpus* (*C. saxicola*, *C. fangshanensis*, and *C. tomentella*) displayed different types of relocation (downstream of the *atpH* gene) from other subgenera (downstream of the *trnK-UUU* gene). Then, in the cp-genome of *C. adunca* (subg. *Cremnocarpus*), block 2 with 1 kb of the *rps16* gene relocated from the typical LSC region to downstream of the *ndhF* gene in the IR region. In addition, blocks 5–7 with approximately 13 kb in the IR region contained 11 genes (*ndhB-trnR-ACG*) inverted uniformly in *C. pinnata*, *C. mucronata*, *C. hsiaowutaishanensis* (subg. *Corydalis*), *C. sheareri* (subg. *Rapiferae*), *C. adunca* (subg. *Cremnocarpus*), *C. saxicola*, and *C. fangshanensis*. In *C. conspersa* (subg. *Rapiferae*) and *C. davidii* (subg. *Fasciculatae*), block 6 was also inverted but blocks 5 and 7 were lost. Blocks 12–15 (~8 kb) in the SSC region contained five genes (*ndhA-ycf1*) inverted uniformly in *C. hsiaowutaishanensis*, *C. conspersa*, *C. davidii*, *C. adunca*, *C. saxicola*, and *C. inopinata*. Moreover, blocks 9–11 were inverted with blocks 12–15 in *C. hsiaowutaishanensis*, *C. conspersa*, and *C. inopinata*, whereas blocks 9–11 and 16 underwent various degrees of loss in *C. davidii*, *C. adunca*, *C. saxicola*, and *C. fangshanensis*. In addition, blocks 3–8, with approximately 36 kb in the IR region, contained 54 genes (*trnN-GUU-psaI*) and were inverted uniformly in *C. tomentella* (subg. *Sophorocarpus*) compared with *C. fangshanensis* and *C. tomentella* from the same subgenus.

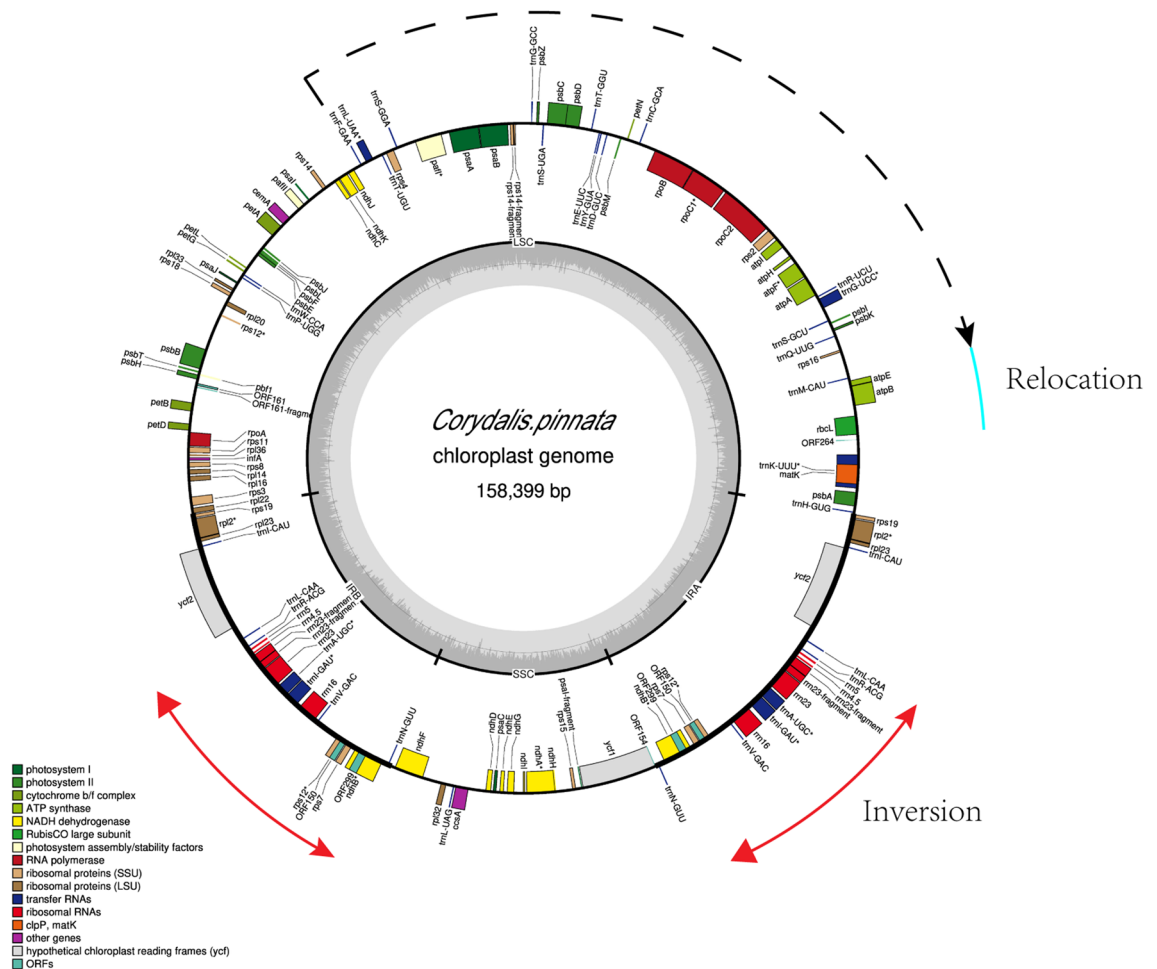


Figure 1. Gene map of the chloroplast genome of *C. pinnata*. Genes within the circle are transcribed clockwise, and those outside are transcribed counterclockwise. Genes belonging to different functional groups are colour coded. The dark grey in the inner circle corresponds to DNA G+C content, and the light grey corresponds to A+T content.

Comparison of genomic variation in the three newly sequenced *Corydalis cp*-genomes and *C. edulis cp*-genome. Previous studies reported a marked IR region expansion in some *Corydalis cp*-genomes; the IR region expanded into the simple sequence repeat (SSR) region and led to IR–SSC boundary variations^{1,2}. In the present study, three newly sequenced *Corydalis cp*-genomes were compared with the *C. edulis cp*-genome, which exhibited a typical angiosperm quadripartite cp-genome structure (Fig. 3). The location of the IR region in the three newly sequenced *Corydalis cp*-genomes was relatively conservative (Fig. 3). In these three species, *rps19* was located in the LSC region, and *ndhF* was in the SSC region. The coding region of *rpl2* was in the IR region of the *C. pinnata cp*-genome but spanned the LSC and IRa regions of the *C. mucronata* and *C. shearerii cp*-genomes; therefore, the IRb/LSC boundary (the 5' end was lost) region created a pseudogene.

The *C. edulis cp*-genome was used as a reference to ascertain differences in the genomic sequences of the three newly sequenced *Corydalis cp*-genomes (Fig. 4a,b). The rearranged regions exhibited higher variability compared with the other regions of the four *Corydalis cp*-genomes studied (Fig. 4a). Similar to other cp-genomes of angiosperms, most of the protein-coding genes were highly conserved, except for the large variation in the protein-coding genes of some genes (e.g., *rps19*, *rpl22*, *ycf1* and *ycf2*), intron regions (*paf1*, *ndhA* and *rpl2*), and intergenic regions (*trnQ-UUG-psbK*, *psbK-psbI*, *atpF-aptH*, *atpH-atpI*, *rpoB-trnC-GCA*, *trnC-GCA-petN*, *trnT-GGU-psbD*, *trnE-UUC-trnT-GGU*, *trnD-GUC-trnY-GUA*, *psaA-pafI*, *pafI-trnS-GGA*, *rps4-trnT-UGU*, *trnT-UGU-trnL-UAA*, *trnR-ACG-trnL-CAA*, and *trnN-GUU-ndhB*) among the chloroplast genomic sequences with a higher degree of variation. Such higher-resolution loci have the potential to be used as barcodes in species identification.

Analyses of long repetitive sequences and SSRs. Interspersed repeated sequences (IRSs) with a repeat unit length of ≥ 39 bp were evaluated in the chloroplast genomes of *C. pinnata*, *C. mucronata*, and *C. shearerii*. These repeats comprised only forward and palindromic repeats and lacked reverse and complementary repeats that are common in other species. Fifty IRSs were found, and among these, the sequence lengths in *C. pinnata*, *C. mucronata*, and *C. shearerii* were 40–49, > 80 , and $\leq 49/\geq 80$ bp, respectively. The IRS analyses of the chloroplast genomes are shown in Fig. 5a–c.

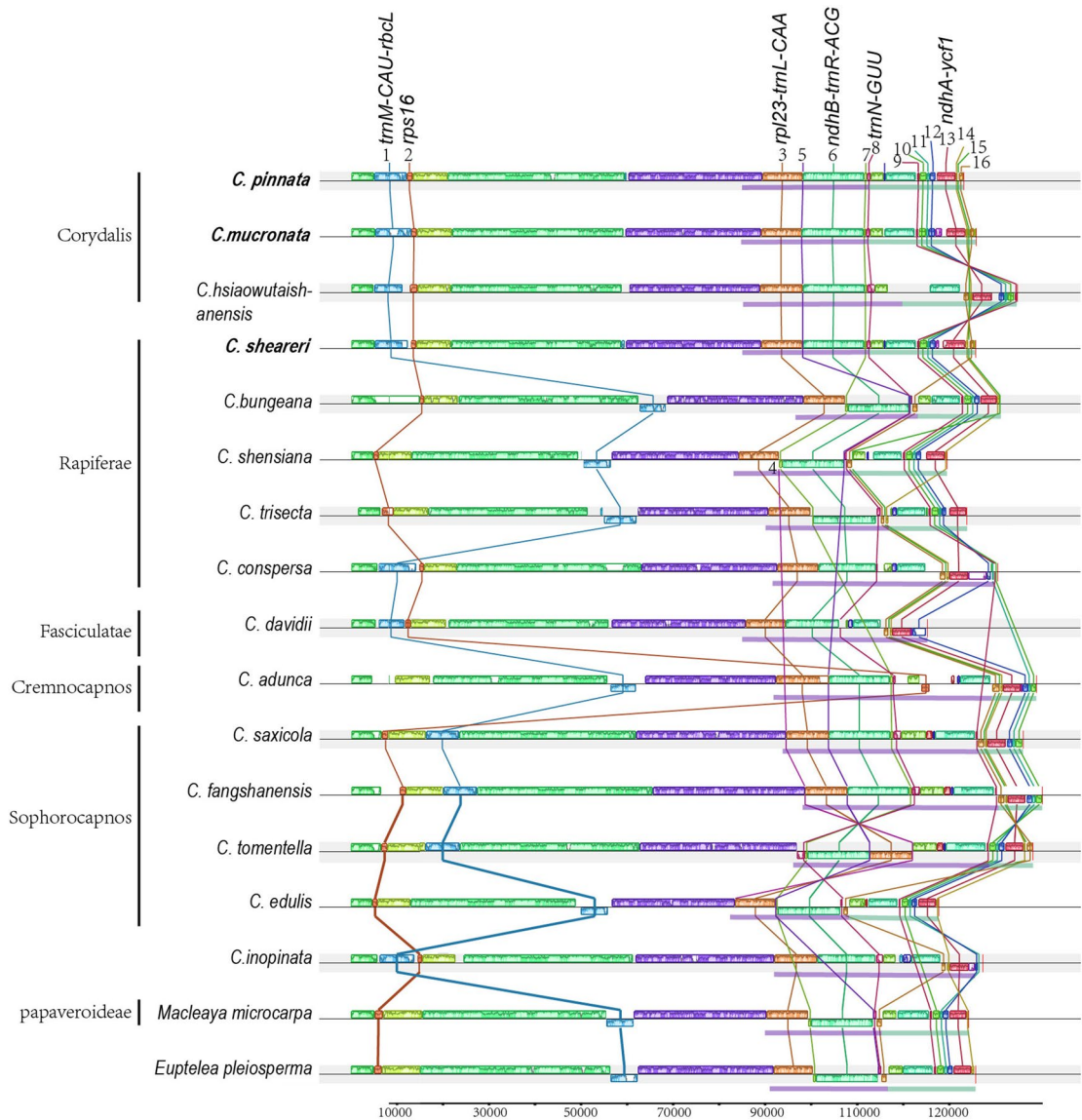


Figure 2. Chloroplast structural alignment of *Corydalis* and *E. pleiosperma* using Mauve. LCBs are represented by coloured blocks, and the blocks connected by lines indicate homology. Violet lines below each cp-genome represent IR regions, and green lines represent SSR regions. Species in bold are newly sequenced.

In total, 46 SSRs were found in *C. pinnata*, including 38 mononucleotide repeats, 1 dinucleotide repeat, and 5 trinucleotide repeats; 51 SSRs were identified in *C. mucronata*, including 43 mononucleotide repeats, 1 dinucleotide repeat, 5 trinucleotide repeats, and 2 pentanucleotide repeats; and 46 SSRs were found in *C. sheareri*, including 35 mononucleotide repeats, 1 dinucleotide repeat, 5 trinucleotide repeats, 2 tetranucleotide repeats, and 3 hexanucleotides (Fig. 5d).

Phylogenetic analyses. Using concatenated single-copy orthologous proteins to resolve phylogenetic relationships could avoid rearrangement-misled phylogenetic tree reconstruction and provide a more reliable evolutionary framework compared with using several specific genes¹⁸. Therefore, the predicted proteome was used in the phylogenetic analyses rather than the whole cp-genome sequence. Based on 31 single-copy orthologous proteins conserved in 27 species with *E. pleiosperma* as the outgroup, a maximum-likelihood (ML) phylogenetic tree was reconstructed to illuminate the evolutionary history of the compared species (Fig. 6). The ML tree had three major clades: the Fumarioideae clade, Papaveroideae clade, and the clade with the rest of the Ranunculales family members. *Corydalis* constituted a monophyletic sub-clade nested within the Fumarioideae clade. All lineages within *Corydalis* were strongly supported. The three newly sequenced *Corydalis* cp-genomes, namely, *C. pinnata* (Sect. *Mucronatae*), *C. mucronata* (Sect. *Mucronatae*), and *C. sheareri* (Sect. *Asterostigmata*), were closely related.

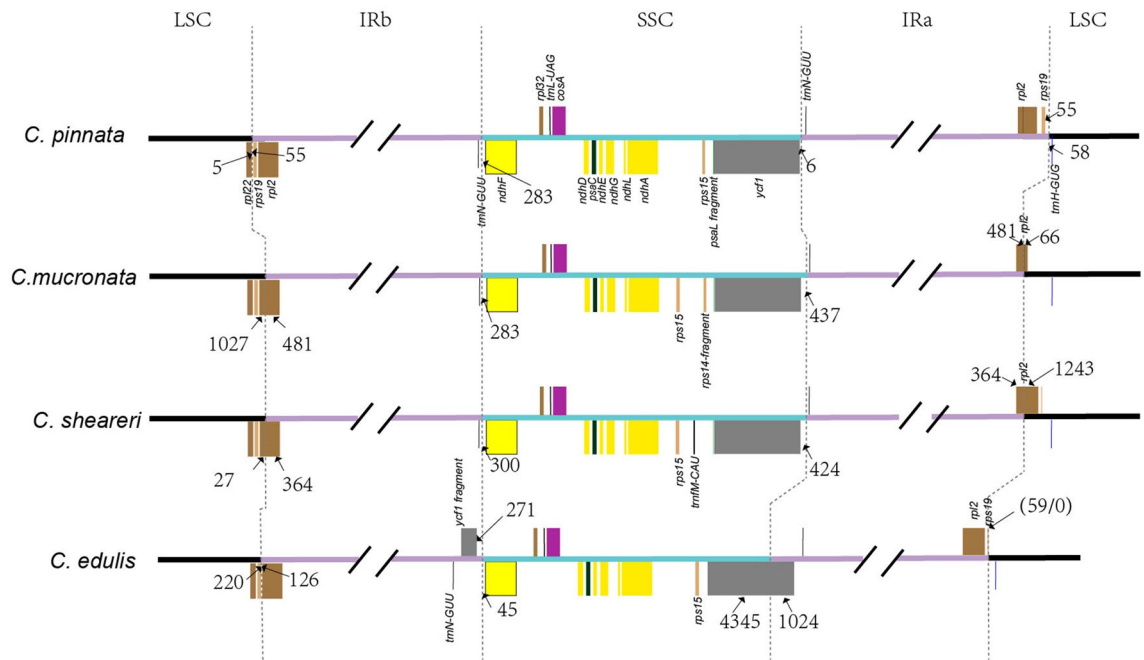


Figure 3. Comparison of genome boundaries in chloroplasts from *C. pinnata*, *C. mucronata*, *C. sheareri* and *C. edulis*.

Discussion

Although the three newly sequenced *Corydalis* cp-genomes from the same geographic region belong to two different subgenera of *Corydalis*, the sizes and structures of their LSC, IR, and SSC regions, as well as their total genomes, are highly similar. This includes similar gene losses, inversions, and relocations (Fig. 1 and Supplementary Table 1), which are common features in the *Corydalis* cp-genomes and are considered to be responsible for the variation in cp-genome sizes¹.

The loss of three genes (*accD*, *clpP*, and *trnV-UAC*) is a synapomorphic characteristic in the *Corydalis* cp-genomes (Supplementary Table 1). Xu et al.¹ speculated that the loss of the *accD* gene occurred before divergence of the genus *Corydalis*. However, in the present study, the *accD* gene was found in the cp-genomes of a few species of the subgenus Rapiferae (Supplementary Table 1), which indicated that the loss event happened after divergence of the genus *Corydalis*. The exact time of the loss event should be further explored by gathering more information on *Corydalis* cp-genomes. The *accD* gene is relocated to the nucleus in some species, such as some members of the family Campanulaceae^{19,20}. The pseudogenization or loss of 11 chloroplast *ndh* genes that encode NADH dehydrogenase subunits only occurred in a few species of the genus *Corydalis* (*C. conspersa*, *C. davidii*, *C. adunca*, and *C. inopinata*; Supplementary Table 1). Strikingly, these species are all located in high-altitude areas (1000–5200 m a.s.l.)²¹. Therefore, extreme changes in the environment may result in gene deletions or pseudogenization; this phenomenon has been observed in other species²². Further studies are required to determine whether or not the pseudogenization or loss of *ndh* genes will affect photosynthesis in those plants.

The chloroplast genome, as a photosynthetic organelle, is highly conserved in terms of structure, gene content, and arrangement^{23–25}. Large-scale rearrangement exists only occasionally in a few lineages, such as Campanulaceae^{16,17,26–28}, Ranunculaceae^{29,30}, Geraniaceae^{31–36}, Fabaceae^{15,37–44}, Oleaceae⁴⁵, Asteraceae^{46–49}, Plantaginaceae^{50–52}, Euphorbiaceae⁵³ and Poaceae^{14,54–57}. In the present study, rearrangement predominantly occurred in 16 regions (blocks 1–16, Fig. 2) of *Corydalis* plants, which determine the diversity in *Corydalis* cp-genomes. Repeat sequences may contribute to structural variations in relatively stable rearrangement regions^{58–60}. Relocation only occurred in the LSC region of the *Corydalis* cp-genomes, and inversion only occurred in the IR and SSC regions (Fig. 2). This suggested that the patterns of relocation and inversion were regulated in different ways. In addition, blocks 1–16 are likely active rearrangement regions because they have various rearrangement patterns. *C. hsiaowutaishanensis* (subg. *Corydalis*), *C. adunca* (subg. *Cremnocapnos*), *C. Saxicola*, and *C. fangshanensis* (subg. *Sophorocapnos*) all underwent the inversion of blocks 10–16, but the inversion boundaries of *C. hsiaowutaishanensis* expanded into block 9, suggesting that the inversion of blocks 9–16 in *C. hsiaowutaishanensis* was an independent event. Furthermore, some species from different subgenera have the same relocation or inversion pattern, such as the three *Corydalis* plants (*C. pinnata*, *C. mucronata*, and *C. sheareri*) collected from Qingcheng Mountain in the current study. Although they represent two subgenera, these three species have an almost identical relocation/inversion pattern in their cp-genomes (Fig. 2). Moreover, blocks 5–7 underwent at least two inversions in *C. tomentella*; blocks 5–7 initially inverted independently and then inverted with blocks 3, 4, and 8. This active rearrangement suggested that relocation or inversion in *Corydalis* cp-genomes might be affected by the geographical environment.

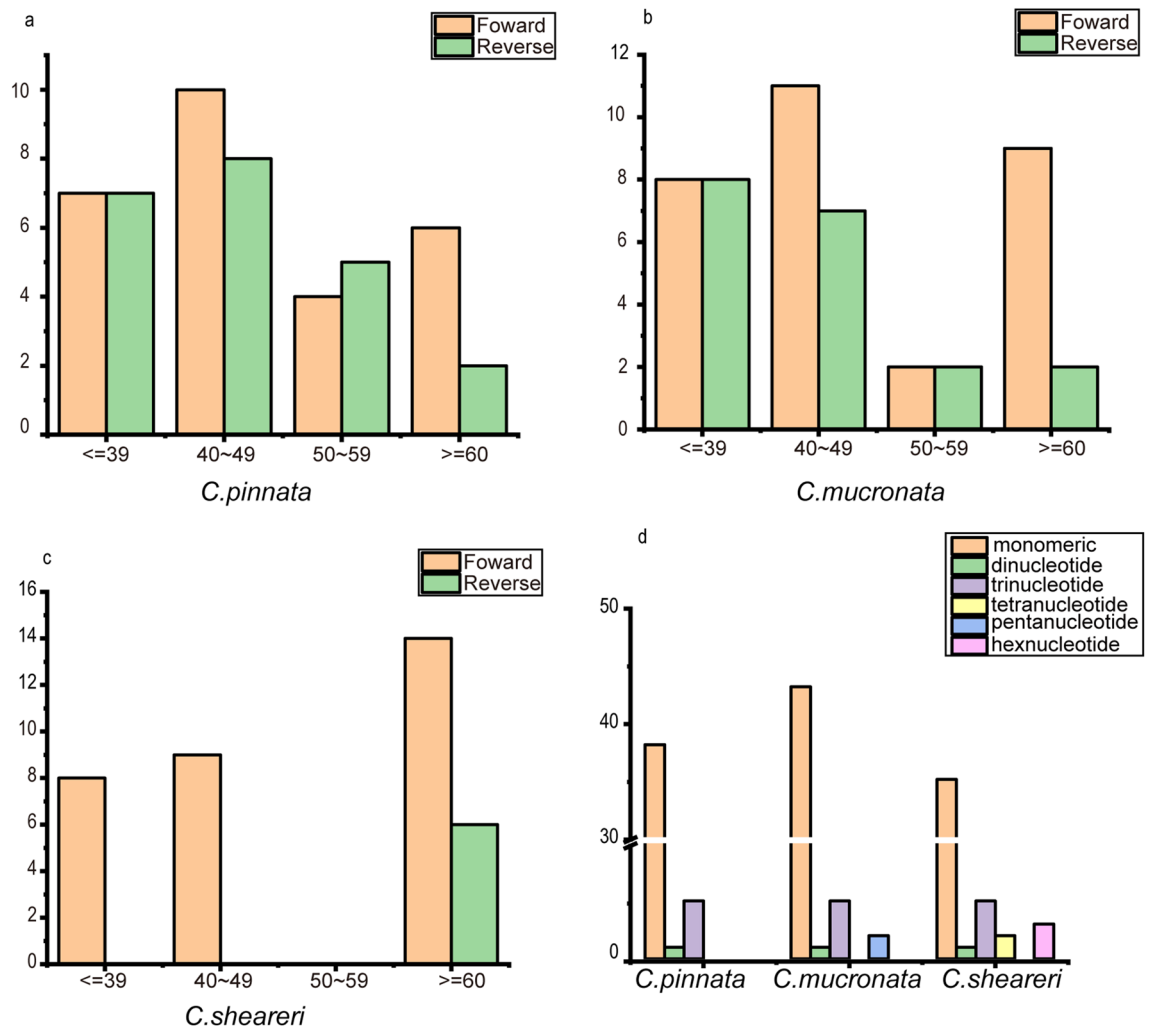


Figure 5. Long repetitive sequences and SSR distribution in the chloroplast genomes of *C. pinnata*, *C. mucronata* and *C. shearerii*.

Loss of introns and/or genes is instrumental in the regulation of gene expression and can control gene expression temporally and in a tissue-specific manner^{61–63}. The regulation mechanisms of introns for gene expression in plants and animals have been reported^{63–65}. However, the implications or link between gene expression and intron loss for *Corydalis* have not been published. Further experimental work on the roles of introns in *Corydalis* is therefore essential and should prove interesting. Highly variable DNA barcodes play an important role in species identification and phylogenetic analyses. In the current study, protein-coding genes (*rps19*, *rpl22*, *ycf1*, and *ycf2*), intron regions (*paf1*, *ndhA*, and *rpl2*), and the intergenic regions (*trnQ-UUG-psbK*, *psbK-psbI*, *atpF-aptH*, *atpH-atpI*, *rpoB-trnC-GCA*, *trnC-GCA-petN*, *trnT-GGU-psbD*, *trnE-UUC-trnT-GGU*, *trnD-GUC-trnY-GUA*, *psaA-paf1*, *paf1-trnS-GGA*, *rps4-trnT-UGU*, *trnT-UGU-trnL-UAA*, *trnR-ACG-trnL-CAA*, and *trnN-GUU-ndhB*) exhibited some extent of variation and have great potential as DNA markers (Fig. 4b).

Cp-genomes have made marked contributions to the phylogenetic studies of angiosperms and to resolving the evolutionary relationships within phylogenetic clades^{66,67}. However, active rearrangement in *Corydalis* cp-genomes may mislead the reconstruction of species phylogenetic relationships based on DNA sequence of cp-genomes. Phylogenetic reconstruction of the genus *Corydalis* was previously explored with DNA barcoding⁶⁸ or relatively conserved nucleotide fragments in cp-genomes¹. However, deep relationships remained poorly resolved by this phylogenetic approach applying a few plastid markers. Some studies reported that the protein-coding genes shared by all taxa could be used to reconstruct a phylogeny^{2,34}. However, single-copy genes (SCGs) have subsequently emerged as candidates for phylogenetic analysis because paralogues are derived from duplication events other than speciation events and should therefore be discarded from phylogenetic analyses^{69,70}. Therefore, the 31 single-copy orthologous proteins in all 27 cp-genomes were used to reconstruct the phylogeny of the genus *Corydalis*. Three distinct clades were defined by high bootstrap values (Fig. 6) in the resulting phylogenetic tree, which is consistent with previous studies based on molecular markers^{1,71}. This indicated that the application of the single-copy orthologous proteins of cp-genomes can improve the resolution of the phylogeny and taxonomy of the genus *Corydalis*. Findings from the study also provide a reference for the taxonomy and identification of other plants with extensive rearrangement in cp-genomes.

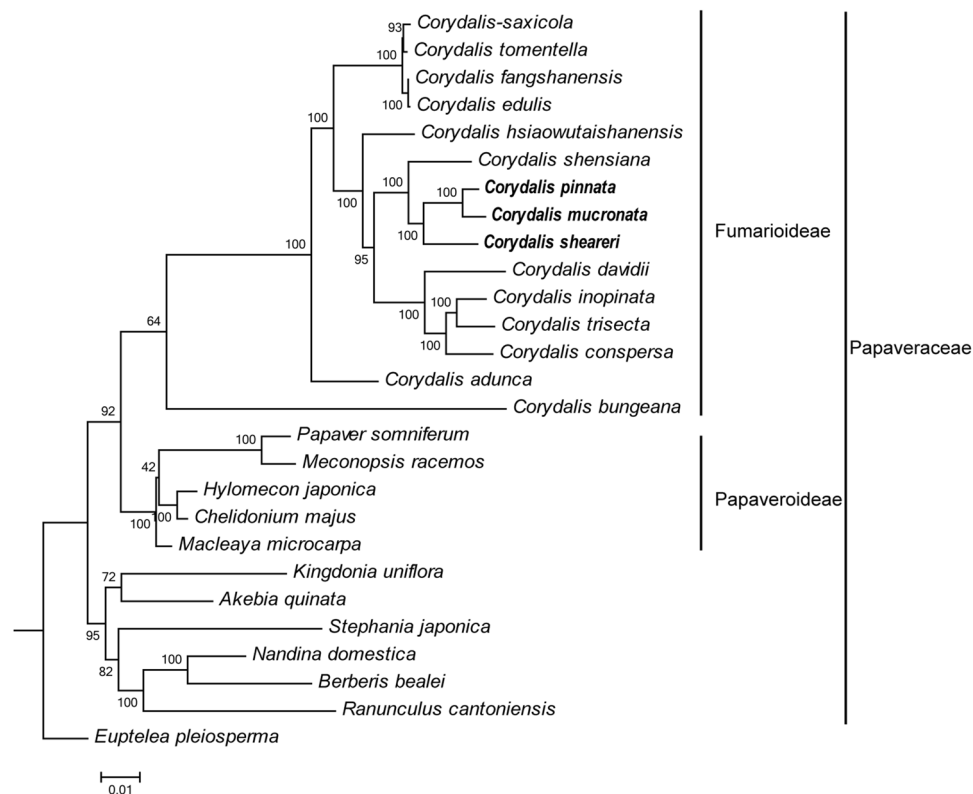


Figure 6. ML tree of *Corydalis* and its relative species based on single-copy orthologous proteins. *Euptelea pleiosperma* was used as outgroup. References for the classification of the genus *Corydalis* were World Flora Online. Bootstrap supports on the branches were calculated from 1000 replicates. Three newly sequenced *Corydalis* marked with bold.

Conclusions

The cp-genomes of three species of the genus *Corydalis* (*C. pinnata*, *C. mucronata*, and *C. sheareri*) from the Qingcheng Mountain in southwest China, including a narrow endemic species (*C. pinnata*), were characterized. The cp-genomes of the three species exhibited a large-scale rearrangement, including the relocation of four genes (*trnM-CAU-rbcL*) in the LSC region, the inversion of an 11–14-kb segment (*ndhB-trnR-ACG*) in the IR region, and the loss of three genes (*accD*, *clpP*, and *trnV-UAC*). The three *Corydalis* cp-genomes showed high similarity in terms of genome size, gene classes, gene sequences, rearrangement pattern, and distribution of repeat sequences. In addition, the structural alignment of 17 *Corydalis* cp-genomes with the typical chloroplast genomic structure of angiosperms (*E. pleiosperma*) revealed a frequent and extensive large-scale rearrangement in the *Corydalis* cp-genomes. Among them, the relocation of two blocks (*trnM-CAU-rbcL* and *rps16*) frequently appeared in the LSC region, and the inversion of four blocks (*rpl23-trnL-CAA*, *ndhB-trnR-ACG*, *trnN-GUU*, and *ndhA-ycf1*) frequently appeared in the IR and SSC regions. The extensive large-scale cp-genome rearrangement may mislead phylogenetic analysis based on cp-genomes. Single-copy orthologous proteins of cp-genomes were therefore used to reconstruct the phylogeny of the genus *Corydalis*. This method was concluded to have good prospects for elucidating the phylogeny and taxonomy of *Corydalis* and could potentially be employed for the phylogenetic analysis of other lineages with extensive rearranged cp-genomes in future studies. Findings from this study provide a reference for further studies on the taxonomy, identification, and evolution of the genus *Corydalis*.

Materials and methods

Plant collection and sampling. The aboveground parts of the three plant species were collected from Qingcheng Mountain, Sichuan Province, China (*C. sheareri*, location: E 103°32'4" N 30°54'5", altitude: 720 m a.s.l.; *C. mucronata*, location: E 103°28'35" N 30°28'35", altitude: 980 m a.s.l.; *C. pinnata*, location: E 103°25'27" N 30°65'5", altitude: 1350 m a.s.l.). The voucher specimens were deposited in the herbarium of the College of Pharmacy, Chengdu University of Traditional Chinese Medicine, China (deposition numbers: *C. sheareri*, CDCM0005283; *C. mucronata*, CDCM0005284; *C. pinnata*, CDCM0005285). The collection of samples conformed to the management provisions of the List of State-protected Wild Plants and was approved by the National Forestry and Grassland Administration of China (Supplementary Fig. 1). The specimens were identified by Professor Guihua Jiang.

DNA sequencing, assembly and validation of the chloroplast genome. A modified cetyltrimethylammonium bromide (CTAB) method was used for DNA extraction and the NEBNext Ultra DNA Library Prep Kit for Illumina sequencing was used for 500-bp paired-end library construction. A shotgun library (250 bp) was constructed according to the manufacturer's (Vazyme Biotech, Nanjing, China) instructions. Sequencing was accomplished with the X[™] Ten platform (Illumina, San Diego, CA, USA) using the double terminal sequencing method (pair-end 150)¹⁰. Total raw data from a sample was approximately 10.0 G, and > 300 million paired-end reads were attained.

Raw data were filtered by Skewer-0.2.2⁷². The resulting reads were used for genome assembly by GetOrganelle version 1.7.5⁷³. Another assembly for each species of the genus *Corydalis* was performed by ABYSS with *C. edulis* as the reference to confirm the GetOrganelle assemblies. The draft genome was used to map clean reads by BWA version 0.7.17⁷⁴, and then clean reads were filtered using SAMtools version 1.7⁷⁵. Mapping was visualized by IGV version 2.10.0⁷⁶ to check the concatenation of contigs¹. Furthermore, junction splicing sites were verified with polymerase chain reaction (PCR) and Sanger sequencing. All of the contigs were aligned to the reference cp-genome of *C. edulis* with MUMmer version 4.0⁷⁷. Finally, the sequences were extended and gaps were filled with SSPACE-3.0⁷⁸.

Gene annotation and sequence analyses. Sequence annotation was achieved by Plann version 1.1.2⁷⁹ using the cp-genome of *C. conspersa* as a reference and some manual correction. BLAST and Apollo⁸⁰ were used to check the start and stop codons and the intron/exon boundaries with the cp-genome of *C. conspersa* as a reference sequence. Complete cp-genome sequences were submitted to the NCBI. A physical map of the cp-genomes was generated with Organellar Genome OGDraw⁸¹ (<http://ogdraw.mpimp-golm.mpg.de/>).

Genome structure analyses. To determine synteny and identify possible rearrangements, 19 cp-genomes were compared using Mauve 2.4.0⁸² with the “progressiveMauve” algorithm, including 17 *Corydalis* cp-genomes, the cp-genome of *Macleaya microcarpa* (NC_039623) representing Papaveraceae, and the cp-genome of *Euptelea pleiosperma* (NC_029429) representing a typical angiosperm cp-genome. The Mauve result was then manually modified to show the notable rearrangements. The cp-genomes of species of the genus *Corydalis* were completed by mVISTA⁸³ (Shuffle-LAGAN mode) using the genome of *C. edulis* as the reference. Tandem Repeats Finder⁸⁴ was used to detect tandem repeats, forward repeats, and palindromic repeats as tested by REPuter⁸⁵. SSRs were detected by Misa.pl⁸⁶ using search parameters of mononucleotides set to ≥ 10 repeat units, dinucleotides ≥ 8 repeat units, trinucleotides and tetranucleotides ≥ 4 repeat units, and pentanucleotides and hexanucleotides ≥ 3 repeat units.

Phylogenetic analyses. Twenty-seven cp-genomes were used to reconstruct a phylogenetic tree. First, single-copy orthologous proteins were extracted by OrthoFinder version 2.3.8⁸⁷. Next, genes were aligned by MUSCLE version 3.8, and then the best-fit models of amino acid substitution were estimated by ProtTest version 3.4⁸⁸ with the best corrected Akaike Information Criterion (AICc) value selected. Finally, a ML phylogenetic tree was reconstructed by RAXML version 8.2.12⁸⁹ including tree robustness assessment using 1000 replicates of rapid bootstrap with the HIVb + I + G + F substitution model based on the results of ProtTest.

Received: 28 October 2021; Accepted: 29 July 2022

Published online: 20 August 2022

References

- Xu, X. & Wang, D. Comparative chloroplast genomics of *Corydalis* Species (Papaveraceae): Evolutionary perspectives on their unusual large scale rearrangements. *Front. Plant Sci.* **11**, 2243 (2021).
- Ren, F. *et al.* Highly variable chloroplast genome from two endangered Papaveraceae lithophytes *Corydalis tomentella* and *Corydalis saxicola*. *Ecol. Evol.* **11**, 4158–4171 (2021).
- Yu, Z., Zhou, T., Li, N. & Wang, D. The complete chloroplast genome and phylogenetic analysis of *Corydalis fangshanensis* W.T. Wang ex S.Y. He (Papaveraceae). *Mitochondrial DNA B* **6**, 3171–3173. <https://doi.org/10.1080/23802359.2021.1987172> (2021).
- Kanwal, N. *et al.* Complete chloroplast genome of a Chinese endemic species *Corydalis trisecta* Franch (Papaveraceae). *Mitochondrial DNA B* **4**, 2291–2292. <https://doi.org/10.1080/23802359.2019.1627930> (2019).
- Medicine, E. B. O. C. T. *Chinese Tibetan Medicine* (Shanghai Science and Technology Press, 1996).
- Kubo, M., Matsuda, H., ToKUoKA, K., Ma, S. & Shiomoto, H. Anti-inflammatory activities of methanolic extract and alkaloidal components from *Corydalis tuber*. *Biol. Pharm. Bull.* **17**, 262–265 (1994).
- Guo, Y. *et al.* The traditional uses, phytochemistry, pharmacokinetics, pharmacology, toxicity, and applications of *Corydalis saxicola* bunting: A review. *Front. Pharmacol.* **13**, 822792. <https://doi.org/10.3389/fphar.2022.822792> (2022).
- Lidén, M., Fukuhara, T. & Axberg, T. *Systematics and Evolution of the Ranunculiflorae* 183–188 (Springer, 1995).
- Bruneau, A., Starr, J. R. & Joly, S. Phylogenetic relationships in the genus *Rosa*: New evidence from chloroplast DNA sequences and an appraisal of current knowledge. *Syst. Bot.* **32**, 366–378 (2007).
- Yin, X. *et al.* The chloroplasts genomic analyses of *Rosa laevigata*, *R. rugosa* and *R. canina*. *Chin. Med.* **15**, 1–11 (2020).
- Ning, C. *et al.* Complete chloroplast genome of *Salvia plebeia*: Organization, specific barcode and phylogenetic analysis. *Chin. J. Nat. Med.* **18**, 563–572 (2020).
- Zhang, Z. L., Zhang, Y., Song, M. F., Guan, Y. H. & Ma, X. J. Species identification of dracaena using the complete chloroplast genome as a super-barcode. *Front. Pharmacol.* **11**, 1441 (2020).
- Wu, L. *et al.* Plant super-barcode: A case study on genome-based identification for closely related species of *Fritillaria*. *Chin. Med.* **16**, 52. <https://doi.org/10.1186/s13020-021-00460-z> (2021).
- Doyle, J. J., Davis, J. L., Soreng, R. J., Garvin, D. & Anderson, M. J. Chloroplast DNA inversions and the origin of the grass family (Poaceae). *Proc. Natl. Acad. Sci.* **89**, 7722–7726 (1992).

15. Doyle, J. J., Doyle, J. L., Ballenger, J. & Palmer, J. The distribution and phylogenetic significance of a 50-kb chloroplast DNA inversion in the flowering plant family Leguminosae. *Mol. Phylogenet. Evol.* **5**, 429–438 (1996).
16. Cosner, M. E., Raubeson, L. A. & Jansen, R. K. Chloroplast DNA rearrangements in Campanulaceae: Phylogenetic utility of highly rearranged genomes. *Bmc Evol. Biol.* **4**, 1–27 (2004).
17. Knox, E., Downie, S. & Palmer, J. Chloroplast genome rearrangements and the evolution of giant lobelias from herbaceous ancestors. *Mol. Biol. Evol.* **10**, 414–430 (1993).
18. Zhang, N., Zeng, L. P., Shan, H. Y. & Ma, H. Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. *New Phytol.* **195**, 923–937 (2012).
19. Hong, C. P. *et al.* accD nuclear transfer of *Platycodon grandiflorum* and the plastid of early Campanulaceae. *BMC Genomics* **18**, 1–13 (2017).
20. Rousseau-Gueutin, M. *et al.* Potential functional replacement of the plastidic acetyl-CoA carboxylase subunit (accD) gene by recent transfers to the nucleus in some angiosperm lineages. *Plant Physiol.* **161**, 1918–1929 (2013).
21. Li, J. Flora of China. *Harv. Pap. Bot.* **13**, 301–302 (2007).
22. Lin, C.-S. *et al.* The location and translocation of ndh genes of chloroplast origin in the Orchidaceae family. *Sci. Rep.* **5**, 1–10 (2015).
23. Mower, J. P. & Vickrey, T. L. Structural diversity among plastid genomes of land plants. *Adv. Bot. Res.* **85**, 263–292 (2018).
24. Palmer, J. D. Comparative organization of chloroplast genomes. *Annu. Rev. Genet.* **19**, 325–354 (1985).
25. Wicke, S., Schneeweiss, G. M., Depamphilis, C. W., Müller, K. F. & Quandt, D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol.* **76**, 273–297 (2011).
26. Uribe-Convers, S., Carlsen, M. M., Lagomarsino, L. P. & Muchhala, N. Phylogenetic relationships of Burmeistera (Campanulaceae: Lobelioideae): Combining whole plastome with targeted loci data in a recent radiation. *Mol. Phylogenet. Evol.* **107**, 551–563 (2017).
27. Knox, E. B. The dynamic history of plastid genomes in the *Campanulaceae* sensu lato is unique among angiosperms. *Proc. Natl. Acad. Sci.* **111**, 11097–11102 (2014).
28. Knox, E. B. & Li, C. The East Asian origin of the giant lobelias. *Am. J. Bot.* **104**, 924–938 (2017).
29. Choi, K. S. *et al.* Two Korean endemic Clematis chloroplast genomes: Inversion, reposition, expansion of the inverted repeat region, phylogenetic analysis, and nucleotide substitution rates. *Plants* **10**, 397 (2021).
30. Liu, H. *et al.* Comparative analysis of complete chloroplast genomes of Anemoclema, Anemone, Pulsatilla, and Hepatica revealing structural variations among genera in tribe Anemoneae (Ranunculaceae). *Front. Plant Sci.* **9**, 1097 (2018).
31. Palmer, J. D., Nugent, J. M. & Herbon, L. A. Unusual structure of geranium chloroplast DNA: A triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. *Proc. Natl. Acad. Sci.* **84**, 769–773 (1987).
32. Chumley, T. W. *et al.* The complete chloroplast genome sequence of Pelargonium × hortorum: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol. Biol. Evol.* **23**, 2175–2190 (2006).
33. Guisinger, M. M., Kuehl, J. V., Boore, J. L. & Jansen, R. K. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: Rearrangements, repeats, and codon usage. *Mol. Biol. Evol.* **28**, 583–600 (2011).
34. Weng, M.-L., Blazier, J. C., Govinduraj, M. & Jansen, R. K. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Mol. Biol. Evol.* **31**, 645–659 (2014).
35. Röschenbleck, J., Wicke, S., Weinl, S., Kudla, J. & Müller, K. F. Genus-wide screening reveals four distinct types of structural plastid genome organization in Pelargonium (Geraniaceae). *Genome Biol. Evol.* **9**, 64–76 (2017).
36. Weng, M. L., Ruhlman, T. A. & Jansen, R. K. Expansion of inverted repeat does not decrease substitution rates in Pelargonium plastid genomes. *New Phytol.* **214**, 842–851 (2017).
37. Kolodner, R. & Tewari, K. Inverted repeats in chloroplast DNA from higher plants. *Proc. Natl. Acad. Sci.* **76**, 41–45 (1979).
38. Palmer, J. D. & Thompson, W. F. Rearrangements in the chloroplast genomes of mung bean and pea. *Proc. Natl. Acad. Sci.* **78**, 5533–5537 (1981).
39. Lavin, M., Doyle, J. J. & Palmer, J. D. Evolutionary significance of the loss of the chloroplast-DNA inverted repeat in the Leguminosae subfamily Papilionoideae. *Evolution* **44**, 390–402 (1990).
40. Cai, Z. *et al.* Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. *J. Mol. Evol.* **67**, 696–704 (2008).
41. Martin, G. E. *et al.* The first complete chloroplast genome of the Genistein legume *Lupinus luteus*: Evidence for a novel major lineage-specific rearrangement and new insights regarding plastome evolution in the legume family. *Ann. Bot.* **113**, 1197–1210 (2014).
42. Schwarz, E. N. *et al.* Plastid genome sequences of legumes reveal parallel inversions and multiple losses of rps16 in papilionoids. *J. Syst. Evol.* **53**, 458–468 (2015).
43. Wang, Y.-H., Qu, X.-J., Chen, S.-Y., Li, D.-Z. & Yi, T.-S. Plastomes of Mimosoideae: structural and size variation, sequence divergence, and phylogenetic implication. *Tree Genet. Genomes* **13**, 41 (2017).
44. Charboneau, J. L., Cronn, R. C., Liston, A., Wojciechowski, M. F. & Sanderson, M. J. Plastome structural evolution and homoplastic inversions in Neo-Astragalus (Fabaceae). *Genome Biol. Evol.* **13**, 215 (2021).
45. Lee, H.-L., Jansen, R. K., Chumley, T. W. & Kim, K.-J. Gene relocations within chloroplast genomes of Jasminum and Menodora (Oleaceae) are due to multiple, overlapping inversions. *Mol. Biol. Evol.* **24**, 1161–1180 (2007).
46. Jansen, R. K. & Palmer, J. D. A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). *Proc. Natl. Acad. Sci.* **84**, 5818–5822 (1987).
47. Kim, K.-J., Choi, K.-S. & Jansen, R. K. Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). *Mol. Biol. Evol.* **22**, 1783–1792 (2005).
48. Sablok, G., Amiroussi, A., He, X., Hyvönen, J. & Poczai, P. Sequencing the plastid genome of giant ragweed (Ambrosia trifida, Asteraceae) from a herbarium specimen. *Front. Plant Sci.* **10**, 218 (2019).
49. Mehmood, F., Rahim, A., Heidari, P., Ahmed, I. & Poczai, P. Comparative plastome analysis of Blumea, with implications for genome evolution and phylogeny of Asteroideae. *Ecol. Evol.* **11**, 7810–7826 (2021).
50. Zhu, A., Guo, W., Gupta, S., Fan, W. & Mower, J. P. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytol.* **209**, 1747–1756 (2016).
51. Kwon, W., Kim, Y., Park, C.-H. & Park, J. The complete chloroplast genome sequence of traditional medical herb, *Plantago depressa* Willd. (Plantaginaceae). *Mitochondrial DNA B* **4**, 437–438 (2019).
52. Asaf, S. *et al.* Expanded inverted repeat region with large scale inversion in the first complete plastid genome sequence of *Plantago ovata*. *Sci. Rep.* **10**, 1–16 (2020).
53. Wei, N. *et al.* Plastome evolution in the hyperdiverse Genus Euphorbia (Euphorbiaceae) using phylogenomic and comparative analyses: Large-scale expansion and contraction of the inverted repeat region. *Front. Plant Sci.* **12**, 1555 (2021).
54. Palmer, J. D. & Thompson, W. F. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell* **29**, 537–550 (1982).
55. Michelangeli, F. A., Davis, J. I. & Stevenson, D. W. Phylogenetic relationships among Poaceae and related families as inferred from morphology, inversions in the plastid genome, and sequence data from the mitochondrial and plastid genomes. *Am. J. Bot.* **90**, 93–106 (2003).
56. Burke, S. V., Lin, C.-S., Wysocki, W. P., Clark, L. G. & Duvall, M. R. Phylogenomics and plastome evolution of tropical forest grasses (Leptaspis, Streptochaeta: Poaceae). *Front. Plant Sci.* **7**, 1993 (2016).

57. Liu, Q. *et al.* Comparative chloroplast genome analyses of *Avena*: Insights into evolutionary dynamics and phylogeny. *BMC Plant Biol.* **20**, 1–20 (2020).
58. Ogihara, Y., Terachi, T. & Sasakuma, T. Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. *Proc. Natl. Acad. Sci.* **85**, 8573–8577 (1988).
59. Milligan, B. G., Hampton, J. N. & Palmer, J. D. Dispersed repeats and structural reorganization in subclover chloroplast DNA. *Mol. Biol. Evol.* **6**, 355–368 (1989).
60. Li, J. *et al.* Assembly of the complete mitochondrial genome of an endemic plant, *Scutellaria tsinyunensis*, revealed the existence of two conformations generated by a repeat-mediated recombination. *Planta* **254**, 1–16 (2021).
61. Le Hir, H., Nott, A. & Moore, M. J. How introns influence and enhance eukaryotic gene expression. *Trends Biochem. Sci.* **28**, 215–220 (2003).
62. Niu, D.-K. & Yang, Y.-F. Why eukaryotic cells use introns to enhance gene expression: Splicing reduces transcription-associated mutagenesis by inhibiting topoisomerase I cutting activity. *Biol. Direct* **6**, 1–10 (2011).
63. Callis, J., Fromm, M. & Walbot, V. Introns increase gene expression in cultured maize cells. *Genes Dev.* **1**, 1183–1200 (1987).
64. Emami, S., Arumainayagam, D., Korf, I. & Rose, A. B. The effects of a stimulating intron on the expression of heterologous genes in *A. thaliana*. *Plant Biotechnol. J.* **11**, 555–563 (2013).
65. Choi, T., Huang, M., Gorman, C. & Jaenisch, R. A generic intron increases gene expression in transgenic mice. *Mol. Cell. Biol.* **11**, 3070–3074 (1991).
66. Haberle, R. C., Fourcade, H. M., Boore, J. L. & Jansen, R. K. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *J. Mol. Evol.* **66**, 350–361 (2008).
67. Yue, F., Cui, L., Claude, W. D., Moret, B. M. & Tang, J. Gene rearrangement analysis and ancestral order inference from chloroplast genomes with inverted repeat. *BMC Genomics* **9**, 1–9 (2008).
68. Ren, F. M. *et al.* DNA barcoding of *Corydalis*, the most taxonomically complicated genus of Papaveraceae. *Ecol. Evol.* **9**, 1934–1945 (2019).
69. Sang, T. Utility of low-copy nuclear gene sequences in plant phylogenetics. *Crit. Rev. Biochem. Mol. Biol.* **37**, 121–147 (2002).
70. Debray, K. *et al.* Identification and assessment of variable single-copy orthologous (SCO) nuclear loci for low-level phylogenomics: A case study in the genus *Rosa* (Rosaceae). *BMC Evol. Biol.* **19**, 1–19 (2019).
71. Wang, Y. W. *Systematics of Corydalis DC. (Fumariaceae)* (The Chinese Academy of Sciences, 2006).
72. Jiang, H., Lei, R., Ding, S.-W. & Zhu, S. Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinform.* **15**, 1–12 (2014).
73. Jin, J.-J. *et al.* GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **21**, 1–31 (2020).
74. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. [arXiv:1303.3997](https://arxiv.org/abs/1303.3997) (2013).
75. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
76. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
77. Delcher, A. L., Salzberg, S. L. & Phillippy, A. M. Using MUMmer to identify similar regions in large sequence sets. *Curr. Protoc. Bioinform.* **1**, 10.13.11–10.13.18 (2003).
78. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579 (2011).
79. Huang, D. I. & Cronk, Q. C. Plann: A command-line application for annotating plastome sequences. *Appl. Plant Sci.* **3**, 1500026 (2015).
80. Misra, S. & Harris, N. Using Apollo to browse and edit genome annotations. *Curr. Protoc. Bioinform.* **12**, 1–28 (2005).
81. Lohse, M., Drechsel, O. & Bock, R. OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* **52**, 267–274 (2007).
82. Darling, A. C., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394–1403 (2004).
83. Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. & Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **32**, W273–W279 (2004).
84. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
85. Kurtz, S. *et al.* REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **29**, 4633–4642 (2001).
86. Beier, S., Thiel, T., Münch, T., Scholz, U. & Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **33**, 2583–2585 (2017).
87. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 1–14 (2019).
88. Abascal, F., Zardoya, R. & Posada, D. ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104–2105 (2005).
89. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

Acknowledgements

This work was supported by grants from the The National Key Research and Development Program of China (2019YFC1712302, 2019YFC1712305); Department of Science and Technology of Shichuan Province (2020YJ0369); Research on the authentic scientific connotation of *Angelicae Dahuricae Radix* based on the difference in efficacy of the "gut-brain axis" in the treatment of migraine (82173928).

Author contributions

G.H.J and H.X.Y. conceived and designed the research framework; X.F.L and J.C.G. collected and identified the sample; X.M.Y and F.H. performed the experiments; Y.L. and N.C. analyzed the data; X.M.Y and F.H. wrote the paper; C.L.L., J.J.D. and H.W. made revisions to the final manuscript. All the authors have read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-17721-y>.

Correspondence and requests for materials should be addressed to H.Y. or G.J.

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.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022