



# Comparative Chloroplast Genomics of *Corydalis* Species (Papaveraceae): Evolutionary Perspectives on Their Unusual Large Scale Rearrangements

Xiaodong Xu<sup>1†</sup> and Dong Wang<sup>1,2\*†</sup>

<sup>1</sup> School of Life Sciences, Central China Normal University, Key Laboratory for Geographical Process Analysis and Simulation, Wuhan, China, <sup>2</sup> Bio-Resources Key Laboratory of Shaanxi Province, Shaanxi University of Technology, Hanzhong, China

## OPEN ACCESS

### Edited by:

Hervé Sauquet,  
Royal Botanic Gardens and Domain  
Trust, Australia

### Reviewed by:

Yanxia Sun,  
Wuhan Botanical Garden, Chinese  
Academy of Sciences, China  
Domingos Cardoso,  
Federal University of Bahia, Brazil

### \*Correspondence:

Dong Wang  
dw@mail.ccnu.edu.cn;  
395351998@qq.com

### †ORCID:

Xiaodong Xu  
orcid.org/0000-0001-8218-1068  
Dong Wang  
orcid.org/0000-0003-1874-7690

### Specialty section:

This article was submitted to  
Plant Systematics and Evolution,  
a section of the journal  
Frontiers in Plant Science

**Received:** 29 August 2020

**Accepted:** 21 December 2020

**Published:** 27 January 2021

### Citation:

Xu X and Wang D (2021)  
Comparative Chloroplast Genomics  
of *Corydalis* Species (Papaveraceae):  
Evolutionary Perspectives on Their  
Unusual Large Scale  
Rearrangements.  
*Front. Plant Sci.* 11:600354.  
doi: 10.3389/fpls.2020.600354

The chloroplast genome (plastome) of angiosperms (particularly photosynthetic members) is generally highly conserved, although structural rearrangements have been reported in a few lineages. In this study, we revealed *Corydalis* to be another unusual lineage with extensive large-scale plastome rearrangements. In the four newly sequenced *Corydalis* plastomes that represent all the three subgenera of *Corydalis*, we detected (1) two independent relocations of the same five genes (*trnV-UAC-rbcL*) from the typically posterior part of the large single-copy (LSC) region to the front, downstream of either the *atpH* gene in *Corydalis saxicola* or the *trnK-UUU* gene in both *Corydalis davidii* and *Corydalis hsiaowutaishanensis*; (2) relocation of the *rps16* gene from the LSC region to the inverted repeat (IR) region in *Corydalis adunca*; (3) uniform inversion of an 11–14 kb segment (*ndhB-trnR-ACG*) in the IR region of all the four *Corydalis* species (the same below); (4) expansions (> 10 kb) of IR into the small single-copy (SSC) region and corresponding contractions of SSC region; and (5) extensive pseudogenizations or losses of 13 genes (*accD*, *clpP*, and 11 *ndh* genes). In addition, we also found that the four *Corydalis* plastomes exhibited elevated GC content in both gene and intergenic regions and high number of dispersed repeats. Phylogenomic analyses generated a well-supported topology that was consistent with the result of previous studies based on a few DNA markers but contradicted with the morphological character-based taxonomy to some extent. This study provided insights into the evolution of plastomes throughout the three *Corydalis* subgenera and will be of value for further study on taxonomy, phylogeny, and evolution of *Corydalis*.

**Keywords:** *Corydalis*, plastome, rearrangement, relocation, inversion, IR expansion, pseudogenization

## INTRODUCTION

The chloroplast genome (plastome) of angiosperms (particularly photosynthetic members) is generally highly conserved in terms of structural organization, gene content, and gene arrangement (Palmer, 1985; Wicke et al., 2011; Ruhlman and Jansen, 2014; Mower and Vickrey, 2018). Plastome rearrangements, if they occurred, tend to be relatively minor. Large-scale rearrangement is rare,

but it occasionally exists in a few lineages, such as Campanulaceae (Knox et al., 1993; Cosner et al., 2004; Knox, 2014; Knox and Li, 2017; Uribe-Convers et al., 2017, etc.), Geraniaceae (Palmer et al., 1987; Chumley et al., 2006; Guisinger et al., 2011; Weng et al., 2014, 2017; Röschenbleck et al., 2016, etc.), Fabaceae (Kolodner and Tewari, 1979; Palmer and Thompson, 1981; Lavin et al., 1990; Doyle et al., 1996; Cai et al., 2008; Martin et al., 2014; Schwarz et al., 2015; Wang et al., 2017, etc.), Oleaceae (Lee et al., 2007), Asteraceae (Jansen and Palmer, 1987; Kim et al., 2005; Sablok et al., 2019, etc.), Plantaginaceae (Zhu et al., 2016; Kwon et al., 2019; Asaf et al., 2020, etc.), and Poaceae (Palmer and Thompson, 1982; Doyle et al., 1992; Michelangeli et al., 2003; Burke et al., 2016; Liu et al., 2020, etc.). Given that the structure of the two already known plastomes from the genus *Corydalis* was highly variable (Wu et al., 2020; Xu and Wang, 2020), this genus is expected to represent an overlooked lineage with extensive large-scale plastome rearrangements. However, understanding of the pattern, origin, and evolution of plastome rearrangements within *Corydalis* is presently limited by the paucity of plastome sequences.

*Corydalis* (~465 species) is the largest genus within Papaveraceae and one of the largest genera in Chinese flora (Wu et al., 1999; Zhang et al., 2008). *Corydalis* species, with a wide distribution in north temperate regions, are particularly diverse in the Hengduan Mountains and Qinghai–Tibet Plateau and adjacent areas (Wu et al., 1999; Zhang et al., 2008). Along with the uplift of the Hengduan Mountains and Qinghai–Tibet Plateau, *Corydalis* species have experienced intensive and rapid differentiation, and their plastomes must have also undergone a series of genetic shift to adapt to the drastically changed environment. *Corydalis* may represent an appropriate group to explore how the plastome content and structure have varied in a fine scale in the evolution history, when the unusual plastome rearrangements have originated, and why those changes have happened. In addition, some *Corydalis* plants have the potential to be exploited as medicine for their anti-hepatitis, antitumor, cardiovascular disease treatment, and pain-releasing effects (Luo et al., 1984; Editorial Board of Chinese Tibetan Medicine, 1996; Chinese Pharmacopoeia Commission, 2015; Zhang B. et al., 2016, etc.). However, the complexity of their morphological characters has greatly challenged our understanding of their taxonomy, ecology, evolution, and utilization. Based on morphological characters and/or a few molecular markers, previous studies (Fedde, 1936; Lidén, 1986; Lidén et al., 1995, 1997; Wu et al., 1996, 1999; Wang, 2006; Zhang et al., 2008; Zhang Z.X. et al., 2016; Jiang et al., 2018; Ren et al., 2018; Xu and Wang, 2018, etc.) have made positive contributions to the taxonomy and systematics of *Corydalis*. Until now, a robust backbone phylogeny of this genus, which is instructive for taxonomy and systematics, is still not completed due to the lack of enough genetic resources.

Complete plastome data have been applied to resolve long-standing controversies at different taxonomic levels (e.g., Jansen et al., 2007; Moore et al., 2007, 2010; Ma et al., 2014; Barrett et al., 2016; Zhai et al., 2019). The highly divergent regions of the plastome can be identified as DNA barcodes for future phylogenetic and population genetic analyses. Meanwhile, the plastome rearrangements can also be useful phylogenetic

markers, because they typically lack homoplasy and are easily identified (Jansen and Palmer, 1987; Downie and Palmer, 1992; Doyle et al., 1996; Cosner et al., 2004). So far, only two *Corydalis* plastomes were formally, but simply, reported (Wu et al., 2020; Xu and Wang, 2020).

In the present study, we newly sequenced the plastomes of four *Corydalis* species representing all the three subgenera of *Corydalis* and conducted detailed comparative genomic analyses of them with the two previously reported *Corydalis* plastomes and the rest of the Ranunculales plastomes as well. Specifically, we aimed to (1) determine the plastome structure, gene content, and gene arrangement of all the three subgenera of *Corydalis*; (2) explore the pattern, origin, evolution, and phylogenetic utility of plastome rearrangements in *Corydalis*; (3) evaluate the effectiveness of complete plastome in phylogenetic analyses; and (4) screen the highly informative plastome DNA regions for future Sanger-based studies.

## MATERIALS AND METHODS

### Taxon Sampling, DNA Extraction, Library Construction, and Sequencing

Samples of four *Corydalis* species (i.e., *Corydalis adunca* Maxim., *Corydalis saxicola* Bunting, *Corydalis hsiaowutaishanensis* T. P. Wang, and *Corydalis davidii* Franch.), representing all the three subgenera of *Corydalis*, were collected from their wild population (Table 1). Voucher specimens were deposited in the herbarium of the Central China Normal University (CCNU), Wuhan, China. Fresh leaves were dried in the wild using silica gel and preserved at  $-20^{\circ}\text{C}$  until DNA extraction. Plants were identified following the treatment of *Corydalis* in the *Flora of China* (Zhang et al., 2008). DNA extraction, library preparation, and sequencing were conducted at Novogene (Tianjin, China). Total DNA was isolated using the modified cetyl trimethylammonium bromide (CTAB) method (Doyle and Doyle, 1987). DNA degradation and contamination were monitored on 1% agarose gels. DNA concentration was measured using a Qubit 2.0 fluorometer (Life Technologies, CA, United States). Approximately 1.5  $\mu\text{g}$  of each DNA sample was fragmented by sonication to an average size of 350 bp. Sequencing libraries were generated using the NEBNext Ultra DNA Library Prep Kit for Illumina (NEB, United States) following the manufacturer's recommendations. Libraries were sequenced using Illumina NovaSeq 6000 (Illumina, San Diego, CA, United States) with  $2 \times 150$  bp paired-end reads.

Together with the above four newly sequenced samples, we included another two previously published *Corydalis* plastomes in our analyses (*Corydalis conspersa* Maxim., NC\_047208; Wu et al., 2020; and *Corydalis inopinata* Prain ex Fedde MT755641; Xu and Wang, 2020). Another five already-reported *Corydalis* plastomes were not included, because the one plastome of *Corydalis trisecta* Franch. (MK713939; Kanwal et al., 2020) was assigned as unverified in the National Center for Biotechnology Information (NCBI), and the other four plastomes from *C. saxicola* and *Corydalis tomentella* Franch. sequenced in a preprint paper (Ren et al., 2020) may have not sustained to

final verification. Besides, the rest of the Ranunculales plastomes (267, excluding unverified sequences and source sequences of the reference sequences) in NCBI (accessed on June 5, 2020) were also included in our analyses. In total, 273 plastomes were used in the present study.

## Plastome Assembly and Annotation

The adaptors in raw data were removed, and low-quality sequences were trimmed using fastp v0.20.1 (Chen et al., 2018) with default parameters. Read quality of clean reads was assessed using FastQC v0.11.9 (Andrews, 2010). The clean reads were assembled using GetOrganelle v1.6.2 (Jin et al., 2018) with the chloroplast genome of *C. inopinata* (MT755641) as reference. Another assembly for each *Corydalis* species was performed using NOVOPlasty (Dierckxsens et al., 2017), with *matK* sequence (MH319908) of *Corydalis temulifolia* as seed, to confirm the GetOrganelle assemblies. The clean reads were mapped to the draft genome using BWA v0.7.17 (Li, 2013), filtered using SAMtools 1.10 (Li et al., 2009), and visualized using IGV 2.8.0 (Robinson et al., 2011) to check the concatenation of contigs. Furthermore, the rearrangements and quadripartite junction sites in the four newly sequenced *Corydalis* plastomes were verified with PCR and Sanger sequencing. The newly designed primers and their location on the plastomes were listed in **Supplementary Table S1**.

The complete chloroplast genomes were annotated using PGA (Qu et al., 2019), with the plastome of *Euptelea pleiosperma* Hook. f. et Thoms. (NC\_029429), *Papaver somniferum* L. (NC\_029434), and *Lamprocapnos spectabilis* (L.) Fukuhara (NC\_039756) as reference. All tRNAs were verified by

ARAGORN v1.2.38 (Laslett and Canback, 2004) and tRNAscan-SE v2.0 (Lowe and Chan, 2016). The annotation results were contrasted with the three annotation reference plastomes and with another three model plant plastomes [*Arabidopsis thaliana* (L.) Heynh., NC\_000932; *Nicotiana tabacum* L., NC\_001879; and *Gossypium hirsutum* L., NC\_007944] and adjusted when necessary. The schematic diagrams of the chloroplast genomes were drawn in OGDRAW v1.3.1 (Greiner et al., 2019). All the newly annotated genomes were submitted to NCBI, and the accession numbers are shown in **Table 1**.

The GC content, large single-copy (LSC) size, inverted repeat (IR) size, small single-copy (SSC) size, and total genome size of all the 273 plastomes used in the present study were counted using Biopython v1.77 (Cock et al., 2009) and drawn in boxplot using Matplotlib v3.2.1 (Hunter, 2007). In addition, we also calculated GC content of each of the 97 genes (not including intron) shared between *Corydalis* plastomes and other representative Ranunculales plastomes which are listed in **Table 1**.

## Genome Structure Analyses

To determine synteny and identify possible rearrangements, we compared the four newly sequenced and two previously published *Corydalis* plastomes (those six plastomes were hereafter referred to as “the *Corydalis* plastomes”) with the plastome of *E. pleiosperma* (NC\_029429) using Mauve 2.4.0 (Darling et al., 2004) with the “progressiveMauve” algorithm. We employed *E. pleiosperma* as reference, because it exhibited typical angiosperm quadripartite plastome structure and was sister to the remaining Ranunculales. The Mauve result was manually modified to make the notable rearrangements clear and

**TABLE 1** | Taxa, vouchers/references, and GenBank accession numbers used in phylogenetic analyses.

Family	Species	Subgenus	Voucher/references	GenBank accession
Papaveraceae	<b><i>C. adunca</i> Maxim.</b>	<i>Gremnocapnos</i>	Maoxian, Sichuan, Wang et al., 200008	MT920559
	<b><i>C. saxicola</i> Bunting</b>	<i>Sophorocapnos</i>	Zigui, Hubei, Wang et al., 160219	MT920562
	<b><i>C. hsiaowutaishanensis</i> T. P. Wang</b>	<i>Corydalis</i>	Diebu, Gansu, Wang et al., 140388	MT920561
	<b><i>C. davidii</i> Franch.</b>	<i>Corydalis</i>	Baoxing, Sichuan, Wang et al., 140275	MT920560
	<i>C. inopinata</i> Prain ex Fedde	<i>Corydalis</i>	Langkazi, Xizang, Wang et al., 140599	MT755641
	<i>C. conspersa</i> Maxim.	<i>Corydalis</i>	Wu et al., 2020	NC_047208
	<i>L. spectabilis</i> (L.) Fukuhara	–	Park et al., 2018	NC_039756
	<i>Chelidonium majus</i> L.	–	Shi et al., 2019	NC_046829
	<i>Coreanomecon hylomeconoides</i> Nakai	–	Kim and Kim, 2016	NC_031446
	<i>Hylomecon japonica</i> (Thunb.) Prantl and Kündig	–	Zhang et al., 2019	NC_045388
	<i>Macleaya microcarpa</i> (Maxim.) Fedde	–	Zeng et al., 2018	NC_039623
	<i>Meconopsis racemosa</i> Maxim.	–	Zeng et al., 2018	NC_039625
	<i>Papaver somniferum</i> L.	–	Sun et al., 2016	NC_029434
	Circaeasteraceae	<i>K. uniflora</i> Balf. f. et W. W. Smith	–	Sun et al., 2017
Lardizabalaceae	<i>Akebia quinata</i> (Thunb. ex Hoult.) Decne.	–	Li et al., 2016	NC_033913
Menispermaceae	<i>Stephania japonica</i> (Thunb.) Miers	–	Sun et al., 2016	NC_029432
Ranunculaceae	<i>Glaucidium palmatum</i> Siebold and Zucc.	–	Zhai et al., 2019	NC_041539
	<i>Ranunculus cantoniensis</i> DC.	–	Li et al., 2019	NC_045920
Berberidaceae	<i>Mahonia bealei</i> (Fortune) Carrière	–	Ma et al., 2013	NC_022457
	<i>Nandina domestica</i> Thunb.	–	Moore et al., 2006	NC_008336
Eupteleaceae	<i>E. pleiosperma</i> Hook. f. et Thoms.	–	Sun et al., 2016	NC_029429

Species in bold are newly sequenced.

concise. To assess the expansion/contraction of the IR regions in detail, we compared the single-copy (SC)/IR junctions and their adjacent genes of the *Corydalis* plastomes with the *E. pleiosperma* plastome. The schematic diagram was manually modified on the basis of the plastome gene map drawn in OGDRAW. The whole sequence similarity among the *Corydalis* plastomes was comparatively analyzed and plotted using mVISTA (Frazer et al., 2004) in Shuffle-LAGAN mode, with default parameters and with the plastome of *C. conspersa* (NC\_047208) as reference.

## Repetitive Sequence Analyses

Three types of repetitive sequence were searched for in the *Corydalis* plastomes, that is, simple sequence repeats (SSR), tandem repeat, and dispersed repeat. Before those analyses were performed, one copy of the IR (IRa) was removed. The SSRs were detected using MISA v2.0 (Thiel et al., 2003) with the minimum numbers of repeats set to 10, 6, 5, 5, and 5 for mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide repeats, respectively. Tandem repeats were detected using the Tandem Repeats Finder v4.09 (Benson, 1999). The alignment parameters match, mismatch, indel, minimum alignment score, and maximum period size were set to 2, 7, 7, 80, and 500, respectively. Four types of dispersed repeat (forward, reverse, complement, and palindromic) were detected using REPuter (Kurtz et al., 2001) with the minimal repeat size set to 30 bp and with a Hamming distance of 1.

## Positive Selection Analyses

To investigate selection pressures from the environment on the *Corydalis* plastomes, we calculated non-synonymous ( $K_a$ ), synonymous ( $K_s$ ), and  $K_a/K_s$  ratios of 66 protein coding genes shared among the *Corydalis* plastomes, with the plastome of *L. spectabilis* (NC\_039756) as reference. *L. spectabilis* was chosen as the reference, because it is closest to *Corydalis* among the taxa whose complete plastome have been reported. The protein-coding DNA sequences were extracted and translated into protein sequences using Biopython v1.77 (Cock et al., 2009). Protein sequences of each gene were aligned using MAFFT v7.450 (Katoh and Standley, 2013) and then converted into their corresponding codon-based nucleotide alignments by the Perl script PAL2NAL v14 (Suyama et al., 2006). Then  $K_a$ ,  $K_s$ , and  $K_a/K_s$  ratios were calculated using the KaKs Calculator version 2.0 (Wang et al., 2010) with default parameters, except that we used the 11th genetic code (-c 11). The two anomalously large  $K_a/K_s$  values ( $K_a/K_s = 50$ ), which represent extremely low synonymous substitutions in the alignment, were changed to 0. The ratios  $K_a/K_s > 1$ ,  $K_a/K_s = 1$ , and  $K_a/K_s < 1$  suggest positive selection, neutral selection, and purifying selection, respectively.

## Phylogenetic Analyses

In total, 21 plastomes were included in our phylogenetic analyses, including all six *Corydalis* plastomes and 15 representatives of Ranunculales plastomes (Table 1). Among them, *E. pleiosperma* (NC\_029429), which was sister to the rest of the Ranunculales, was used as an outgroup to root the trees. The data matrix contains coding regions of 64 protein-coding genes and four rRNA genes, which were shared among the 21 plastomes. Those

genes including *infA* gene, *rps15* gene, and the genes listed in Table 3 were excluded, because of pseudogenization or loss of the genes in one or more analyzed plastomes. Codon-based nucleotide alignments of each individual genes were aligned using MAFFT v7.450 (Katoh and Standley, 2013) and PAL2NAL v14 (Suyama et al., 2006) collectively. After manual adjustment, the nucleotide alignments were concatenated. The optimal partitioning scheme and corresponding best fit substitution model for Bayesian inference (BI) analysis were determined using PartitionFinder v2.1.1 (AICc criterion and greedy algorithm for all the mrbayes models, Lanfear et al., 2016) with the concatenated dataset initially partitioned by gene and codon position. Another three similar PartitionFinder analyses was lunched with only one model, GTR, GTR + G, or GTR + I + G; the resulting partitioning schemes with the lowest AICc values were used in maximum likelihood (ML) analysis. Three different phylogenetic methods, that is, BI, ML, and maximum parsimony (MP), were performed. The BI analysis was performed using MrBayes v3.2.7 (Ronquist et al., 2012), with the partitioned dataset and corresponding substitution model, one cold and three heated chains for one-million-generation Markov chains, sampling every 100 generations, and 25% discarded as burn-in. The ML analysis was performed using RAxML v8.2.12 (Stamatakis, 2014) with the partitioned dataset, GTRGAMMA model, and 1,000 bootstrap replicates. The MP analysis was performed using PAUP 4.0a166 (Swofford, 2003), with the directly concatenated dataset, heuristic search with 1,000 random taxon stepwise addition sequences, tree bisection reconnection branch swapping, and 1,000 bootstrap replications.

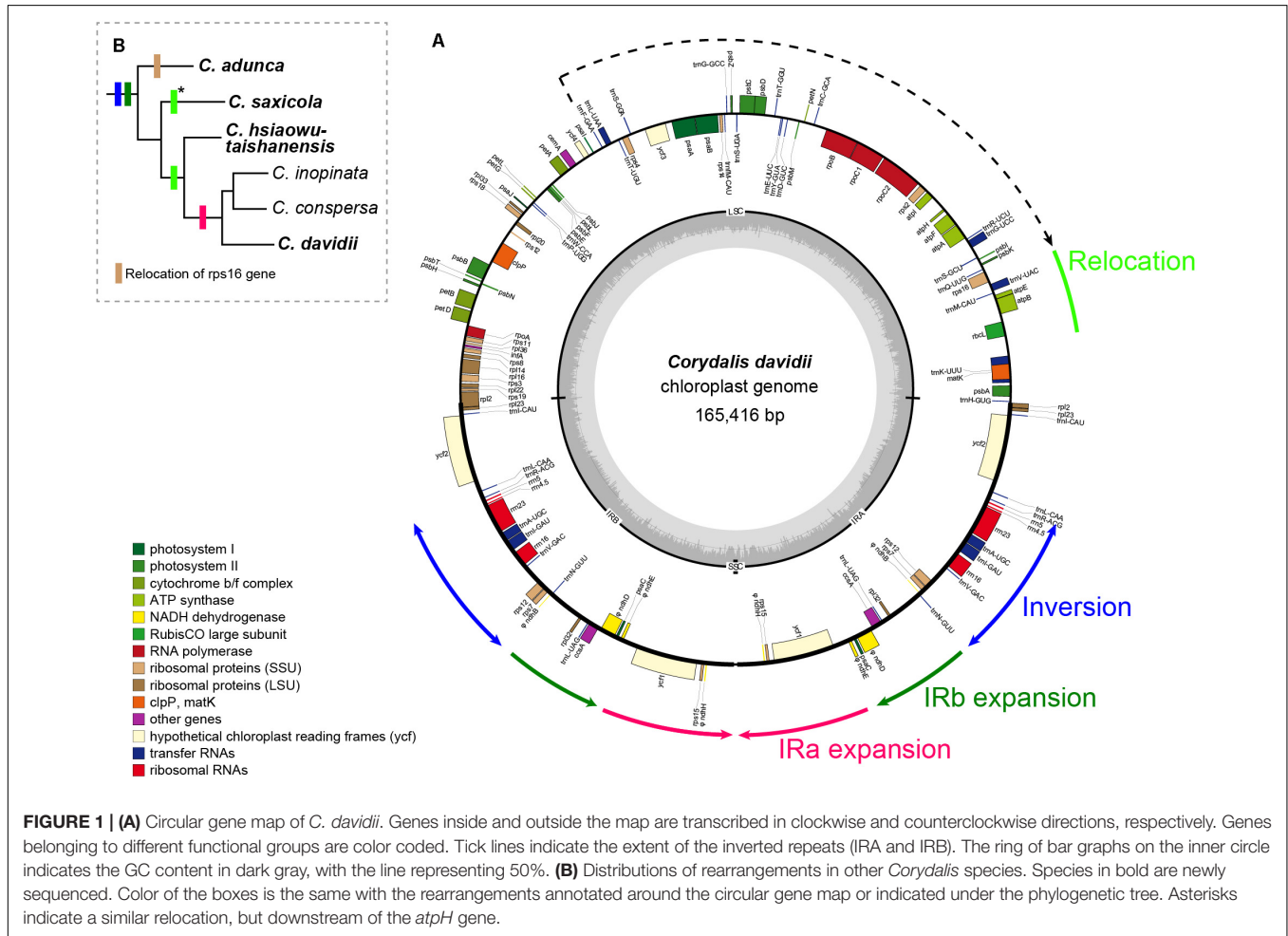
## RESULTS

### Sequencing Results and Characteristics of the Newly Sequenced *Corydalis* Plastomes

For the four newly sampled *Corydalis* species, 18,463,522–22,599,330 raw 150 nt paired-end reads ( $Q30 > 90\%$ ) were obtained, with their average coverage depth ranging from 1,273 to 7,154-folds (Table 2). The plastomes assembled using GetOrganelle and NOVOPlasty were identical. All the Sanger sequencing results were also identical to the corresponding sequences of the assembled plastomes. The four newly sequenced complete plastomes ranged in length from 165,416 to 196,128 bp (Table 2). The four newly sequenced *Corydalis* plastomes exhibited the typical angiosperm quadripartite structure, with two IR regions (39,867–47,226 bp each) separating the LSC (85,352–94,289 bp) and SSC (330–12,086 bp) regions. The chloroplast genome schematic diagram of *C. davidii* is taken as an example and shown in Figure 1. The GC content of the four newly sequenced *Corydalis* plastomes ranged from 40.21 to 41.03% (Table 2). As viewed from the boxplots, the GC content, LSC size (except for *C. davidii* and *C. hsiawutaishanensis*), IR size, and total genome size (except for *C. davidii*) of the four newly sequenced *Corydalis* plastomes were all among the largest, while their SSC sizes were all among the smallest within

**TABLE 2** | Sequencing results (raw data), genome structure, and gene content of the four newly sequenced *Corydalis* plastomes.

Species	Total reads	Total base (G)	Q30	Average depth	Genome size (bp)	LSC (bp)	IR (bp)	SSC (bp)	GC content	Coding gene	tRNA	rRNA	Pseudo gene
<i>C. adunca</i>	18,463,522	5.54	91.48%	5,316	196,128	92,145	47,226	9,531	41.03%	71	30	4	7
<i>C. saxicola</i>	21,189,298	6.36	92.86%	7,154	188,060	94,289	41,969	9,833	40.21%	78	30	4	0
<i>C. hsiaowutaishanensis</i>	19,924,311	5.98	90.03%	6,388	188,784	88,558	44,070	12,086	40.82%	77	30	4	0
<i>C. davidii</i>	22,599,330	6.78	91.91%	1,273	165,416	85,352	39,867	330	40.69%	67	30	4	4



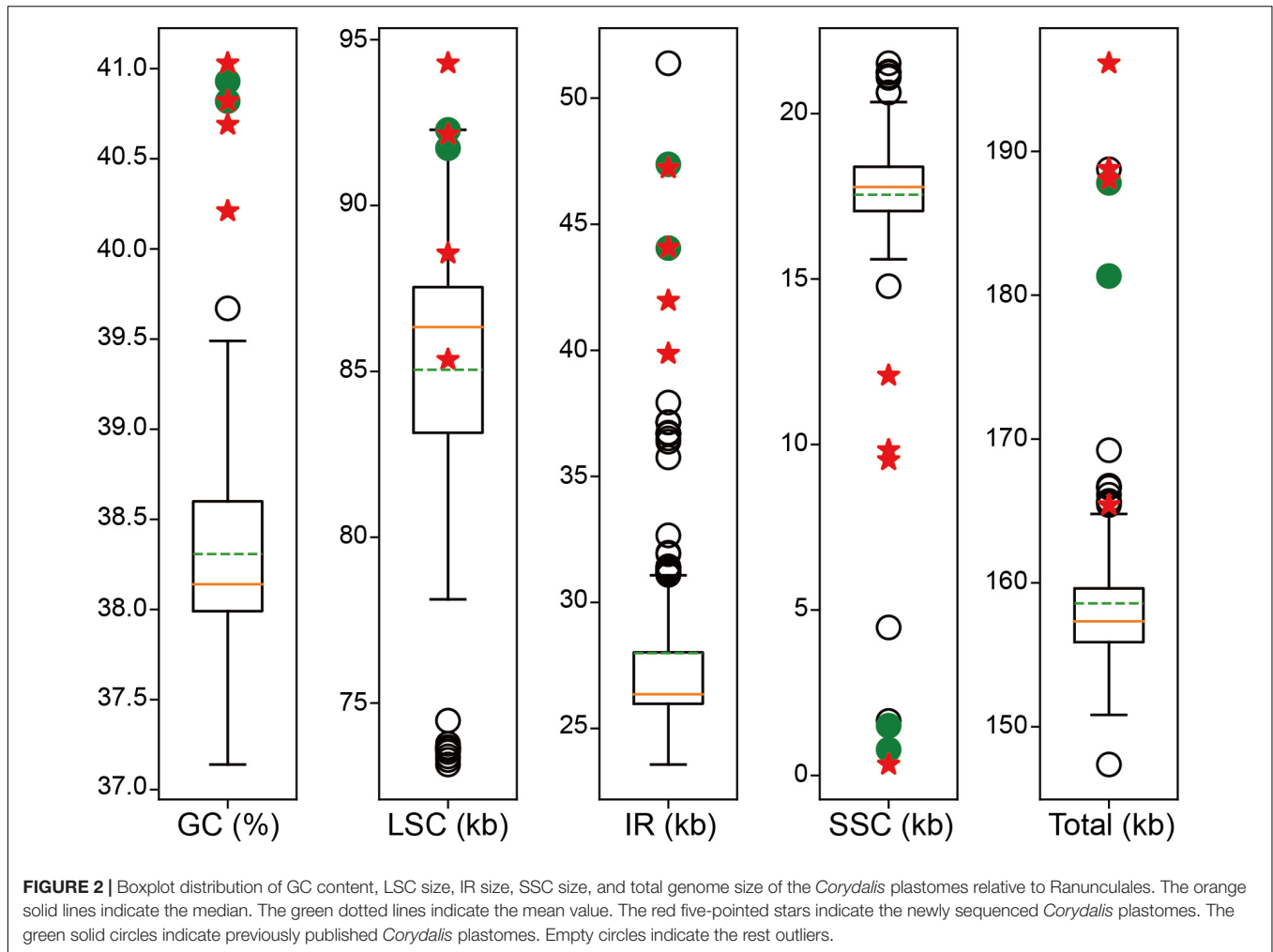
Ranunculales (**Figure 2**). The GC content of *C. adunca* (41.03%) is the highest, the LSC size of *C. saxicola* (94,289 bp) is the largest, and the SSC size of *C. davidii* (330 bp) is the smallest ever recorded in this order. For the 97 shared genes in the four newly sequenced *Corydalis* plastomes, 71.65% (278 out of 388) exhibited elevated GC content, as compared with the average GC content of 15 representatives of Ranunculales plastomes (**Supplementary Table S2**). The four newly sequenced *Corydalis* plastomes contained 101–112 unique genes, consisting of 67–78 protein-coding genes, 30 tRNA genes, and four rRNA genes, and 0–7 pseudogenes (**Table 2**). In total, 13 genes (*accD*, *clpP*, and 11 *ndh* genes) were detected to be pseudogenized or lost in one or more *Corydalis* plastomes (**Table 3**). The *accD* gene was lost in the four newly sequenced *Corydalis* plastomes, and the *clpP* gene

was lost in *C. hsiaowutaishanensis*. *C. adunca* contained truncated versions of *ndhC*, *ndhB*, *ndhF*, *ndhD*, *ndhI*, *ndhH*, and *ndhA*. *C. davidii* contained truncated versions of *ndhB*, *ndhD*, *ndhE*, and *ndhH* but lacked sequences for *ndhJ*, *ndhK*, *ndhC*, *ndhF*, *ndhG*, *ndhI*, and *ndhA*. Besides, the *rrn16* gene was duplicated in *C. adunca*, and the *psaI* gene was duplicated in both *C. saxicola* and *C. hsiaowutaishanensis*.

### Genome Structure Variations

Mauve alignment identified 16 locally collinear blocks, from which five unusual rearrangements in the four newly sequenced *Corydalis* plastomes were deduced (**Figure 3**).

Firstly, one block (~6–7 kb) contains five genes (*trnV-UAC*, *trnM-CAU*, *atpE*, *atpB*, and *rbcl*) and the associated non-coding



**TABLE 3 |** Pseudogenes and lost genes in the *Corydalis* plastomes.

Species	<i>ndhJ</i>	<i>ndhK</i>	<i>ndhC</i>	<i>ndhB</i>	<i>ndhF</i>	<i>ndhD</i>	<i>ndhE</i>	<i>ndhG</i>	<i>ndhI</i>	<i>ndhH</i>	<i>ndhA</i>	<i>accD</i>	<i>clpP</i>	<i>trnV-UAC</i>
<b><i>C. adunca</i></b>	+	+	φ	φ	φ	φ	+	+	φ	φ	φ	-	+	+
<b><i>C. saxicola</i></b>	+	+	+	+	+	+	+	+	+	+	+	-	+	+
<b><i>C. hsiaowutaishanensis</i></b>	+	+	+	+	+	+	+	+	+	+	+	-	-	+
<b><i>C. davidii</i></b>	-	-	-	φ	-	φ	φ	-	-	φ	-	-	+	+
<i>C. inopinata</i>	-	φ	φ	φ	φ	φ	+	φ	-	φ	φ	-	+	-
<i>C. conspersa</i>	-	-	-	+	+	φ	φ	+	φ	φ	+	-	+	+

Species in bold are newly sequenced. φ, pseudogene; -, lost gene; +, normal gene.

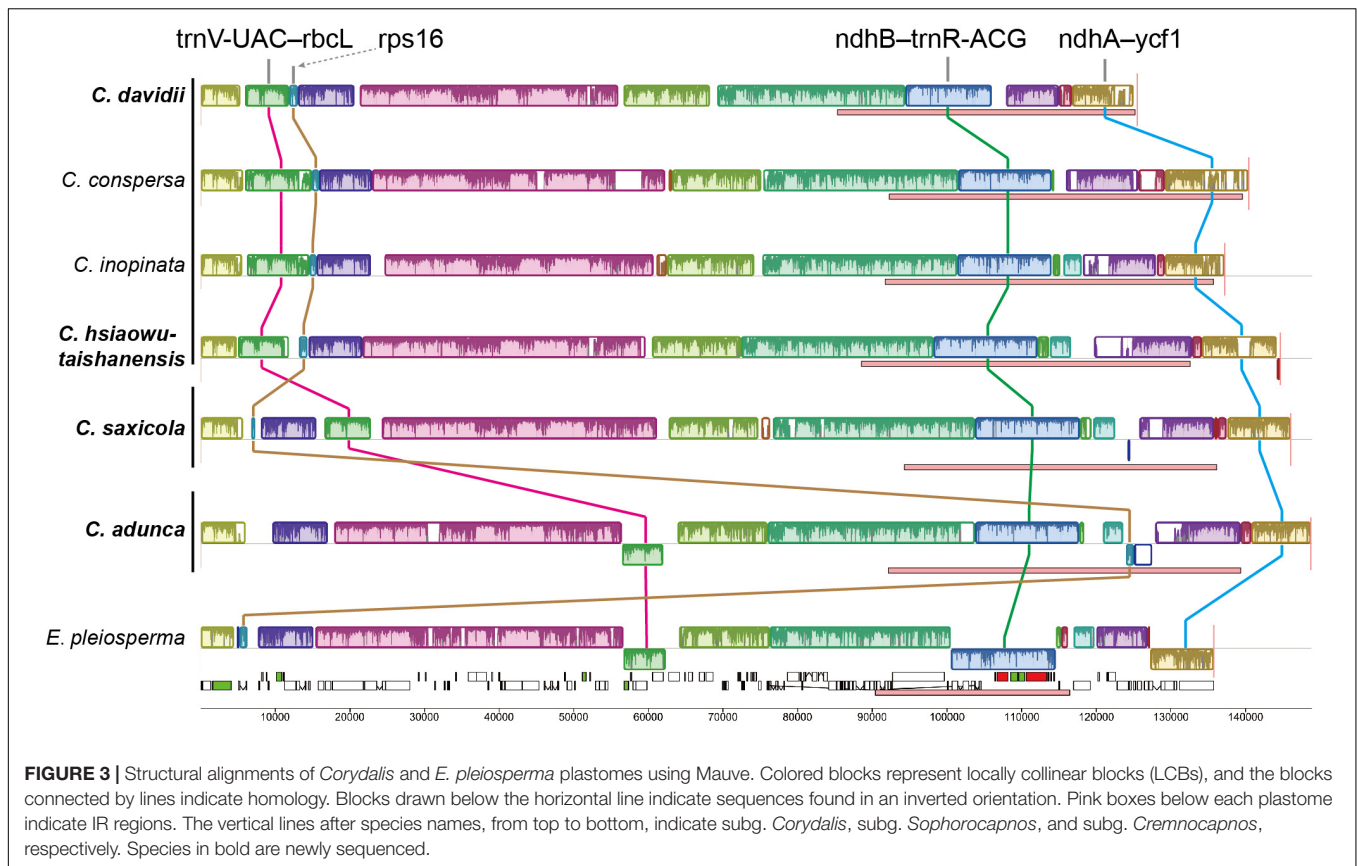
sequences, relocated from the typically posterior part of the LSC region to the front. In *C. adunca* (subg. *Cremonocarpnos*), those five genes displayed a typical angiosperm location, within the 56–62 kb region, downstream of the *ndhC* gene. In *C. saxicola* (subg. *Sophorocarpnos*), those five genes relocated to the 17–23 kb region, downstream of the *atpH* gene. In *C. davidii* and *C. hsiaowutaishanensis* (subg. *Corydalis*), those five genes relocated to the 5–12 kb region, downstream of the *trnK-UUU* gene.

Secondly, one small block (~1 kb) contained only the *rps16* gene relocated from the LSC region to downstream of the *ndhF*

gene in the IR region in *C. adunca*. Meanwhile, one duplicate of the *rrn16* gene in *C. adunca* transferred to the original *rps16* gene loci (not shown in the Mauve result).

Thirdly, one block (~11–14 kb) in the IR region contains 11 genes (*ndhB-trnR-ACG*), inverted uniformly in the four newly sequenced *Corydalis* plastomes.

Fourthly, the IRs expanded greatly, with the absorption of more than half of the SSC region in different types (Figure 4). The IRb region expanded downstream of the pseudogenized *ndhI* gene in *C. adunca* or the middle of the *ndhA* gene in *C. saxicola* or the middle of the *ndhI* gene



in *C. hsiaowutaishanensis*, resulting in the transfer of 9–10 genes (though some were pseudogenized) from the SSC region to the IR region. In contrast, in *C. davidii*, IRb expanded downstream of the pseudogenized *ndhE* gene, and successively, IRa expanded downstream of the pseudogenized *ndhH* gene. As a result, the IR regions of the four newly sequenced *Corydalis* plastomes ranged from 39,867 to 44,070 bp in length, which were more than 10 kb longer than the average (28,127 bp) of Ranunculales plastomes.

Lastly, the SSC regions were inverted uniformly in the four newly sequenced *Corydalis* plastomes.

Except for the IR–SSC boundary variations mentioned above, the IR shrank slightly at the LSC/IRb border, from the middle of the *rps19* gene to the middle of the *rpl2* gene (Figure 4), creating a 392–436 bp truncated *rpl2* gene copy (pseudogene) in the IRA region of the four newly sequenced *Corydalis* plastomes.

The alignment and conserved regions of the *Corydalis* plastomes are shown in Figure 5. Overall, they exhibited high similarity, especially in the gene regions, with the exception that some specific regions that involved plastome rearrangements were highly divergent. For the detected conserved regions, the average identity values of untranslated (tRNA and rRNA), exon, intron, and intergenic regions were 99.22, 93.33, 93.0, and 88.27%, respectively. Some intron regions (*rps16*, *rpl16*, and *trnK-UUU*) and intergenic regions (*trnQ-UUG-psbK*, *psbK-psbI*, *atpH-atpI*, *rpoB-trnC-GCA*, *trnC-GCA-petN*, *psbM-trnD-GUC*, *trnD-GUC-trnY-GUA*, *trnE-UUC-trnT-GGU*, *trnT-GGU-psbD*,

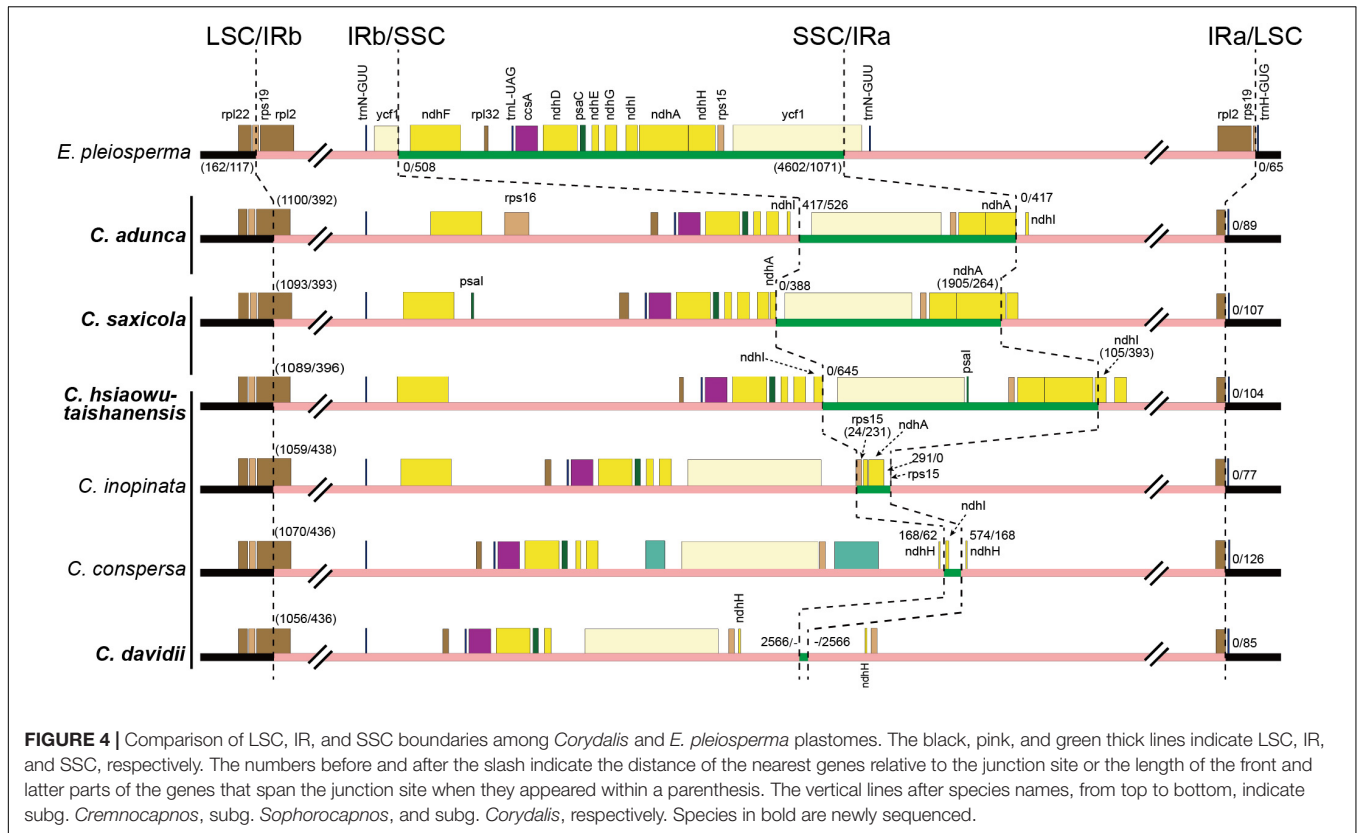
*trnT-UGU-trnL-UAA*, *rbcL-atpB*, *psaI-ycf4*, *petA-psbJ*, and *psbE-petL*), which exhibited some extent of divergence, were identified, and they have the potential to be DNA markers in future phylogenetic analyses.

## Repeat Sequences

In total, we detected 150 SSRs including 135 mononucleotide repeats and 15 dinucleotide repeats in the four newly sequenced *Corydalis* plastomes (not including IRa; Table 4). Almost all mononucleotide repeats were composed of A/T (94.07%), while only 5.93% were composed of C/G. The 15 dinucleotide repeats consisted of 14 AT/TA repeats and one AG repeat. The total numbers of SSRs in the four newly sequenced *Corydalis* plastomes were similar to those of *C. inopinata*, *C. conspersa*, and *L. spectabilis*. It was, however, larger than that of *P. somniferum* and smaller than that of *E. pleiosperma*. In total, 26–28 tandem repeats and 156–274 dispersed repeats were detected in each of the four newly sequenced *Corydalis* plastomes (Table 4). *C. adunca* has the most dispersed repeats (274), which were ~16 times of *E. pleiosperma* (17) and ~30 times of *P. somniferum* (9).

## Nucleotide Substitution Rates

The Ka/Ks ratios of the 66 protein-coding genes are shown in the bar graph (Figure 6). The majority (85.1%) of Ka/Ks ratios was between 0 and 0.5. The average Ka/Ks ratio for all those genes was 0.26. Higher Ka/Ks values were detected mainly in the *rpl*



and *rps* gene families. The Ka/Ks ratios of four genes (*psaI*, *rpl23*, *rpl36*, and *rps7*) were greater than 1 in one or more pairwise comparisons. In total, 13 Ka/Ks ratios, that is, those of all the four genes in *C. adunca*; *psaI*, *rpl36*, and *rps7* in *C. saxicola*; *rpl36* and *rps7* in both *C. hsiaowutaishanensis* and *C. conspersa*; and *psaI* and *rps7* in *C. davidii*, were larger than 1.

### Phylogenetic Analyses

The data matrix used in phylogenetic analyses consisted of 63,027 nucleotide sites; of these, 8,642 (13.7%) were parsimony informative. The PartitionFinder2 analyses using mrbayes models partitioned the concatenated dataset into 70 subsets and 14 substitution models (Supplementary Table S3). The AICc values of another three independent PartitionFinder2 analyses using GTR, GTR + G, or GTR + I + G were 524,755.977691, 518,714.822997, and 518,513.525426, respectively. The partitioning scheme determined using GTR + I + G (62 subsets; Supplementary Table S4) has the lowest AICc value and was used in ML analysis. The three phylogenetic analyses (BI, ML, and MP) generated identical topologies, and the vast majority branches received full support. Thus, only the BI phylogenetic tree is presented in Figure 7. Three clades can be defined in the phylogenetic tree, that is, the Eupteleaceae clade, the Papaveraceae clade, and the clade composed of the rest of the Ranunculales families. Within Papaveraceae, two sub-clades corresponding to the subfamilies (Papaveroideae and Fumarioideae) were resolved with full support. *Corydalis* belongs to the

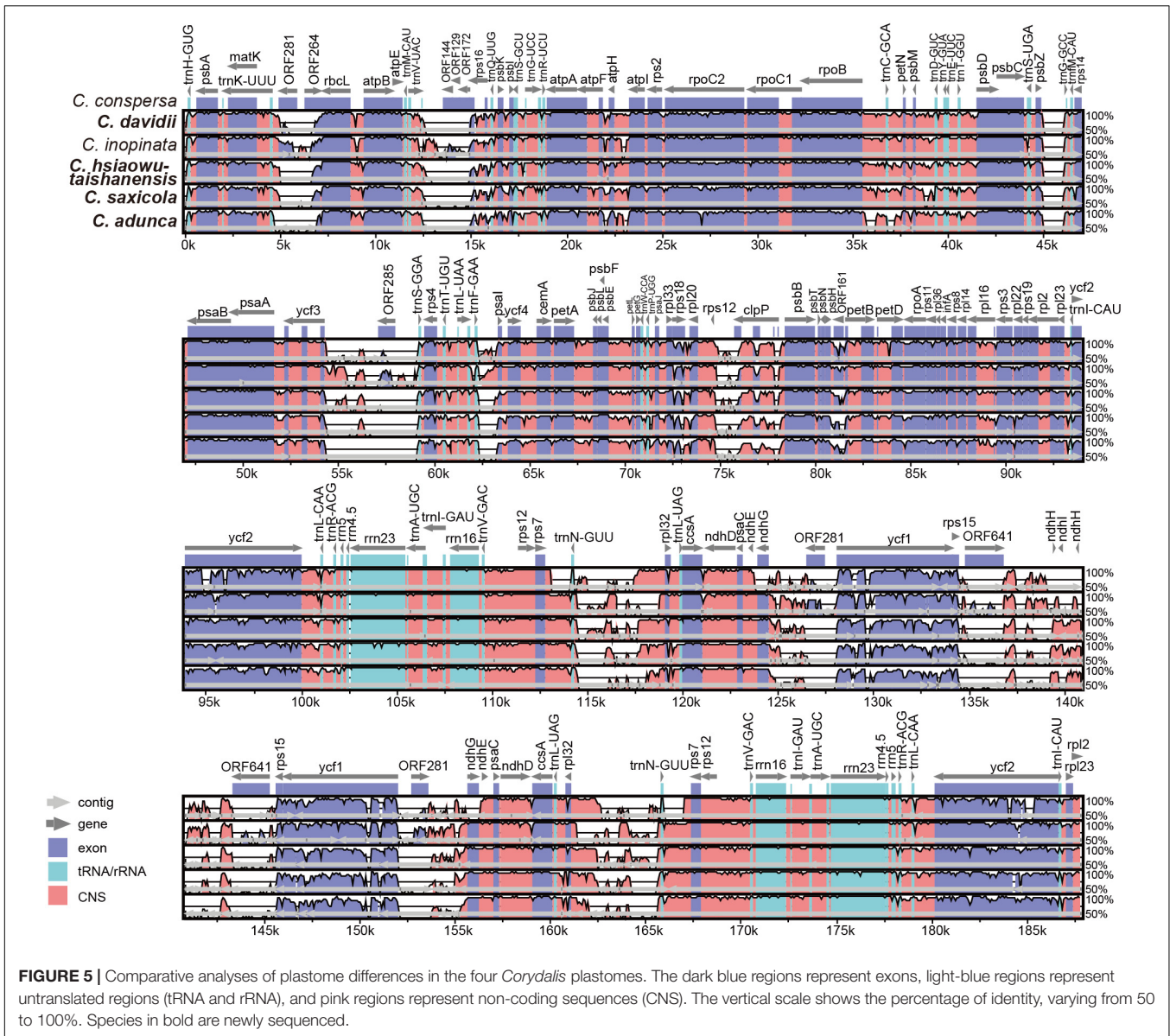
Fumarioideae sub-clade, and all lineages within *Corydalis* were fully supported. The four newly sequenced *Corydalis* species, that is, *C. adunca* (sect. *Strictae*), *C. saxicola* (sect. *Thalictrifoliae*), *C. hsiaowutaishanensis* (sect. *Dactylotuber*), and *C. davidii* (sect. *Davidianae*), served as the successively diverged lineages in this genus.

### DISCUSSION

#### Abnormal Genome Sizes, Reduced Gene Composition, Elevated GC Content, and High Number of Repeat Sequences

The LSC size, IR size, SSC size, and total genome size of the *Corydalis* plastomes were proven to be abnormally larger or smaller than those of the other Ranunculales, though a few exceptions existed (Figure 2). Several factors, for example, gene duplications, gene losses, and IR expansions, may be responsible for the variation in *Corydalis* plastome sizes. Among them, the expansions of IR into the SSC region (for details, see section “Various IR Expansions”) have contributed the most to the increase of plastome size. Gene losses (and/or pseudogenizations), on the contrary, have partially counteracted the increase of plastome size. Owing to its pseudogenization or loss of all eight *ndh* genes in the SSC and IR regions, *C. davidii* had the smallest IR and total plastome size, albeit its IR expansion was most pronounced in *Corydalis*.





**FIGURE 5 |** Comparative analyses of plastome differences in the four *Corydalis* plastomes. The dark blue regions represent exons, light-blue regions represent untranslated regions (tRNA and rRNA), and pink regions represent non-coding sequences (CNS). The vertical scale shows the percentage of identity, varying from 50 to 100%. Species in bold are newly sequenced.

**TABLE 4 |** Numbers of SSR, tandem repeat, and dispersed repeat in the plastomes of *Corydalis* species and related species.

Species	Simple sequence repeat											Tandem repeat		Dispersed repeat				Total
	A	C	G	T	AT	TA	AG	TG	AAT	GAG	TTC	Forward	Reverse	Complement	Palindromic			
<b>C. adunca</b>	20	1	-	15	2	1	1	-	-	-	-	27	137	9	4	124	341	
<b>C. saxicola</b>	7	2	-	21	2	3	-	-	-	-	-	26	180	-	-	15	256	
<b>C. hsiaowutaishanensis</b>	13	3	1	16	2	1	-	-	-	-	-	28	121	-	-	35	220	
<b>C. davidii</b>	24	1	-	11	1	2	-	-	-	-	-	28	188	1	-	2	258	
<i>C. inopinata</i>	15	2	-	9	1	-	-	-	-	-	-	19	76	-	-	5	127	
<i>C. conspersa</i>	11	1	-	13	1	-	-	-	-	-	-	29	276	-	-	4	335	
<i>L. spectabilis</i>	17	1	1	23	1	-	-	1	-	1	1	19	83	-	-	4	152	
<i>P. somniferum</i>	7	-	-	8	-	-	-	-	-	-	-	5	6	-	-	3	29	
<i>E. pleiosperma</i>	33	-	2	35	1	2	-	-	1	-	-	31	11	1	-	5	122	

Species in bold are newly sequenced.

With respect to gene content, the *Corydalis* plastomes exhibited extensive pseudogenization or loss of some of the 13 genes (11 *ndh*, *accD*, and *clpP* genes). The 11 chloroplast *ndh* genes that encode NADH dehydrogenase subunits are essential for the function of chloroplast and are widely present in angiosperms (Martin and Sabater, 2010). However, we found that 7–11 of the *ndh* genes were either pseudogenized or lost in the plastome of *C. adunca*, *C. davidii*, *C. inopinata*, and *C. conspersa*. The pseudogenization and/or loss of *ndhJ*, *ndhK*, and *ndhC* were possibly associated with the adjacent relocation event in the LSC region, while the pseudogenization and/or loss of the rest of the *ndh* genes were possibly associated with the IR boundary shift, as evidenced in another Ranunculales species (*Kingdonia uniflora*, Sun et al., 2017), where similar plastome rearrangements were also accompanied by pseudogenization and/or loss of *ndh* genes. In a phylogenetic perspective, the pseudogenization or loss of *ndh* genes in the *Corydalis* and Ranunculales may have occurred independently after the split of those species, which was probably a similar case found in the research of Orchidaceae (Lin et al., 2015). It is necessary to use more species to elucidate the pattern and mechanism of pseudogenization and loss of *ndh* genes within *Corydalis*. Another unusual case is the uniform loss of the *accD* gene in the *Corydalis* plastomes. We speculate that the loss of the *accD* gene probably occurred in the common ancestor of the *Corydalis* species. The *accD* gene encodes one of four subunits of the acetyl-CoA carboxylase enzyme (ACC), which is a rate-limiting enzyme in the first committed step of fatty acid synthesis (Cronan and Waldrop, 2002; Sasaki and Nagano, 2004). The loss of the *accD* gene is lethal, as evidenced by the study of tobacco (Kode et al., 2005). Within angiosperms, the lost *accD* has been found to relocate to the nucleus in some species, such as *Trifolium repens* (Magee et al., 2010), some Campanulaceae species (Rousseau et al., 2013), and *Platycodon grandiflorum* (Hong et al., 2017). We are not sure whether the plastid-encoded *accD* gene is lost entirely or functionally transferred to the nucleus in *Corydalis*, but there must be some kind of compensatory mechanism that can explain their loss from the plastome.

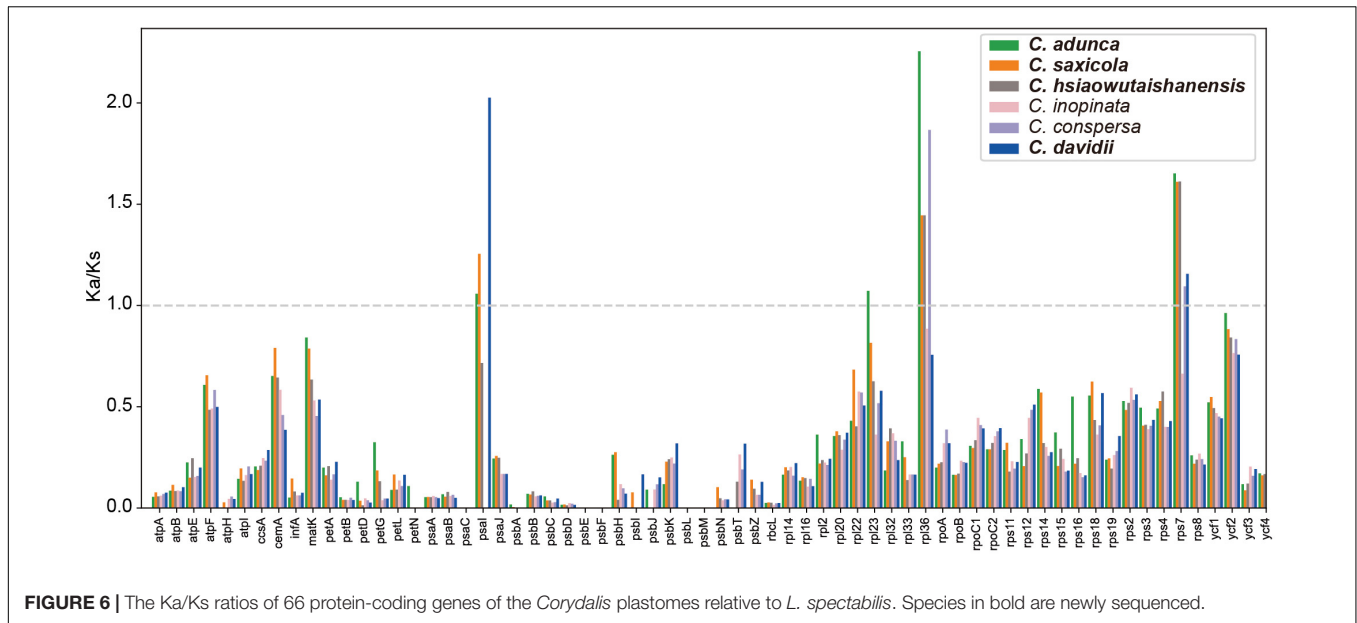
The GC content of the *Corydalis* plastomes (40.21–41.03%) was slightly higher than the average of Ranunculales (38.31%). It seems not unusual, but when examining the NCBI RefSeq database (accessed on October 19, 2020), we can find that only ~130 plastomes out of more than 5,000 land plant plastomes have GC content higher than 40%. The increase of GC content in the *Corydalis* plastomes has affected not only the gene regions but also the intergenic and intron regions. This is because the GC content of the gene region and intergenic and intron regions of the *Corydalis* plastomes has increased, by 0.8–1.75% and 3.32–4.76%, respectively, as compared with the average of the 15 representatives of Ranunculales plastomes. It is well known that the GC base pair is more stable than the AT. Thus, the overall increase of GC content may potentially enhance the stability of the plastomes and consequently contribute to their adaptation to some harsh conditions. Here, the two largest GC contents within Ranunculales, kept by *C. adunca* and *C. inopinata*, may have contributed to their adaptation to the Qinghai–Tibet Plateau.

The pairwise Ka/Ks ratios were widely used as an effective way to detect positive selection or adaptive evolution in plant species (Wang et al., 2010; Gao et al., 2019). For the protein-coding genes in the *Corydalis* plastomes, the majority (85.1%) of Ka/Ks ratios obtained in our analyses were between 0 and 0.5, which was consistent with the result of previous researches (Nei et al., 2010; Yang et al., 2020). This suggests that the majority of genes in the *Corydalis* plastomes were probably under purifying selection. The *rpl* and *rps* gene families, which are involved in self-replication, showed higher Ka/Ks ratios compared with the majority of the rest of the genes in the *Corydalis* plastome. This may contribute to their adaptation to diverse habitats. In the present study, only four genes (*psaI*, *rpl23*, *rpl36*, and *rps7*) showed Ka/Ks ratios greater than 1 (Figure 6) in one or more pairwise comparisons, indicating that they may undergo some selective pressure. All four genes (*psaI*, *rpl23*, *rpl36*, and *rps7*) and three genes (except for *rpl23*) in the plastomes of *C. adunca* and *C. saxicola*, respectively, showed Ka/Ks ratios greater than 1. Considering that these two species readily grow in dry habitats, which is different from the rest of the *Corydalis* species and *L. spectabilis*, we speculate that the chloroplast functional genes may have facilitated their adaptation to such a dry environment.

Another unique feature of the *Corydalis* plastomes was its high number of repeat sequences, especially the dispersed repeats (81–280), which can be ~30 times of the number of the dispersed repeats in *P. somniferum* (9). Some researchers have suggested that repeat sequences may contribute to plastome rearrangement (Ogihara et al., 1988; Milligan et al., 1989; Cole et al., 2018). In the present study, we also detected that the dispersed repeat sequences often located around the loci that involved rearrangement.

## Rare Gene Relocations and Inversions

Among the various chloroplast genome rearrangements, relocation is probably the least common in angiosperms (Mower and Vickrey, 2018). Until now, it was only reported within a few lineages, such as Oleaceae (Lee et al., 2007), Campanulaceae (Knox, 2014; Knox and Li, 2017; Uribe-Convers et al., 2017), and Ranunculaceae (Liu et al., 2018). In the present study, a 5 kb segment containing five genes (*trnV-UAC-rbcL*), which were traditionally located downstream of the *ndhC* gene, has unconventionally inverted and relocated either downstream of the *atpH* gene in *C. saxicola* (subg. *Sophorocapnos*) or downstream of the *trnK-UUU* gene in both *C. davidii* and *C. hsiaowutaishanensis* (subg. *Corydalis*) (Figure 3). The latter relocation event was also detected in the two previously published *Corydalis* plastomes (subg. *Corydalis*; Wu et al., 2020; Xu and Wang, 2020). In a phylogenetic view, it seems that the locations of those five genes (*trnV-UAC-rbcL*) were subgenus specific, and the relocation event probably occurred after the divergence of subg. *Cremonocapnos*. Moreover, subg. *Sophorocapnos* and subg. *Corydalis* displayed different types of *trnV-UAC-rbcL* location, suggesting that the two relocation events occurred independently in these two subgenera. Until now, the relocation of those five genes (*trnV-UAC-rbcL*) was only reported from *Corydalis* within Ranunculales. Research from some Oleaceae species (Lee et al., 2007) revealed that the



plastome relocation resulted from two overlapping inversions: the first inversion inverted the entire segment, and the second restored some part to its original gene order. It is possible that the relocation of those five genes (*trnV-UAC-rbcL*) in *Corydalis* is also the result of two overlapping inversions. Putatively, the first inversion for subg. *Corydalis* is ~50 kb, comprising *rps16-rbcL*, which translocates *trnV-UAC-rbcL* to downstream of *trnK-UUU*, and the second inversion restores the subsequent genes (*ndhC-rps16*) to their original order. Additionally, an inversion of *trnQ-UUG-rbcL*, which was similar to the first “putative” subg. *Corydalis* inversion, has been reported in two Ranunculales species (*Circaea agrestis* Maxim. and *K. uniflora*, Sun et al., 2017). It is likely that this “putative” inversion may also be found in some *Corydalis* species. There were reports of similar relocations of *trnV-UAC-rbcL* to downstream the *psbA* gene in at least 22 genera of Campanulaceae (Knox, 2014; Knox and Li, 2017; Uribe-Convers et al., 2017). Significantly, similar plastome rearrangements have happened independently in those two distantly related lineages. Elucidating the origin, evolution, and mechanism under such similar but independent genome rearrangements seems an exciting topic to explore in further comparative genomic studies.

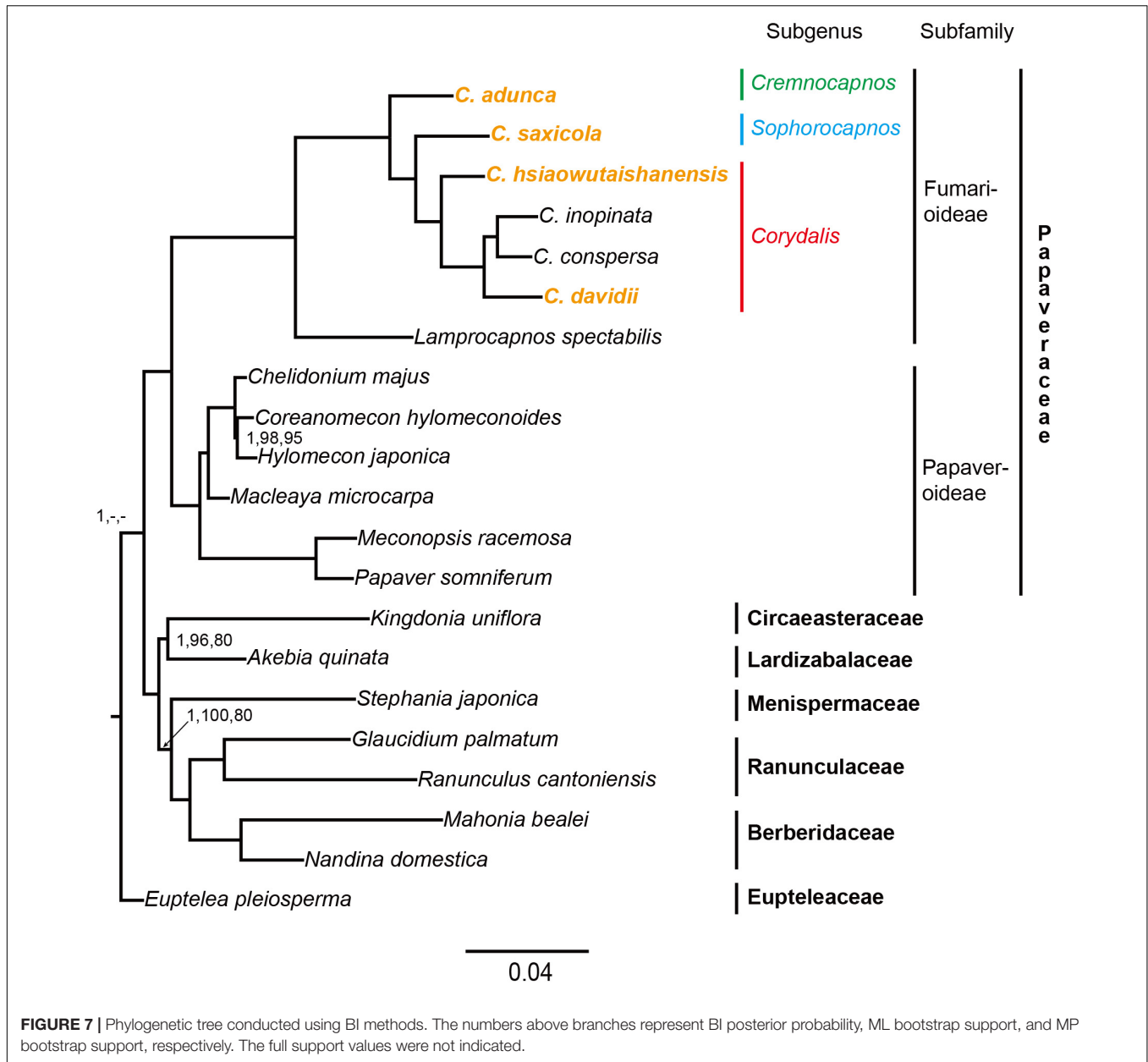
One segment that contains 11 genes (*ndhB-trnR-ACG*) in the IR region, accounting for about 60% of the typical IR region, inverted uniformly in the *Corydalis* plastomes and *L. spectabilis* plastome (Park et al., 2018). This inversion was shared by all the sequenced Fumarioideae plastomes, but it was absent from the subfamily Papaveroideae. This indicated that the inversion has occurred after the divergence of those two Papaveraceae subfamilies and existed in the common ancestor of *Lamprocapnos* and *Corydalis*. In the nearby region of the above inversion, *C. adunca*, which was sister to the rest of the *Corydalis*, was characterized by a unique relocation of the *rps16* gene from the LSC region to the IR region and by the transfer of a replicate of its *rrn16* gene to the typical *rps16*

gene site. Whether those two rearrangements occurred in the plastome of *C. adunca* simultaneously or independently, extra hidden plastome rearrangements must have happened in the evolutionary history. *L. spectabilis*, a species that diverged before *Corydalis* within the subfamily Fumarioideae, also exhibited some extra rearrangements around the inverted *ndhB-trnR-ACG* segment (Park et al., 2018). Those portend that the common ancestor of *Corydalis* and *Lamprocapnos* may possess a variable IR region. With a variable IR region in the ancestor, it may be interesting to research why only the inversion of the *ndhB-trnR-ACG* segment in the IR regions was kept throughout all those species of the subfamily Fumarioideae.

The inversion in the SSC regions, by the way, has been detected in the *Corydalis* plastomes, as compared with the plastome of *E. pleiosperma* and most other angiosperms. However, it is probably not a noticeable feature, because previous studies have found that most plastomes often occur in two distinct haplotypes (with equal frequency) differing in the orientation of SC regions (Palmer, 1983; Wang and Lanfear, 2019).

### Various IR Expansions

The IR regions of angiosperm plastomes usually start around the *rps19* gene and terminate almost uniformly at either downstream of the *trnN-GUU* or truncated *ycf1* gene. IR expansion has been reported in some lineages, but it was usually into the LSC region (e.g., in Ma et al., 2013; Sun et al., 2013; Liu et al., 2018; Mower and Vickrey, 2018; Zhai et al., 2019). The IR/SSC junctions were thought to be relatively stable (Sun et al., 2016; Zhu et al., 2016; Mower and Vickrey, 2018; Park et al., 2018). In the present study, we found that the IR of the *Corydalis* plastomes expanded markedly at the IR/SSC boundaries (Figures 3, 4), which has resulted in the second to seventh largest IR sizes



and the three smallest and sixth to eighth smallest SSC sizes in Ranunculales. The three earlier diverged *Corydalis* species (*C. adunca*, *C. saxicola*, and *C. hsiaowutaishanensis*; **Figure 7**) have only IRb/SSC expansion, while *C. inopinata*, *C. conspersa*, and *C. davidii* have a further IRa/SSC expansion. *L. spectabilis*, which also belongs to the subfamily Fumarioideae (Papaveraceae) but diverged earlier (**Figure 7**) than *Corydalis*, was characterized with the expansion of IRa into the SSC region. Thus, three types of IR/SSC expansion can be defined within the subfamily Fumarioideae, that is, IRa/SSC expansion, IRb/SSC expansion, and IRb/SSC + IRa/SSC expansion. While in the subfamily Papaverioideae (Papaveraceae), no IR expansion was observed. Hereby, we can deduce that IR expansion occurred after the diversification of those two Papaveraceae subfamilies. After a

further examination of the IR/SSC boundary within *Corydalis*, we found that the adjacent genes differed between lineages. These results suggest that more than one IR expansion event has occurred and that each lineage has its specific IR expansion history. More extensive sampling is needed to elucidate the IR expansion pattern within *Corydalis*.

### Phylogenetic Implications and New DNA Markers

The plastid phylogenomic analyses generated highly supported phylogeny with three distinct clades (i.e., Eupteleaceae, Papaveraceae, and the rest of the Ranunculales families), which is consistent with the results of previous studies (Wang et al., 2009; Pérez-Gutiérrez et al., 2015; Sauquet et al., 2015;

Sun et al., 2016). All the *Corydalis* lineages were fully supported in the phylogenetic tree, which indicates that the application of complete plastome data can improve the resolution of the phylogeny of *Corydalis*. This suggests that the complete plastome data have the potential to be employed in the construction of a robust phylogeny for *Corydalis*, which would be instructive for resolving the taxonomic controversy in this genus. Importantly, we resolved *C. adunca* (sect. *Strictae*) to be a lineage with early divergence in this genus and *C. hsiaowutaishanensis* (sect. *Dactylotuber*) to be an earlier diverged lineage in subg. *Corydalis*. This was consistent with the results of previous studies based on two DNA markers (*rps16* and *matK*; Wang, 2006) but contradicted with the morphological character-based taxonomy to some extent (Wu et al., 1996, 1999).

Highly variable DNA markers are useful, especially in fast species identification and wide-range phylogenetic analyses. In the present study, the introns of *rps16*, *rpl16*, and *trnK-UUU* and intergenic regions of *trnQ-UUG-psbK*, *psbK-psbI*, *atpH-atpI*, *rpoB-trnC-GCA*, etc., which exhibited some extent of divergence, were identified and have great potential to be exploited as DNA markers. Some of them have already been used as markers in previous phylogenetic analyses of *Corydalis* (e.g., Wang, 2006; Zhang Z.X. et al., 2016; Jiang et al., 2018; Ren et al., 2018), while the rest is in need of exploitation in the future. Nevertheless, they should be used with caution in *Corydalis*, because genes that stayed in the SC region or relocated to the IR region usually obtain different substitution rates (Zhu et al., 2016), which may mislead the phylogenetic reference. For example, the *rps16* gene, which has been used in the previous phylogenetic analysis of *Corydalis* (Lidén et al., 1997; Wang, 2006), was found to have relocated to IR in *C. adunca*, while it still remained in the LSC region in the rest of the sequenced *Corydalis* species.

## CONCLUSION

We provided insights into the plastome structure of *Corydalis* and resolved *Corydalis* to be an overlooked lineage that exhibited some extraordinary rearrangements (relocations, inversions, IR expansions, pseudogenizations, gene losses, gene duplications, etc.). Although similar rearrangements have occurred multiple times in the history of angiosperm, it is noteworthy that all those rearrangements existed in a single genus. The plastome data

## REFERENCES

- Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed 21 April 2020).
- Asaf, S., Khan, A. L., Lubna, K. A., Khan, A., Khan, G., Lee, I.-G., et al. (2020). Expanded inverted repeat region with large scale inversion in the first complete plastid genome sequence of *Plantago ovata*. *Sci. Rep.* 10:3881. doi: 10.1038/s41598-020-60803-y
- Barrett, C. F., Baker, W. J., Comer, J. R., Conran, J. G., Lahmeyer, S. C., LeebensMack, J. H., et al. (2016). Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytol.* 209, 855–870. doi: 10.1111/nph.13617
- Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. doi: 10.1093/nar/27.2.573

also turned out to be effective in enhancing the resolution of phylogenies in *Corydalis*, and it could be alternatively employed to construct a robust phylogeny for *Corydalis* in further studies. The results obtained in this study may be valuable for further study on the taxonomy, phylogeny, and evolution of *Corydalis*, a taxonomically difficult but fascinating genus.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in National Center for Biotechnology Information (NCBI) under the accession numbers: MT920559–MT920562.

## AUTHOR CONTRIBUTIONS

XX performed the analyses and drafted the manuscript. DW provided the suggestions on structuring the manuscript and the main points of the discussion and revised the manuscript. Both authors conceived the study, obtained the samples, designed the experiments, read, and approved the final manuscript.

## FUNDING

This work was supported by grants from the National Natural Science Foundation of China (31170310), Science and Technology Basic Work (2013FY112100 and 2013FY112300), the Special Foundation for the Specimen Platform of China, teaching specimen sub-platform, <http://mnh.scu.edu.cn/> (2005DKA21403-JK), Fundamental Research Funds for the Central Universities (Excellent doctoral dissertation development program of CCNU, 202050185085), and funds from the Shaanxi Provincial Bio-Resource Key Laboratory, Shaanxi University of Technology (SLGPT2019KF03-02).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2020.600354/full#supplementary-material>

- Burke, S. V., Lin, C. S., Wysocki, W. P., Clark, L. G., and Duvall, M. R. (2016). Phylogenomics and plastome evolution of tropical forest Grasses (*Leptaspis*, *Streptochoaeta*: Poaceae). *Front. Plant Sci.* 7:1993. doi: 10.3389/fpls.2016.01993
- Cai, Z. Q., Guisinger, M., Kim, H. G., Ruck, E., Blazier, J. C., McMurtry, V., et al. (2008). 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. doi: 10.1007/s00239-008-9180-7
- Chen, S. F., Zhou, Y. Q., Chen, Y. R., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560
- Chinese Pharmacopoeia Commission (2015). *Pharmacopoeia of the People's Republic of China*, Vol. 1. Beijing: China Medical Science Press.
- Chumley, T. W., Palmer, J. D., Mower, J. P., Fourcade, H. M., Calie, P. J., Boore, J. L., et al. (2006). 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. doi: 10.1093/molbev/msl089
- Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422–1423. doi: 10.1093/bioinformatics/btp163
- Cole, L. W., Guo, W., Mower, J. P., and Palmer, J. D. (2018). High and variable rates of repeat-mediated mitochondrial genome rearrangement in a genus of plants. *Mol. Biol. Evol.* 35, 2773–2785. doi: 10.1093/molbev/msy176
- Cosner, M. E., Raubeson, L. A., and Jansen, R. K. (2004). Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. *BMC Evol. Biol.* 4:27. doi: 10.1186/1471-2148-4-27
- Cronan, J. E., and Waldrop, G. L. (2002). Multi-subunit acetyl-CoA carboxylase. *Prog. Lipid Res.* 41, 407–435. doi: 10.1016/S0163-7827(02)00007-3
- Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704
- Dierckxens, N., Mardulyn, P., and Smits, G. (2017). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45:e18. doi: 10.1093/nar/gkw955
- Downie, S. R., and Palmer, J. D. (1992). “Use of chloroplast DNA rearrangements in reconstructing plant phylogeny,” in *Plant Molecular Systematics*, eds P. Soltis, D. Soltis, and J. J. Doyle (New York, NY: Chapman and Hall), 14–35. doi: 10.1007/978-1-4615-3276-7\_2
- Doyle, J. J., Davis, J. I., Soreng, R. J., Garvin, D., and Anderson, M. J. (1992). Chloroplast DNA inversions and the origin of the grass family (Poaceae). *Proc. Natl. Acad. Sci. U.S.A.* 89, 7722–7726. doi: 10.1073/pnas.89.16.7722
- Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh tissue. *Phytochem. Bull.* 19, 11–15.
- Doyle, J. J., Doyle, J. L., Ballenger, J. A., and Palmer, J. D. (1996). The Distribution and Phylogenetic Significance of a 50-kb Chloroplast DNA Inversion in the Flowering Plant Family Leguminosae. *Mol. Phylogenet. Evol.* 5, 429–438. doi: 10.1006/mpev.1996.0038
- Editorial Board of Chinese Tibetan Medicine (1996). *Chinese Tibetan medicine*. Shanghai: Shanghai Science and Technology Press.
- Fedde, F. (1936). “Papaveraceae,” in *Die naturlichen Pflanzenfamilien*, 2nd Edn, eds A. Engler and K. Prantl (Leipzig: W. Engelmann), 5–145.
- Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 32, W273–W279. doi: 10.1093/nar/gkh458
- Gao, L. Z., Liu, Y. L., Zhang, D., Li, W., Gao, J., Liu, Y., et al. (2019). Evolution of *Oryza* chloroplast genomes promoted adaptation to diverse ecological habitats. *Commun. Biol.* 2:278. doi: 10.1038/s42003-019-0531-2
- Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47, W59–W64. doi: 10.1093/nar/gkz238
- Guisinger, M. M., Kuehl, J. V., Boore, J. L., and Jansen, R. K. (2011). Extreme reconfiguration of plastid genomes in the angiosperm family geraniaceae: rearrangements, repeats, and codon usage. *Mol. Biol. Evol.* 28, 583–600. doi: 10.1093/molbev/msq229
- Hong, C. P., Park, J., Lee, Y., Lee, M., Park, S. G., Uhm, Y., et al. (2017). *accD* nuclear transfer of *Platycodon grandiflorum* and the plastid of early Campanulaceae. *BMC Genomics* 18:607. doi: 10.1186/s12864-017-4014-x
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95. doi: 10.1109/MCSE.2007.55
- Jansen, R. K., Cai, Z., Raubeson, L. A., Daniell, H., Leebens-Mack, J., Müller, K. F., et al. (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19369–19374. doi: 10.1073/pnas.0709121104
- Jansen, R. K., and Palmer, J. D. (1987). A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). *Proc. Natl. Acad. Sci. U.S.A.* 84, 5818–5822. doi: 10.1073/pnas.84.16.5818
- Jiang, L., Li, M. H., Zhao, F. X., Chu, S. S., Zhang, L. P., Xu, T., et al. (2018). Molecular identification and taxonomic implication of herbal species in genus *Corydalis* (Papaveraceae). *Molecules* 23, 1–10. doi: 10.3390/molecules23061393
- Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., Yi, T. S., and Li, D. Z. (2018). GetOrganelle: a simple and fast pipeline for de novo assemble of a complete circular chloroplast genome using genome skimming data. *bioRxiv* [Preprint]. doi: 10.1101/256479
- Kanwal, N., Zhang, X., Afzal, N., Yang, J., Li, Z. H., and Zhao, G. F. (2020). Complete chloroplast genome of a Chinese endemic species *Corydalis trisecta* Franch. (Papaveraceae). *Mitochondrial DNA B Resour.* 4, 2291–2292. doi: 10.1080/23802359.2019.1627930
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kim, H. W., and Kim, K. J. (2016). Complete plastid genome sequences of *Coreanomecon hylomeconoides* Nakai (Papaveraceae), a Korea endemic genus. *Mitochondrial DNA B Resour.* 1, 601–602. doi: 10.1080/23802359.2016.1209089
- Kim, K. J., Choi, K. S., and Jansen, R. K. (2005). Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). *Mol. Biol. Evol.* 22, 1783–1792. doi: 10.1093/molbev/msi174
- Knox, E. B. (2014). The dynamic history of plastid genomes in the Campanulaceae sensu lato is unique among angiosperms. *Proc. Natl. Acad. Sci. U.S.A.* 111, 11097–11102. doi: 10.1073/pnas.1403363111
- Knox, E. B., Downie, S. R., and Palmer, J. D. (1993). Chloroplast genome rearrangements and the evolution of giant lobelias from herbaceous ancestors. *Mol. Biol. Evol.* 10, 414–430. doi: 10.1093/oxfordjournals.molbev.a040017
- Knox, E. B., and Li, C. (2017). The East Asian origin of the giant lobelias. *Am. J. Bot.* 104, 924–938. doi: 10.3732/ajb.1700025
- Kode, V., Mudd, E. A., Iamtham, S., and Day, A. (2005). The tobacco plastid *accD* gene is essential and is required for leaf development. *Plant J.* 44, 237–244. doi: 10.1111/j.1365-313X.2005.02533.x
- Kolodner, R., and Tewari, K. K. (1979). Inverted repeats in chloroplast DNA from higher plants. *Proc. Natl. Acad. Sci. U.S.A.* 76, 41–45. doi: 10.1073/pnas.76.1.41
- Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 29, 4633–4642. doi: 10.1093/nar/29.22.4633
- Kwon, W., Kim, Y., Park, C. H., and Park, J. (2019). The complete chloroplast genome sequence of traditional medicinal herb, *Plantago depressa* Willd. (Plantaginaceae). *Mitochondrial DNA B Resour.* 4, 437–438. doi: 10.1080/23802359.2018.1553530
- Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. (2016). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34, 772–773. doi: 10.1093/molbev/msw260
- Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32, 11–16. doi: 10.1093/nar/gkh152
- Lavin, M., Doyle, J. J., and Palmer, J. D. (1990). Evolutionary significance of the loss of the chloroplast-DNA inverted repeat in the Leguminosae subfamily Papilionoideae. *Evolution* 44, 390–402. doi: 10.1111/j.1558-5646.1990.tb05207.x
- Lee, H. L., Jansen, R. K., Chumley, T. W., and Kim, K. J. (2007). Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. *Mol. Biol. Evol.* 24, 1161–1180. doi: 10.1093/molbev/msm036
- Li, B., Li, Y. D., Cai, Q. F., Lin, F. R., Huang, P., and Zheng, Y. Q. (2016). Development of chloroplast genomic resources for *Akebia quinata* (Lardizabalaceae). *Conserv. Genet. Resour.* 8, 447–449. doi: 10.1007/s12686-016-0593-0
- Li, H. (2013). *Aligning Sequence reads, Clone Sequences and Assembly Contigs with BWA-MEM*. arXiv, 1303.3997v2 [q-bio.GN]. Available online at: <https://arxiv.org/abs/1303.3997> (accessed May 15, 2020).
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, T. J., Fu, X. C., Deng, H. S., Han, X. J., Wen, F., and Xu, L. L. (2019). The complete chloroplast genome of *Ranunculus Cantonensis*. *Mitochondrial DNA B Resour.* 4, 1095–1096. doi: 10.1080/23802359.2019.1586483
- Lidén, M. (1986). Synopsis of Fumarioideae with a monograph of the tribe Fumarieae. *Opera Bot.* 88, 1–133. doi: 10.1017/s0264180100000217
- Lidén, M., Fukuhara, T., and Axberg, T. (1995). Phylogeny of *Corydalis*, ITS and morphology. *Plant Syst. Evol.* 9, 183–188. doi: 10.1007/978-3-7091-6612-3\_17

- Lidén, M., Fukuhara, T., Rylander, J., and Oxelman, B. (1997). Phylogeny and classification of Fumariaceae, with emphasis on *Dicentra* s. l., based on the chloroplast gene *rps16* intron. *Plant Syst. Evol.* 206, 411–420. doi: 10.1007/BF00987960
- Lin, C. S., Chen, J. J., Huang, Y. T., Chan, M. T., Daniell, H., Chang, W. J., et al. (2015). The location and translocation of *ndh* genes of chloroplast origin in the Orchidaceae family. *Sci. Rep.* 5:9040. doi: 10.1038/srep09040
- Liu, H., He, J., Ding, C., Lyu, R., Pei, L., Cheng, J., et al. (2018). 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. doi: 10.3389/fpls.2018.01097
- Liu, Q., Li, X., Li, M. Z., Xu, W. K., Schwarzacher, T., and Heslop-Harrison, J. S. (2020). Comparative chloroplast genome analyses of *Avena*: insights into evolutionary dynamics and phylogeny. *BMC Plant Biol.* 20:406. doi: 10.1186/s12870-020-02621-y
- Lowe, T. M., and Chan, P. P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 44, W54–W57. doi: 10.1093/nar/gkw413
- Luo, D. S., Feng, C. H., and Xia, G. C. (1984). The resources of the Tibetan drugs in Qinghai-Xizang Plateau — Preliminary studies on the plants of *Corydalis*. *Zhong Cao Yao* 15, 33–36.
- Ma, J., Yang, B., Zhu, W., Sun, L., Tian, J., and Wang, X. (2013). The complete chloroplast genome sequence of *Mahonia bealei* (Berberidaceae) reveals a significant expansion of the inverted repeat and phylogenetic relationship with other angiosperms. *Gene* 528, 120–131. doi: 10.1016/j.gene.2013.07.037
- Ma, P. F., Zhang, Y. X., Zeng, C. X., Guo, Z. H., and Li, D. Z. (2014). Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (Poaceae). *Syst. Biol.* 63, 933–950. doi: 10.1093/sysbio/syu054
- Magee, A. M., Aspinall, S., Rice, D. W., Cusack, B. P., Semon, M., Perry, A. S., et al. (2010). Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Res.* 20, 1700–1710. doi: 10.1101/gr.111955.110
- Martin, G. E., Rousseau-Guettin, M., Cordonnier, S., Lima, O., Michon-Coudouel, S., Naquin, D., et al. (2014). The first complete chloroplast genome of the Genistoid 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. doi: 10.1093/aob/mcu050
- Martin, M., and Sabater, B. (2010). Plastid *ndh* genes in plant evolution. *Plant Physiol. Biochem.* 48, 636–645. doi: 10.1016/j.plaphy.2010.04.009
- Michelangeli, F. A., Davis, J. I., and Stevenson, D. W. (2003). 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. doi: 10.3732/ajb.90.1.93
- Milligan, B. G., Hampton, J. N., and Palmer, J. D. (1989). Dispersed repeats and structural reorganization in subclover chloroplast DNA. *Mol. Biol. Evol.* 6, 355–368. doi: 10.1093/oxfordjournals.molbev.a040558
- Moore, M. J., Bell, C. D., Soltis, P. S., and Soltis, D. E. (2007). Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19363–19368. doi: 10.1073/pnas.0708072104
- Moore, M. J., Dhingra, A., Soltis, P. S., Shaw, R., Farmerie, W. G., Foltá, K. M., et al. (2006). Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biol.* 6:17. doi: 10.1186/1471-2229-6-17
- Moore, M. J., Soltis, P. S., Bell, C. D., Burleigh, J. G., and Soltis, D. E. (2010). Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4623–4628. doi: 10.1073/pnas.0907801107
- Mower, J. P., and Vickrey, T. L. (2018). Structural diversity among plastid genomes of land plants. *Adv. Bot. Res.* 85, 263–292. doi: 10.1016/bs.abr.2017.11.013
- Nei, M., Suzuki, Y., and Nozawa, M. (2010). The neutral theory of molecular evolution in the genomic Era. *Annu. Rev. Genomics Hum. Genet.* 11, 265–289. doi: 10.1146/annurev-genom-082908-150129
- Ogihara, Y., Terachi, T., and Sasakuma, T. (1988). Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. *Proc. Natl. Acad. Sci. U.S.A.* 85, 8573–8577. doi: 10.1073/pnas.85.22.8573
- Palmer, J. D. (1983). Chloroplast DNA exists in two orientations. *Nature* 301, 92–93. doi: 10.1038/301092a0
- Palmer, J. D. (1985). Comparative organization of chloroplast genomes. *Annu. Rev. Genet.* 19, 325–354. doi: 10.1146/annurev.ge.19.120185.001545
- Palmer, J. D., Nugent, J. M., and Herbon, L. A. (1987). Unusual structure of geranium chloroplast DNA: a triple-sized repeat, extensive gene duplications, multiple inversions, and new repeat families. *Proc. Natl. Acad. Sci. U.S.A.* 84, 769–773. doi: 10.1073/pnas.84.3.769
- Palmer, J. D., and Thompson, W. F. (1981). Rearrangements in the chloroplast genomes of mung bean and pea. *Proc. Natl. Acad. Sci. U.S.A.* 78, 5533–5537. doi: 10.1073/pnas.78.9.5533
- Palmer, J. D., and Thompson, W. F. (1982). Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell* 29, 537–550. doi: 10.1016/0092-8674(82)90170-2
- Park, S., An, B., and Park, S. (2018). Reconfiguration of the plastid genome in *Lamprocapnos spectabilis*: IR boundary shifting, inversion, and intraspecific variation. *Sci. Rep.* 8:13568. doi: 10.1038/s41598-018-31938-w
- Pérez-Gutiérrez, M. A., Romero-García, A. T., Fernández, M. C., Blanca, G., Salinas-Bonillo, M. J., and Suárez-Santiago, V. N. (2015). Evolutionary history of fumitories (subfamily fumarioideae, papaveraceae): an old story shaped by the main geological and climatic events in the northern hemisphere. *Mol. Phylogenet. Evol.* 88, 75–92. doi: 10.1016/j.ympev.2015.03.026
- Qu, X. J., Moore, M. J., Li, D. Z., and Yi, T. S. (2019). PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods* 15:50. doi: 10.1186/s13007-019-0435-7
- Ren, F. M., Wang, L. Q., Li, Y., Zhuo, W., Xu, Z. C., Guo, H. J., et al. (2020). Highly variable chloroplast genome from two endangered Papaveraceae lithophytes *Corydalis saxicola* and *C. tomentella*. *Res. Square* [Preprint]. doi: 10.21203/rs.3.rs-18411/v1
- Ren, F. M., Wang, Y. W., Xu, Z. C., Li, Y., Xin, T. Y., Zhou, J. G., et al. (2018). DNA barcoding of *Corydalis*, the most taxonomically complicated genus of Papaveraceae. *Ecol. Evol.* 9, 1934–1945. doi: 10.1002/ece3.4886
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. doi: 10.1038/nbt.1754
- Ronquist, F., Teslenk, M., Van der Mark, P., Ayres, D., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sy029
- Röschenbleck, J., Wicke, S., Weinl, S., Kudla, J., and Müller, K. F. (2016). Genus-wide screening reveals four distinct types of structural plastid genome organization in *Pelargonium* (Geraniaceae). *Genome Biol. Evol.* 9, 64–76. doi: 10.1093/gbe/evw271
- Rousseau, G. M., Huang, X., Higginson, E., Ayliffe, M., Day, A., and Timmis, J. N. (2013). 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. doi: 10.1104/pp.113.21.4528
- Ruhlman, T. A., and Jansen, R. K. (2014). “The plastid genomes of flowering plants,” in *Chloroplast Biotechnology: Methods and Protocols*, ed. P. Maliga (New York, NY: Springer), 3–38. doi: 10.1007/978-1-62703-995-6\_1
- Sablok, G., Amiryousefi, A., He, X., Hyvönen, J., and Poczar, P. (2019). Sequencing the plastid genome of giant ragweed (*Ambrosia trifida*, Asteraceae) from a herbarium specimen. *Front. Plant Sci.* 10:218. doi: 10.3389/fpls.2019.00218
- Sasaki, Y., and Nagano, Y. (2004). Plant acetyl-CoA carboxylase: structure, biosynthesis, regulation, and gene manipulation for plant breeding. *Biosci. Biotechnol. Biochem.* 68, 1175–1184. doi: 10.1271/bbb.68.1175
- Sauquet, H., Carrive, L., Poullain, N., Sannier, J., Damerval, C., and Nadot, S. (2015). Zygomorphy evolved from disymmetry in Fumarioideae (Papaveraceae, Ranunculales): new evidence from an expanded molecular phylogenetic framework. *Ann. Bot.* 115, 895–914. doi: 10.1093/aob/mcv020
- Schwarz, E. N., Ruhlman, T. A., Sabir, J. S. M., Hajrah, N. H., Alharbi, N. S., Al-Malki, A. L., et al. (2015). Plastid genome sequences of legumes reveal parallel inversions and multiple losses of *rps16* in papilionoids. *J. Syst. Evol.* 53, 458–468. doi: 10.1111/jse.12179
- Shi, L. C., Chen, H. M., Jiang, M., Wang, L. Q., Wu, X., Huang, L. F., et al. (2019). CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* 47, W65–W73. doi: 10.1093/nar/gkz345
- Stamatakis, A. (2014). RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

- Sun, Y. X., Moore, M. J., Lin, N., Adelalu, K. F., Meng, A., Jian, S., et al. (2017). Complete plastome sequencing of both living species of *Circaeasteraceae* (Ranunculales) reveals unusual rearrangements and the loss of the *ndh* gene family. *BMC Genomics* 18:592. doi: 10.1186/s12864-017-3956-3
- Sun, Y. X., Moore, M. J., Meng, A., Soltis, P. S., Soltis, D. E., Li, J., et al. (2013). Complete plastid genome sequencing of trochodendraceae reveals a significant expansion of the inverted repeat and suggests a paleogene divergence between the two extant species. *PLoS One* 8:e60429. doi: 10.1371/journal.pone.0060429
- Sun, Y. X., Moore, M. J., Zhang, S. J., Soltis, P. S., Soltis, D. E., Zhao, T. T., et al. (2016). Phylogenomic and structural analyses of 18 complete plastomes across nearly all families of early-diverging eudicots, including an angiosperm-wide analysis of IR gene content evolution. *Mol. Phylogenet. Evol.* 96, 93–101. doi: 10.1016/j.ympev.2015.12.006
- Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34, W609–W612. doi: 10.1093/nar/gkl315
- Swofford, D. L. (2003). *PAUP. Phylogenetic Analysis Using Parsimony (\*and Other Methods)*. Version 4. Sunderland: Sinauer Associates Press.
- Thiel, T., Michalek, W., Varshney, R., and Graner, A. (2003). Exploiting EST databases for the development and characterization of genederived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 106, 411–422. doi: 10.1007/s00122-002-1031-0
- Uribe-Convers, S., Carlsen, M. M., Lagomarsino, L. P., and Muchhala, N. (2017). Phylogenetic relationships of *Burmeistera* (Campanulaceae: Lobelioideae): Combining whole plastome with targeted loci data in a recent radiation. *Mol. Phylogenet. Evol.* 107, 551–563. doi: 10.1016/j.ympev.2016.12.011
- Wang, D. P., Zhang, Y. B., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs\_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genom. Proteom. Bioinform.* 8, 77–80. doi: 10.1016/S1672-0229(10)60008-3
- Wang, W., and Lanfear, R. (2019). Long-reads reveal that the chloroplast genome exists in two distinct versions in most plants. *Genome Biol. Evol.* 11, 3372–3381. doi: 10.1093/gbe/evz256
- Wang, W., Lu, A. M., Ren, Y., Endress, M. E., and Chen, Z. D. (2009). Phylogeny and classification of Ranunculales: evidence from four molecular loci and morphological data. *Perspect. Plant Ecol. Evol. Syst.* 11, 81–110. doi: 10.1016/j.ppees.2009.01.001
- Wang, Y. H., Qu, X. J., Chen, S. Y., Li, D. Z., and Yi, T. S. (2017). Plastomes of Mimosoideae structural and size variation, sequence divergence, and phylogenetic implication. *Tree Genet. Genomes* 13:41. doi: 10.1007/s11295-017-1124-1
- Wang, Y. W. (2006). *Systematics of Corydalis DC. (Fumariaceae)*. Ph.D thesis, Institute of Botany, the Chinese Academy of Sciences, Beijing.
- Weng, M. L., Blazier, J. C., Govindu, M., and Jansen, R. K. (2014). 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. doi: 10.1093/molbev/mst257
- Weng, M. L., Ruhlman, T. A., and Jansen, R. K. (2017). Expansion of inverted repeat does not decrease substitution rates in *Pelargonium* plastid genomes. *New Phytol.* 214, 842–851. doi: 10.1111/nph.14375
- Wicke, S., Schneeweiss, G. M., dePamphilis, C. W., Müller, K. F., and Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol. Biol.* 76, 273–297. doi: 10.1007/s11103-011-9762-4
- Wu, J., Lin, P. C., Guo, Y. P., and Liu, M. D. (2020). The complete chloroplast genome of *Corydalis conspersa*. *Mitochondrial DNA B Resour.* 5, 1977–1978. doi: 10.1080/23802359.2020.1756944
- Wu, Z. Y., Zhuang, X., and Su, Z. Y. (1996). The systematic evolution of *Corydalis* in relation to florogenesis and floristic regionalization in the world. *Acta Bot. Yunnan.* 18, 241–267.
- Wu, Z. Y., Zhuang, X., and Su, Z. Y. (1999). “*Corydalis* DC.” in *Flora Reipublicae Popularis Sinicae, Tomus 32, ed. Delecti Flora Reipublicae Popularis Sinicae Agendae Academiae Sinicae* (Beijing: Science Press), 106–479 [in Chinese].
- Xu, X. D., and Wang, D. (2018). *Corydalis ternatifolia* belongs to *C. sect. Asterostigmata*, not *C. sect. Incisae* (Papaveraceae): Evidence from morphological and phylogenetic study. *Phytotaxa* 382, 193–203. doi: 10.11646/phytotaxa.382.2.4
- Xu, X. D., and Wang, D. (2020). Characterization of the complete chloroplast genome of *Corydalis inopinata* Prain ex Fedde (Papaveraceae). *Mitochondrial DNA B Resour.* 5, 3302–3303. doi: 10.1080/23802359.2020.1814887
- Yang, X., Xie, D. F., Chen, J. P., Zhou, S. D., Yu, Y., and He, X. J. (2020). Comparative analysis of the complete chloroplast genomes in *Allium* Subgenus *Cyathophora* (Amaryllidaceae): phylogenetic relationship and adaptive evolution. *Biomed Res. Int.* 2020:1732586. doi: 10.1155/2020/1732586
- Zeng, C. X., Hollingsworth, P. M., Yang, J., He, Z. S., Zhang, Z. R., Li, D. Z., et al. (2018). Genome skimming herbarium specimens for DNA barcoding and phylogenomics. *Plant Methods* 14:43. doi: 10.1186/s13007-018-0300-0
- Zhai, W., Duan, X. S., Zhang, R., Guo, C. C., Li, L., Xu, G. X., et al. (2019). Chloroplast genomic data provide new and robust insights into the phylogeny and evolution of the Ranunculaceae. *Mol. Phylogenet. Evol.* 135, 12–21. doi: 10.1016/j.ympev.2019.02.024
- Zhang, B., Huang, R. Z., Hua, J., Liang, H., Pan, Y. M., Dai, L. M., et al. (2016). Antitumor lignanamides from the aerial parts of *Corydalis saxicola*. *Phytomedicine* 23, 1599–1609. doi: 10.1016/j.phymed.2016.09.006
- Zhang, M. L., Su, Z. Y., and Lidén, M. (2008). “*Corydalis* DC.” in *Flora of China, Vol. 7*, eds Z. Y. Wu, P. H. Raven, and D. Y. Hong (Beijing: Science Press), 295–427.
- Zhang, Y., Lee, J., Liu, X., and Sun, Z. (2019). The first complete chloroplast genome of *Hylomecon japonica* and its phylogenetic position within Papaveraceae. *Mitochondrial DNA B Resour.* 4, 2349–2350. doi: 10.1080/23802359.2019.1573125
- Zhang, Z. X., Wang, D., and Yang, X. (2016). The taxonomic position of *Corydalis parviflora* Su & Lidén (Papaveraceae), a genetically distinct species: evidence from cpDNA and nDNA sequences. *Biochem. Syst. Ecol.* 67, 134–141. doi: 10.1016/j.bse.2016.06.003
- Zhu, A. D., Guo, W. H., Gupta, S., Fan, W. S., and Mower, J. P. (2016). Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytol.* 209, 1747–1756. doi: 10.1111/nph.13743

**Conflict 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.

Copyright © 2021 Xu and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.