

RESEARCH

Open Access



Characterization of the *Dicranostigma leptopodum* chloroplast genome and comparative analysis within subfamily Papaveroideae

Lei Wang^{1*}, Fuxing Li¹, Ning Wang¹, Yongwei Gao², Kangjia Liu², Gangmin Zhang² and Jiahui Sun^{3*}

Abstract

Background: *Dicranostigma leptopodum* (Maxim.) Fedde is a perennial herb with bright yellow flowers, well known as "Hongmao Cao" for its medicinal properties, and is an excellent early spring flower used in urban greening. However, its molecular genomic information remains largely unknown. Here, we sequenced and analyzed the chloroplast genome of *D. leptopodum* to discover its genome structure, organization, and phylogenomic position within the subfamily Papaveroideae.

Results: The chloroplast genome size of *D. leptopodum* was 162,942 bp, and *D. leptopodum* exhibited a characteristic circular quadripartite structure, with a large single-copy (LSC) region (87,565 bp), a small single-copy (SSC) region (18,759 bp) and a pair of inverted repeat (IR) regions (28,309 bp). The *D. leptopodum* chloroplast genome encoded 113 genes, including 79 protein-coding genes, 30 tRNA genes, and four rRNA genes. The dynamics of the genome structures, genes, IR contraction and expansion, long repeats, and single sequence repeats exhibited similarities, with slight differences observed among the eight Papaveroideae species. In addition, seven interspace regions and three coding genes displayed highly variable divergence, signifying their potential to serve as molecular markers for phylogenetic and species identification studies. Molecular evolution analyses indicated that most of the genes were undergoing purifying selection. Phylogenetic analyses revealed that *D. leptopodum* formed a clade with the tribe Chelidoniae.

Conclusions: Our study provides detailed information on the *D. leptopodum* chloroplast genome, expanding the available genomic resources that may be used for future evolution and genetic diversity studies.

Keywords: Chloroplast genome, *Dicranostigma leptopodum*, Papaveroideae, Phylogeny, Genetic diversity

Introduction

Dicranostigma leptopodum (Maxim.) Fedde within Papaveraceae is a perennial herb with bright yellow flowers that is endemic to China and is well known as "Hongmao Cao" due to its medicinal properties [1–3]. The distribution of the species ranges from southwest and northwest to central China, and its fluorescence occurs from March to May, lasting for one or two months (Fig. 1). The basal leaves of *D. leptopodum* are dense and may be evergreen throughout the winter, with a capacity to retain moisture and fertilizer in soil [4, 5]. Some studies have shown that

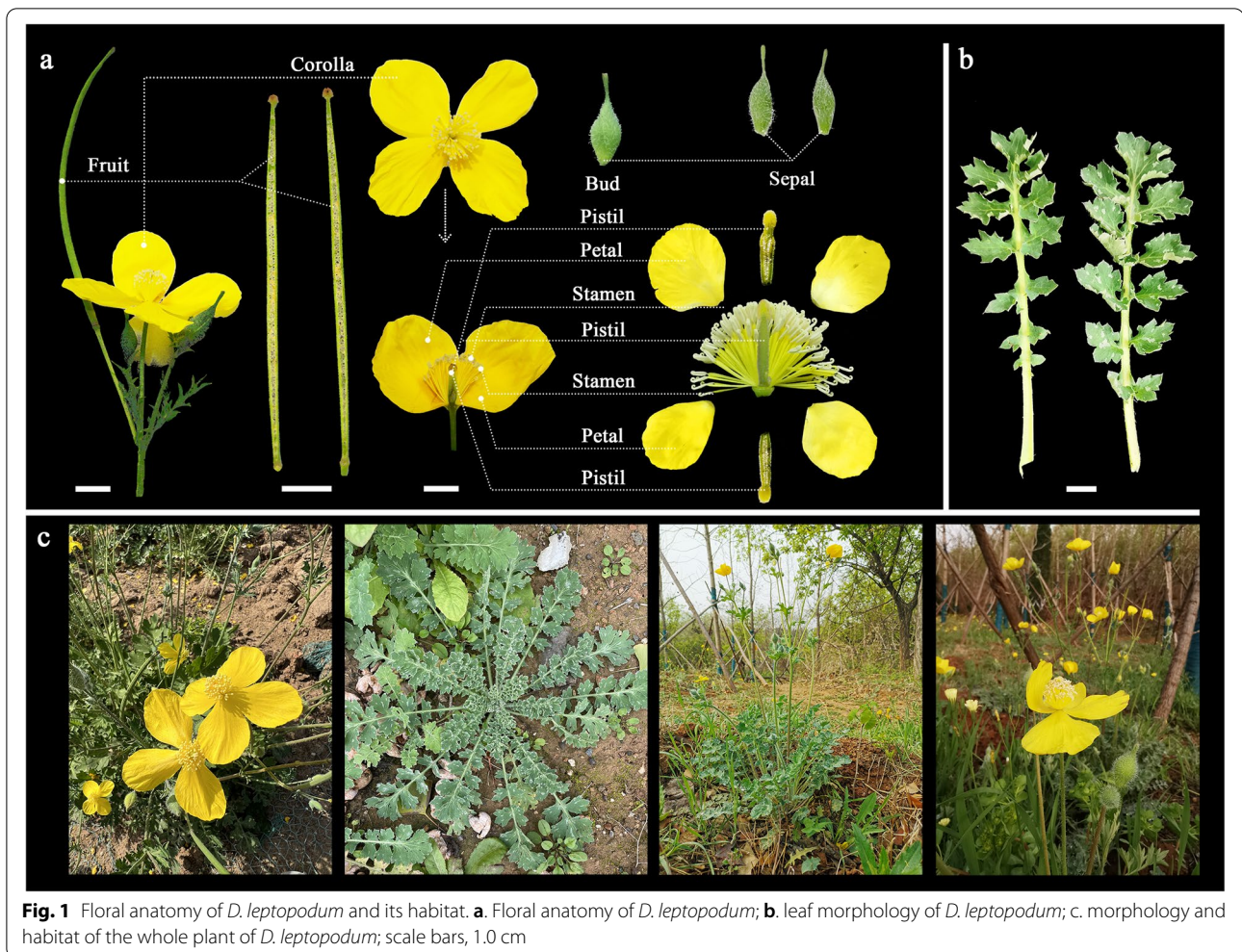
*Correspondence: wlhaust@126.com; sunjh_2010@sina.com

¹ College of Horticulture and Plant Protection, Henan University of Science and Technology, Luoyang 471023, Henan, China

³ State Key Laboratory Breeding Base of Dao-di Herbs, National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China

Full list of author information is available at the end of the article





D. leptopodum has a certain tolerance for and ability to enrich elements such as Cd, Zn, Cu, and Pb and can be used as a plant for remediating heavy-metal-contaminated soils in mining areas, farmland, and rivers [6]. Therefore, from the perspectives of both ornamental characteristics and ecological restoration, *D. leptopodum* and its related species are rare, excellent early spring flowering resources for urban greening [7].

The phylogenetic position of *Dicranostigma* remains unclear. Some classical taxonomic studies suggest that *Dicranostigma* and *Glaucium* are sister groups within Papaveraceae [8]; however, the latest molecular evidence shows that *Dicranostigma* and *Glaucium* are still sister groups but within Chelidoniae [9], which is inconsistent with some previous findings based on morphological evidence [10]. Thus, the phylogenetic position of *Dicranostigma* remains controversial.

The chloroplast genome, which is characterized primarily by maternal inheritance, has been widely used in the reconstruction of phylogenetic relationships at

different taxonomic levels [11–15] and species identification [16] due to its highly conserved nature in terms of stable structure, gene arrangement, GC (guanine and cytosine) content, and lack of recombination during genetic processes. To date, most of the studies on *Dicranostigma* have focused mainly on the medicinal properties of its extracts [1, 2], ecophysiological index, and seed germination characteristics [5]; in contrast, almost no research has been conducted on its phylogeny. Additionally, only a few molecular sequences of species within *Dicranostigma* are recorded in the NCBI database, substantially restricting phylogenetic research and the development and utilization of *D. leptopodum*.

Therefore, this study intends to comprehensively reveal the phylogeny of *D. leptopodum* and its relatives using comparative chloroplast genomic and molecular phylogenetic methods. The main research objectives are as follows: (1) to comprehensively analyze the structural characteristics of the *D. leptopodum* chloroplast genome, (2) to identify similarities and differences in the structural

characteristics of the chloroplast genome of *D. leptopodium* and its relatives, and (3) to reveal the phylogenetic position of *Dicranostigma* within Papaveraceae.

Materials and methods

Plant material, DNA extraction, and genome sequencing

The *D. leptopodium* material was obtained from Luoyang, Henan, China. The specimens were subsequently deposited in the Herbarium of the College of Horticulture and Plant Protection, Henan University of Science and Technology (Voucher: WL1000). These specimens were identified by Dr. Ning Wang at Henan University of Science and Technology. Fresh leaves were preserved in silica gel. Total DNA was extracted using a modified CTAB method [17] and detected by electrophoresis on 0.8% agarose gels. Next, library preparation and next-generation sequencing (Illumina, Nova PE150 sequencing strategy) were conducted at Novogene Biotechnology Co., Ltd. (Tianjin, China). The total sequencing data were 4 Gb.

Chloroplast genome assembly and annotation

Trimmomatic v 0.39 [18] was used to filter the original reads obtained by sequencing and remove those of low quality at the joints and ends. The chloroplast genome was assembled using GetOrganelle v 1.7.6.1 [19] with the default parameters. Plastid Genome Annotator (PGA) software [20] was used to annotate the chloroplast genome with the reference of *Coreanomecon hylomeconoides* Nakai (KT274030). Furthermore, the results annotated by PGA were checked using Geneious Prime v 2021 [21] and Sequin v 9.20 [22] to ensure the accuracy of the annotation results. Some genes with high sequence divergence and genes with introns received special attention, such as *accD*, *petB*, *petD*, *rps16*, *rpl16*, and *ycf1*. We manually checked the annotations of these genes. The chloroplast genome map of *D. leptopodium* was drawn and visualized using OGDRAW v 1.3.1 [23]. The complete annotated sequence was submitted to GenBank with accession number OM994890. Codon usage and relative synonymous codon usage (RSCU) analyses were performed using the MEGA-X program [24] to analyze the codon usage bias of the *D. leptopodium* chloroplast genome.

Comparative analyses of chloroplast genome structures

We compared the chloroplast genome structures within species of the subfamily Papaveroideae. Eight samples of *Dicranostigma leptopodium* (OM994890), *Stylophorum lasiocarpum* (Oliv.) Fedde (MW232434), *Papaver nudicaule* L. (MW411801), *Meconopsis horridula* Hook. f. & Thomson (MK533646), *Macleaya cordata* (Willd.) R. Br. (MT178411), *Hylomecon japonica* (Thunb.)

Prantl & Kündig (MK251463), *Coreanomecon hylomeconoides* Nakai (KT274030), and *Chelidonium majus* L. (MK433200) were used. The mVISTA online program (<https://genome.lbl.gov/vista/mvista/submit.shtml>) was used to compare the genome structures and the sequence similarity of the chloroplast genomes of eight Papaveroideae species. The annotated *D. leptopodium* chloroplast genome was used as a reference. The variation in LSC/IRb/SSC/IRa junction regions was compared using IRscope [25] (<https://irscope.shinyapps.io/IRapp/>).

Repeat sequences were analyzed using the REPuter online program [26] and Perl script MISA [27]. REPuter was used to identify four types of long repeats (forward, reverse, palindromic, and complementary) in eight Papaveroideae chloroplast genomes with a Hamming distance of 3 and a minimum repeat size of 30 bp. MISA was employed to identify the simple sequence repeats (SSRs or microsatellites). Six types of SSRs (mono-, di-, tri-, tetra-, penta-, and hexanucleotides) were analyzed, and the minimum numbers of SSRs were set to 10, 5, 4, 3, 3, and 3, respectively.

Analysis of the nucleotide substitution rate of protein-coding genes

The molecular evolution between *D. leptopodium* and seven other Papaveroideae species was investigated. The values of dN (nonsynonymous substitutions), dS (synonymous substitutions), and ω (dN/dS ratio) were estimated using the YN100 module in PAMLX [28]. All 79 protein-coding genes were extracted from the annotated *D. leptopodium* chloroplast genome. Three strategies were used to infer the molecular evolution: (1) calculate each coding gene for all of the species in a pairwise manner; (2) calculate gene groups with the same function [29], such as *rps*, *pet*, and *psa*; and (3) calculate groups at the species level by combining all 79 coding genes for each species.

Phylogenetic inference

The newly assembled chloroplast genome of *D. leptopodium* and all the published complete chloroplast genomes of the family Papaveraceae were downloaded from GenBank [22]. All 79 protein-coding genes and the four rRNA genes were extracted from the results of the chloroplast genome annotation. All genes were aligned using MAFFT v 7.450 [30].

Maximum likelihood (ML) and Bayesian inference (BI) methods were used to infer the phylogenetic relationships. The ML tree was constructed using RAxML-NG [31] with the GTR+G model, and node support was assessed with 1000 bootstrap inferences. The BI tree was inferred with MrBayes v 3.2 [32]. The Markov chain Monte Carlo algorithm was performed with two independent runs of four simultaneous chains with a random

starting tree and default priors for 5,000,000 generations, with every 1,000 generations used for tree sampling. Tracer v1.6 was used to analyze the convergence to stationary distribution and the effective size of each parameter. The first 25% of generations were discarded as a burn-in. The remaining trees were used to build the BI tree of posterior probabilities.

Results

Assembly and general features of the *D. leptopodum* chloroplast genome

Using the Illumina sequencing platform, we obtained 6,536,061 clean reads. The coverage depth of the *D. leptopodum* chloroplast genome was 40 X. The complete *D. leptopodum* chloroplast genome was 162,942 bp, exhibiting a characteristic circular quadripartite structure. The *D. leptopodum* chloroplast genome consisted of a pair of inverted repeat (IR) regions (28,309 bp) separated by a larger single-copy (LSC) region (87,565 bp) and a small single-copy (SSC) region (18,759 bp). A circular map of the chloroplast genome is shown in Fig. 2.

One hundred thirteen unique gene annotations were identified in the *D. leptopodum* chloroplast genome, including 79 protein-coding genes, 30 tRNAs, and four rRNAs. Ten protein-coding genes, four rRNAs, and seven tRNAs were duplicated in the IR regions. The *ycf1* in IRb is a pseudogene formed due to the expansion of the IR region (Fig. 2). Forty-five of those genes play a role in photosynthesis, and 60 genes are associated with self-replication (Table 1). Ten protein-coding genes (*atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl2*, *rpl16*, *rpoC1*, *rps16*, and *rps12*) and six tRNA genes (*trnA-UGC*, *trnG-UCC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, and *trnV-UAC*) contained a single intron, whereas two genes (*clpP* and *ycf3*) contained two introns. The largest intron was observed in *trnK-UUU* (2,558 bp), which contains the *matK* gene. *rps12* is a trans-spliced gene with the 5' end located at the LSC region and the 3' end located in the IR region. The GC content of the complete chloroplast genome was 39.4%, with a higher GC content in the IR region (43.1%) than in the LSC (38.1%) and SSC (34.4%) regions.

Codon usage analysis

The *D. leptopodum* chloroplast genome encoded 26,915 codons, representing 20 amino acids and the stop codon (Fig. 3). Leucine was the most abundant amino acid, with a frequency of 10.30%, followed by isoleucine (8.12%) and serine (7.62%); tryptophan was less abundant, with a frequency of 1.81%. All 64 codon types were detected. Additionally, AUU showed a high number of occurrences (1,036), followed by AAA (996), GAA (996), and GAU (953).

We detected 31 codons with a relative synonymous codon usage (RSCU) value > 1, indicating their usage bias in the *D. leptopodum* chloroplast genome. The usage bias was toward A or T (U), with high RSCU values. The highest RSCU values were detected for AGA encoding arginine, followed by GCU encoding alanine and UAU encoding tyrosine. The start codon (ATG) and TGG, encoding tryptophan, exhibited no bias (RSCU = 1).

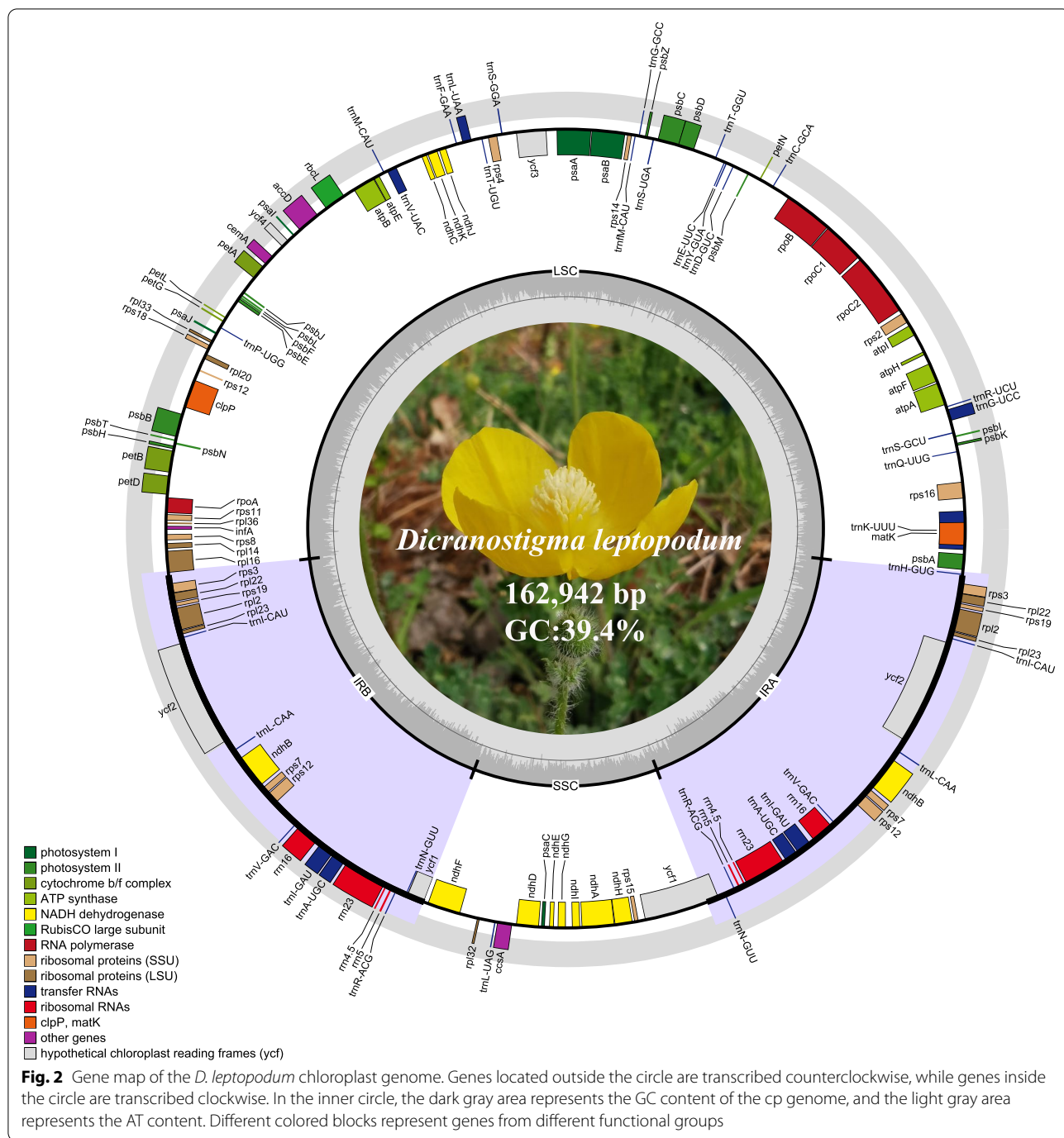
Repeat sequences and SSR analysis

The repeat sequences were investigated among eight Papaveroideae species. Long repeat sequences with a repeat unit longer than 30 bp were analyzed. A total of 210 long repeats (30–190 bp) were identified from the eight chloroplast genomes, consisting of 137 forward, 69 palindromic, two reverse, and one complementary repeat (Fig. 4a). For each chloroplast genome, the number of long repeats ranged from 16 (*Meconopsis horridula*) to 50 (*D. leptopodum*), and the numbers of forward and palindromic repeats were 8–46 and 4–11 (Figs. 4c and 4d), respectively. The two reverse repeats occurred in *Coreanomecon hylomeconoides* and *Hylomecon japonica*, and the only complementary repeat was detected in *Co. hylomeconoides*. Most long repeats were distributed in the noncoding areas (including intergenic and intron regions), and a few existed in shared genes, such as *accD*, *ycf1*, and *ycf2* (Fig. 4b and Table S1). The most frequent repeat size was 30–35 bp (51.92%) (Figs. 4c and 4d), and the longest repeat occurred in the *rbcl-accD* sequence of *D. leptopodum*.

The total number of SSRs in the eight chloroplast genomes was 382 (Fig. 4e), ranging from 20 (*Papaver nudicaule*) to 69 (*Macleaya cordata*). Most of the SSRs, with proportions from 55.0% (*Chelidonium majus*) to 74.2% (*Stylophorum lasiocarpum*) of the total number of SSRs, were distributed in the spacer regions. Mononucleotide repeat units were the most frequent type, and A/T repeat units were the most abundant repeats. AT/TA and ATT/TTA repeats were the most frequent repeat units for the dinucleotide and trinucleotide types. We found that the chloroplast genome SSRs in Papaveroideae species contained a high level of AT repeats. The SSRs were mainly located in the noncoding regions (79.58%) (Fig. 4f). We also detected SSRs in protein-coding gene regions, such as *rpoC2*, *ycf1*, and *cemA* (Table S2).

Analysis of the chloroplast genome structure

We compared the chloroplast genome of *D. leptopodum* to the genomes of seven other species from Papaveroideae. Our comparison indicated that all eight species had similar genome structures and encoded a consistent gene content (Table 2). All eight chloroplast genomes displayed a typical quadripartite structure,



and the genome size ranged from 152,867 bp (*P. nudicaule*) to 163,107 bp (*Ma. cordata*). The LSC regions ranged from 83,104–88,165 bp, and the SSC size was 17,898–18,759 bp. The GC content of the eight chloroplast genomes was very similar (38.7–39.4%) but differed by region. Specifically, the GC content in the IR regions (42.8–43.2%) was higher than the GC content in the LSC (37.2–38.1%) and SSC (32.8–34.4%) regions.

The Papaveroideae chloroplast genomes contained highly similar gene contents, consisting of 79 protein-coding genes, 30 tRNAs, and four rRNAs.

Multiple sequence alignments revealed no genomic rearrangements or large inversions in the eight Papaveroideae chloroplast genomes (Fig. 5). These chloroplast genomes were highly conserved both in gene order and identity. Notably, the coding and IR regions were more

Table 1 Genes in the *D. leptopodum* chloroplast genome

Category of Genes	Group of Genes	Name of Genes
Self-replication	Ribosomal proteins (LSU)	<i>rpl2, rpl14, rpl16, rpl20, rpl22, rpl23, rpl32, rpl33, rpl36</i>
	Ribosomal proteins (SSU)	<i>rps2, rps3, rps4, rps7, rps8, rps11, rps12, rps14, rps15, rps16, rps18, rps19</i>
	RNA polymerase	<i>rpoA, rpoB, rpoC1, rpoC2</i>
	Ribosomal RNAs	<i>rrn16, rrn23, rrn4.5, rrn5</i>
	Transfer RNAs	<i>trnA-UGC, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnG-UCC, trnH-GUG, trnI-CAU, trnI-GAU, trnK-UUU, trnL-CAA, trnL-UAA, trnL-UAG, trnM-CAU, trnM-CAU, trnN-GUU, trnP-UGG, trnQ-UUG, trnR-ACG, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC, trnV-UAC, trnW-CCA, trnY-GUA</i>
Genes for photosynthesis	Translational initiation factor	<i>infA</i>
	ATP synthase	<i>atpA, atpB, atpE, atpF, atpH, atpI</i>
	Protease	<i>clpP1</i>
	RubisCO large subunit	<i>rbcL</i>
	NADH dehydrogenase	<i>ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
	Cytochrome b/f complex	<i>petA, petB, petD, petG, petL, petN</i>
	Photosystem I	<i>psaA, psaB, psaC, psal, psaj</i>
Other genes	Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	Subunit of acetyl-CoA carboxylase	<i>accD</i>
	Maturase	<i>matK</i>
	C-type cytochrome synthesis gene	<i>ccsA</i>
	Envelope membrane protein	<i>cemA</i>
Genes of unknown function	Conserved open reading frames	<i>ycf1, ycf2, ycf3, ycf4</i>

conserved than noncoding and single-copy regions. The intergenic spacer regions with more variation were *rps16-trnQ*, *trnC-petN*, *rbcL-accD*, *ycf4-cemA*, *rps12-clpP*, *rpl16-trnS*, and *rpl32-trnL*. Moreover, the coding genes *accD*, *ndhF* and *ycf1* showed high levels of variation among the coding regions.

Multiple sequence alignment using mVISTA showed that the IR regions of the Papaveroideae chloroplast genomes were highly conserved, and variations in structure were observed in SC/IR boundary regions (Fig. 6). Among the eight chloroplast genomes, two types were identified in the SC/IR boundary. *D. leptopodum* and *Ma. cordata* had similar structures: *rpl16* was located at the boundaries of the LSC and IRb regions in these two species. In the other six species, these boundaries instead contained the *rps19* gene. This result indicated that expansion of the IR caused a duplication of *rps3*, *rpl22*, and *rps19* in *D. leptopodum* and *Ma. cordata* chloroplast genomes. One base pair was identified between the *rps19* and LSC/IRb boundaries in *P. nudicaule*. *Rps19* expanded into the IRb regions in five species, namely, *S. lasiocarpum*, *M. horridula*, *H. japonica*, *Co. hylomeconoides*, and *Ch. majus*, by 72, 72, 74, 72, and 74 bp, respectively. The *ycf1* gene crossed the SSC/IRa boundary in all species, and the length of *ycf1* in the

IRa region varied among the eight Papaveroideae species from 798 bp to 1,297 bp, which indicated dynamic variation in the SSC/IRa boundaries. Notable differences were observed among species in the IRa/LSC boundary. The *rps3* and *trnH* genes were located at this boundary in *D. leptopodum* and *Ma. cordata*, which instead contained the *rps19* and *trnH* genes in *S. lasiocarpum*, *Me. Horridula*, *H. japonica*, *Co. hylomeconoides*, and *Ch. majus*. *P. nudicaule* was the only species in which the *rps19* gene within the LSC was detected at the IRa/LSC boundary.

Analysis of selection pressure

The ω values were calculated for 79 single protein genes, gene groups, and combinations of all 79 coding genes for each of the eight Papaveroideae species, with *D. leptopodum* serving as the reference (Table S3 and Fig. 7). Among the 79 single protein-coding genes, most of the ω values were less than one, except for *rpl20* in *Ch. majus* (1.3599) and *Co. hylomeconoides* (1.2238); *rps7* in *Ch. majus* (3.4119), *Co. hylomeconoides* (3.4119), *H. japonica* (3.4119), *Ma. cordata* (3.4119), and *Me. Horridula* (1.242); and *ycf2* in *Ch. majus* (1.444), *Co. hylomeconoides* (1.4265), *H. japonica* (1.2824), and *Ma. cordata* (1.3469). The ω values of all eight gene groups were less than 0.5 (Fig. 7), indicating the presence of purifying selection



pressure on the gene groups of Papaveroideae species. At the species level, the ω values among the eight species were not significantly different, suggesting uniform evolution rates in the Papaveroideae species.

Phylogenetic inference of the family Papaveraceae

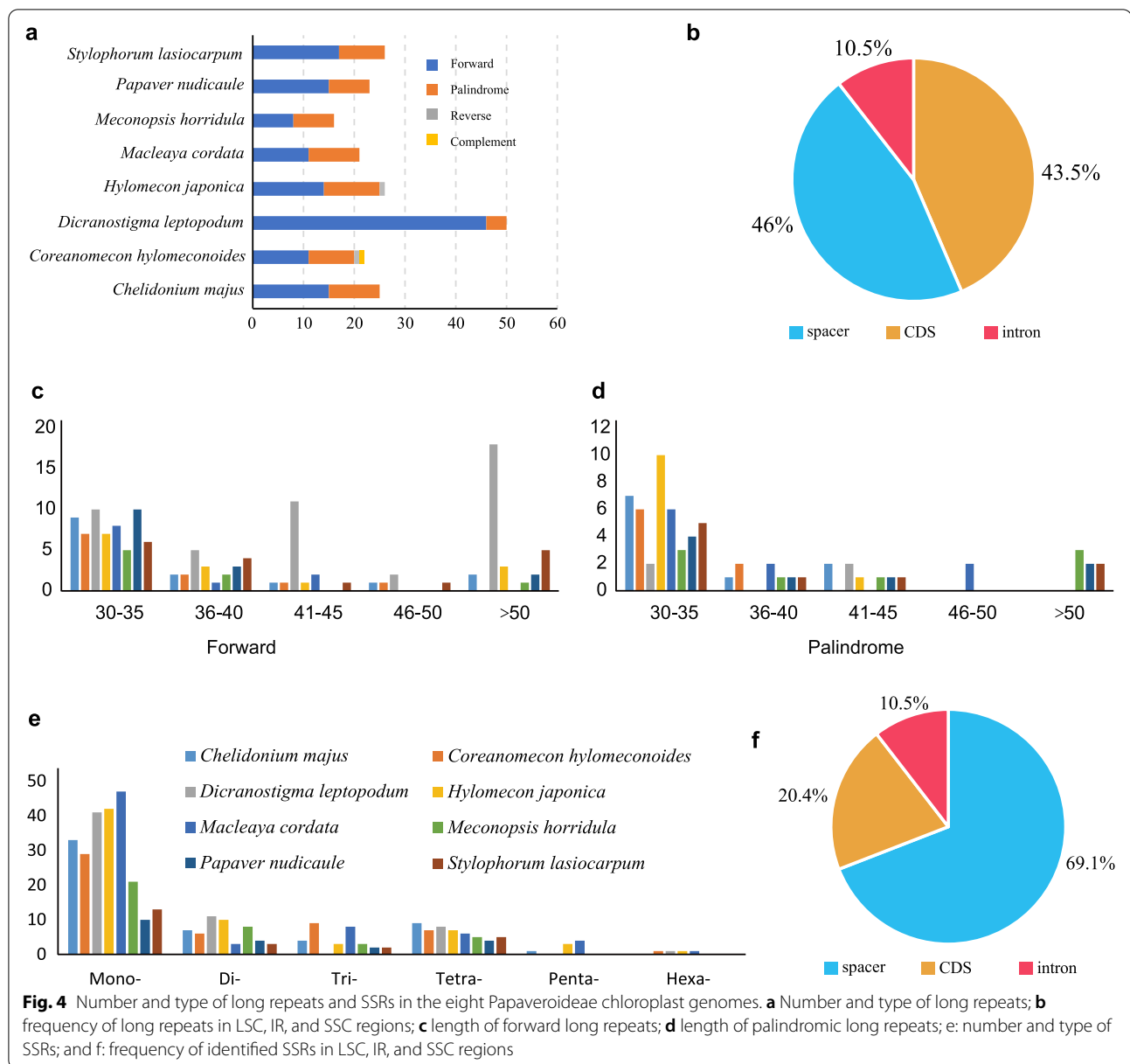
ML and BI trees were constructed based on 83 chloroplast genes to infer the phylogenetic position of *Dicranostigma* in the family Papaveraceae (Fig. 8). This dataset included 20 species of the subfamily Fumarioideae, 19 of the subfamily Papaveroideae, one species (*Eschscholzia californica*) of the subfamily Eschscholzioidae and six species (*Asteropyrum peltatum*, *Semiaquilegia Guangxiensis*, *Circaea agrestis*, *Sinofranchetia chinensis*, *Sargentodoxa cuneata*, and *Euptelea pleiosperma*) as the outgroups. Topological structures generated from the ML and BI analyses were consistent and presented highly resolved phylogenies. Papaveraceae was divided into two clades with strong support (bootstrap (BS) value = 100% for the ML tree and posterior probability (PP) = 1.00 for the BI tree), and subfamilies Papaveroideae and Eschscholzioidae formed a clade.

The subfamily Papaveroideae was shaped into two clades, including the tribe Papavereae and the tribe

Chelidoniaceae, both with strong support (BS/PP = 100/1). Our results showed that species of *Papaver* did not form a clade. *P. nudicaule* was sister to *S. lasiocarpum* and was not clustered in a clade with other *Papaver* species. The tribe Chelidoniaceae was further divided into three groups with strong support, each belonging to a different subtribe. The first group contained two *Macleaya* species, which belong to the subtribe Bocconiinae. The second group contained *Dicranostigma leptopodum*, which was clustered with the subtribe Chelidoniinae that contained three genera.

Discussion

In this study, we sequenced the chloroplast genome of *D. leptopodum* and applied it in comparative analyses with the available chloroplast genomes of the subfamily Papaveroideae. The complete chloroplast genome of *D. leptopodum* showed a typical quadripartite structure—with one LSC and one SSC region, as well as two IR regions—which was highly conserved, similar to the chloroplast genomes previously reported in the subfamily Papaveroideae [33, 34]. The genome size of *D. leptopodum* did not differ substantially from the available chloroplast genomes of Papaveroideae. However, the IR



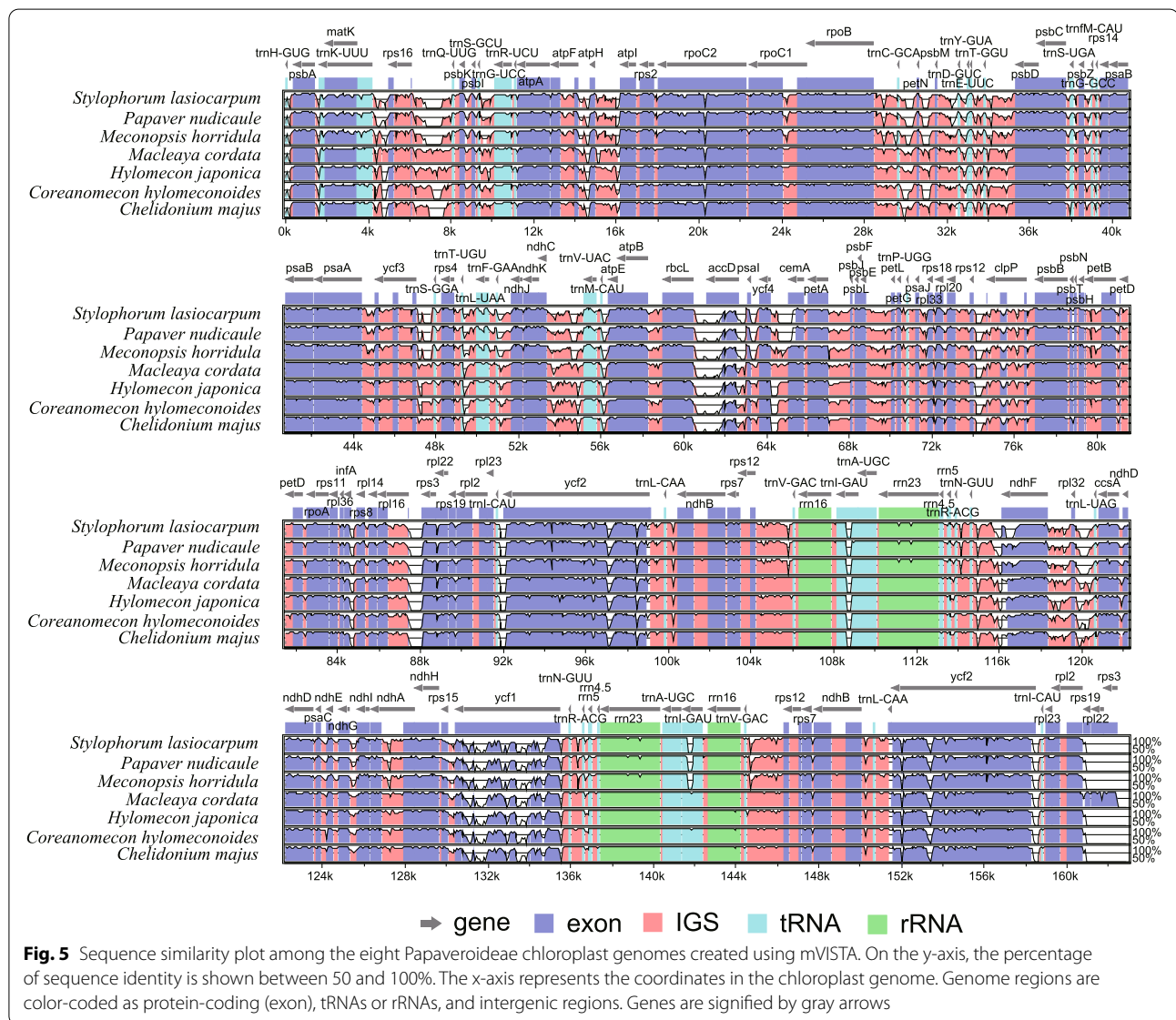
size of *D. leptopodum* was much larger than the IR sizes of the other species, except *Ma. cordata* (Table 2). The contractions and expansions at the borders of the IR regions are considered the most effective processes causing genome size variations (Fig. 6); furthermore, these contractions and expansions are important evolutionary events in the chloroplast genome, with effects on genome size. After comparing eight chloroplast genomes from Papaveroideae species, we noticed that the boundaries between the LSC and IRb regions were divided into two types (Fig. 6). The IR regions of *D. leptopodum* and *Ma. cordata* exhibited an obvious expansion, and those genomes had the largest IR sizes. More genes had

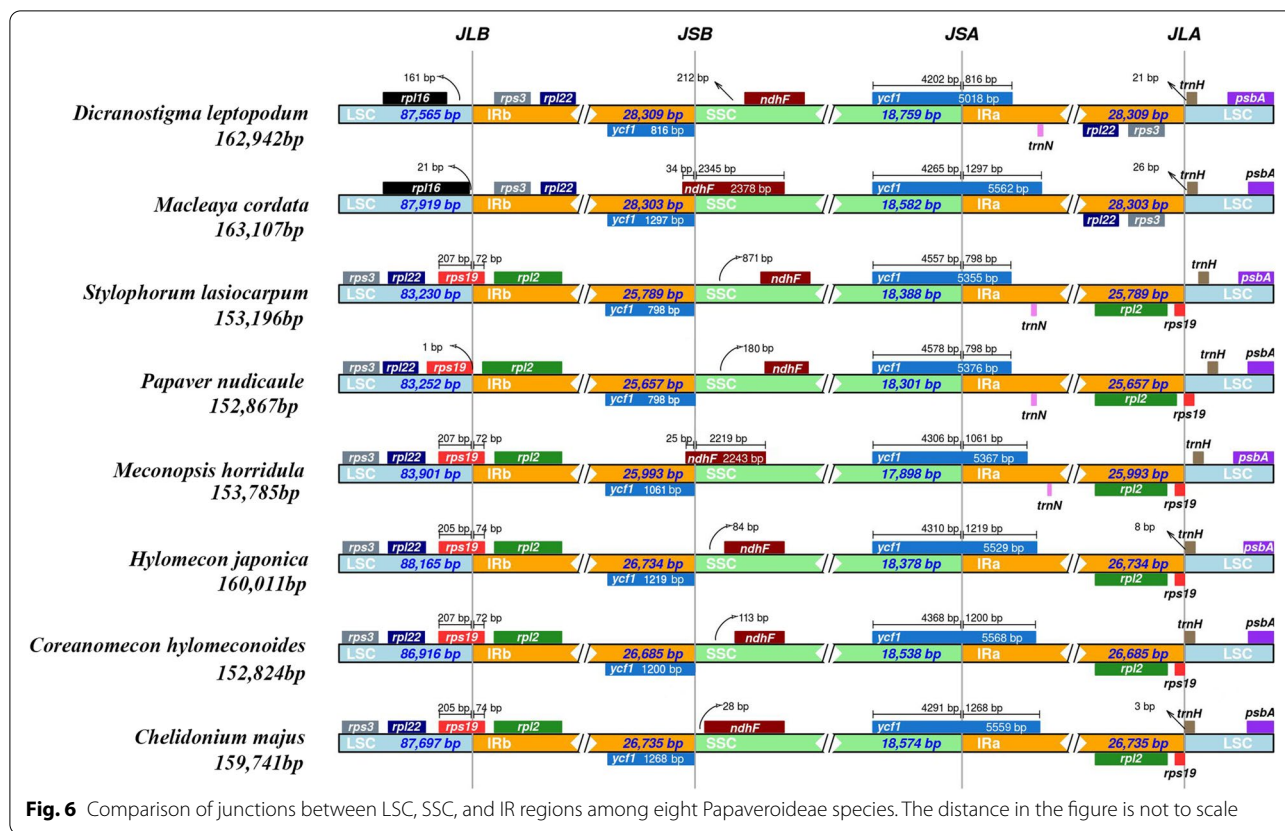
changed from a single copy to two copies. We also evaluated the IRa and LSC junctions, and the distributions and locations of genes in these regions were highly variable. Therefore, changes in the SC/IR boundaries may be the main contributors to genome size variation, especially in the IR regions, in Papaveroideae species.

Similar to typical angiosperm chloroplast genomes [35, 36], *D. leptopodum* shared a high sequence similarity with the other Papaveroideae species. However, some regions of these chloroplast genomes exhibited high sequence divergence. According to the mVISTA results, the sequence divergence of the IR region was lower than the sequence divergences of the LSC and SSC regions

Table 2 Features of the chloroplast genomes of *D. leptopodum* and seven Papaveroideae species

Species	<i>Dicranostigma leptopodum</i>	<i>Macleaya cordata</i>	<i>Stylophorum lasiocarpum</i>	<i>Papaver nudicaule</i>	<i>Meconopsis horridula</i>	<i>Hylomecon japonica</i>	<i>Coreanomecon hylomeconoides</i>	<i>Chelidonium majus</i>
Length (bp)	162,942	163,107	153,196	152,867	153,785	160,011	158,824	159,741
LSC (bp)	87,565	87,919	83,230	83,104	83,899	88,165	86,916	87,697
SSC (bp)	18,759	18,582	18,388	18,301	17,898	18,378	18,538	18,574
IR (bp)	28,309	28,303	25,789	25,731	25,994	26,734	26,685	26,735
LSC (GC%)	38.1	37.0	37.4	37.4	37.2	37.4	32.8	37.2
SSC (GC%)	34.4	33.1	34.0	33.5	33.2	33.2	37.3	33.3
IR (GC%)	43.1	42.8	43.2	43.2	43.1	43.2	43.2	43.1
GC (Total%)	39.4	38.6	38.9	38.9	38.8	38.8	38.7	38.7
Genes	113	113	113	113	113	113	113	113
Coding genes	79	79	79	79	79	79	79	79
tRNA	30	30	30	30	30	30	30	30
rRNA	4	4	4	4	4	4	4	4





(Fig. 5), due to the sequence correction of the two copies in the processes of gene replication and transcription. The other reason for the conservation of the IR regions is that they play an important role in maintaining the conservation and stability of the chloroplast genome structure [37, 38]. We identified several intergenic regions and genes with high sequence divergences, such as *rps16-trnQ*, *trnC-petN*, *rbcL-accD*, *ycf4-cemA*, *rps12-clpP*, *rpl16-trnS*, *rpl32-trnL*, *accD*, *ndhF*, and *ycf1* (Fig. 5). These high sequence divergence regions have previously been reported in other lineages. For example, Dong et al. [39, 40] compared several pairs of chloroplast genome sequences and identified *rps16-trnQ*, *trnC-petN*, *rbcL-accD*, *rpl32-trnL*, *ndhF*, and *ycf1* as mutation hotspot regions that might be used as phylogenetic and species identification markers for angiosperms. Furthermore, these divergent regions are potentially useful genomic markers in phylogenetic studies of the subfamily Papaveroideae. Our results also support the previous report that the LSC region had more highly divergent sequences than the IR and SSC regions, suggesting that LSC regions are evolving rapidly [12].

Repeat sequences play important roles in genomic rearrangement and the provision of a stable chloroplast

genome structure [41, 42]. Because of the variable copy number and the variation in length, repeat sequences have attracted considerable attention for understanding phylogenetic relationships among species, biogeography, and population genetics. A total of 210 long repeats (30–190 bp) and 382 SSRs were identified from the eight Papaveroideae chloroplast genomes (Fig. 4). *D. leptopodum* contained the highest number (50) of long repeat sequences, and *Me. horridula* had the lowest number (16) among the compared species. The numbers of forward and palindromic sequence repeats detected in *D. leptopodum* were 46 and 4, respectively. Similarly, forward repeats were the most frequent repeat sequences observed among the other species. These repeats exhibit a pattern similar to the patterns reported previously, and the complex repeats are pivotal components in studying the evolutionary dynamics of the chloroplast genome.

Chloroplast genome SSRs (cpSSRs) have several essential characteristics, such as abundance, maternal inheritance, and haploid nature. Based on these features, chloroplast genome SSRs are mainly used in population genetic variation and gene flow analyses [43–45] and are considered valuable markers. The significance and applicability of cpSSR markers have been reported in various

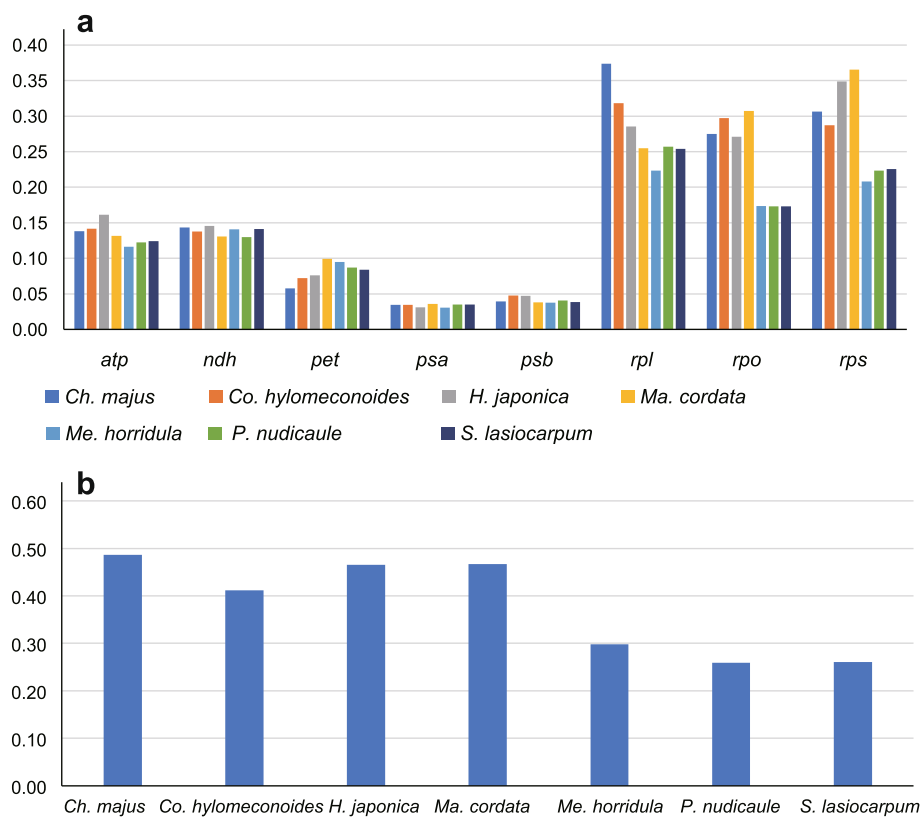


Fig. 7 The dN/dS values of protein-coding genes in the seven comparative combinations. *D. leptopodum* was used as the reference. a: Gene groups and b: the combination of all 79 protein-coding genes. dN nonsynonymous substitution; dS , synonymous substitution

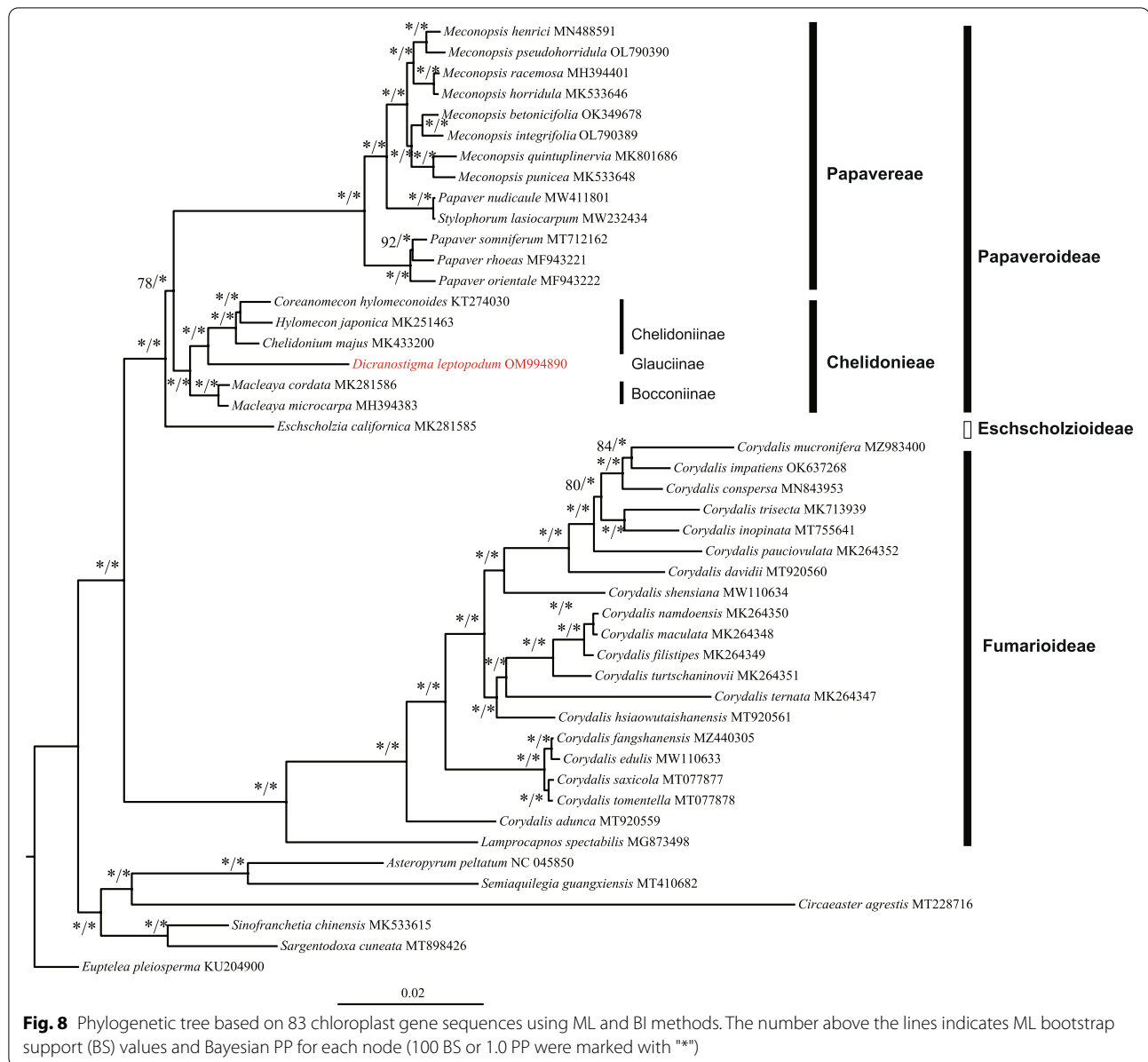
other Papaveroideae species, such as using cpSSR markers to assess the population genetics of opium poppy [46]. In this study, the numbers, types, and distribution of cpSSRs were analyzed in *D. leptopodum* and related chloroplast genomes. The highest number of cpSSRs was detected in *Ma. cordata* (69), while the fewest cpSSRs were observed in *P. nudicaule* (20). Consistent with previous results, we identified that mononucleotide-type SSRs were the most abundant in the chloroplast genome and were biased as A and T nucleotides, resulting in an AT-rich chloroplast genome.

Increasingly, studies have shown that chloroplast genome sequences are suitable for inferring phylogenetic relationships at different taxonomic classification levels [14, 47–50]. Based on whole chloroplast genome sequences, numerous phylogenetic problems at the deep node level have been solved, for example, identifying the earliest-diverged group of angiosperms [51–53] or the phylogenetic relationships among the five tribes of Oleaceae [13]. This approach allows a better understanding of the complex evolutionary links among angiosperms. Meanwhile, the chloroplast genome dataset may also resolve shallow phylogenetic problems. Previously,

Hoot et al. [54] determined the phylogenetic relationships of Papaveraceae based on the chloroplast regions *atpB*, *rbcL*, *matK*, and nuclear 26S ribosomal DNA. The results resolved the relationships of deep nodes (subfamily and tribe levels) in the family Papaveraceae but failed to resolve phylogenetic relationships at the species level using these four markers. In our study, *Dicranostigma* was sister to the tribe Chelidoniinae, with strong support obtained using the chloroplast genome (Fig. 8). Furthermore, our study revealed species clustering within *Meconopsis* and *Corydalis*, all with high bootstrap and posterior probability values.

Conclusions

In this study, we sequenced and assembled the complete chloroplast genome of *D. leptopodum* and compared it with the chloroplast genomes of related species from the subfamily Papaveroideae. We identified important genetic resources and evolutionary dynamics of the chloroplast genome for *D. leptopodum*, such as repetitive sequences, codon usage, SSRs, sequence divergence, IR contraction and expansion, molecular evolution analyses, and phylogenetic



inference. Comparative genomics indicates that the Papaveroideae chloroplast genomes are relatively conserved, with several regions of high sequence divergence identified as potential markers for phylogeny. Phylogenetic results resolved the phylogenetic position of *Dicranostigma*.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-09049-8>.

Additional file 1.

Additional file 2.

Additional file 3.

Acknowledgements

We thank Fanglong Mao and Jia Huang for their assistance with the plant materials and Jian He, Xingyong Cui, and Rudan Lyu for providing technical assistance. We also thank Chao Xu, Yanlei Liu, and Yanan Li for assisting with DNA extraction.

Authors' contributions

Lei Wang designed the experiments, analyzed the data, wrote the manuscript, and obtained financial support. Fuxing Li helped to collect samples and prepare figures. Ning Wang helped to identify specimens and revised the manuscript. Yongwei Gao helped to assemble the chloroplast genomes and prepare the figures. Kangjia Liu prepared samples for DNA extraction and conducted capillary electrophoresis. Gangmin Zhang directed the manuscript writing. Jiahui Sun provided financial support and revised the manuscript. All authors have reviewed and approved the manuscript.

Funding

This work was supported by the Science and Technology Research Project of Henan Province (No. 202102110232) and the CACMS Innovation Fund (No. CI2021A03909).

Availability of data and materials

The complete annotated sequence of *Dicranostigma leptopodum* is deposited in the NCBI database (<https://www.ncbi.nlm.nih.gov/>) (GenBank accession number: OM994890). The *D. leptopodum* material was obtained from Luoyang, Henan, China, and the specimens were subsequently deposited in the Herbarium of the College of Horticulture and Plant Protection, Henan University of Science and Technology (Voucher: WL1000). These specimens were identified by Dr. Ning Wang at Henan University of Science and Technology.

Declarations

Ethical approval and consent to participate

The collection of the sample analyzed in this study followed the Regulations on the Protection of Wild Plants of the People's Republic of China, the IUCN Policy Statement on Research Involving Species at Risk of Extinction, and the Convention on the Trade in Endangered Species of Wild Fauna and Flora.

Consent for publication

Not applicable.

Competing interests

The authors have no competing interests to declare.

Author details

¹College of Horticulture and Plant Protection, Henan University of Science and Technology, Luoyang 471023, Henan, China. ²Laboratory of Systematic Evolution and Biogeography of Woody Plants, School of Ecology and Nature Conservation, Beijing Forestry University, Beijing 100083, China. ³State Key Laboratory Breeding Base of Dao-di Herbs, National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China.

Received: 20 June 2022 Accepted: 25 November 2022

Published online: 02 December 2022

References

- Dang Y, Gong HF, Liu JX, Yu SJ. Alkaloid from *Dicranostigma leptopodum* (Maxim.) Fedde. *Chin Chem Lett*. 2009;20(10):1218–20.
- Sun R, Jiang H, Zhang W, Yang K, Wang C, Fan L, He Q, Feng J, Du S, Deng Z, et al. Cytotoxicity of Aporphine, Protoberberine, and Protopine Alkaloids from *Dicranostigma leptopodum* (Maxim.) Fedde. *Evid-Based Complementary Alter Med*. 2014;2014:580483.
- Dong HJ, Xiang CL. *Dicranostigma platycarpum*, a new synonym of *Dicranostigma erectum* (Papaveraceae). *Phytotaxa*. 2015;230(2):198–200.
- Wei XY, Lian FQ, Cai JH. Application and development of landscape ground covers in Jiangxi. *Acta Agric Univ Jiangxiensis*. 2002;05:680–3.
- Wang N, Chen H, Wang L. Physiological Acclimation of *Dicranostigma henanensis* to Soil Drought Stress and Rewatering. *Acta Soc Bot Pol*. 2021;90(907):1–11.
- Xu YX, Wang QH, Wang HB, Peng YK, Xue L. Research of dominant plants selection for treatment of heavy metal polluted soils surrounding mining areas. *Environ Protect Sci*. 2016;42(6):61–7.
- Yuan J, You FY, Hou CL, Ou HJ, Yin Y. Reconstruction of urban wilderness habitats based on vegetation rewilding: taking wildflower meadows as an example. *Landsc Archit Front*. 2021;9(1):26–39.
- Kadereit JW, Blattner FR, Jork KB, Schwarzbach A. The phylogeny of the Papaveraceae sensu lato: morphological, geographical and ecological implications. In: *Systematics and Evolution of the Ranunculiflorae: 1995/1995*; Vienna: Springer Vienna; 1995: 133–145.
- Hoot SB, Wefferling KM, Wulff JA. Phylogeny and character evolution of papaveraceae s. l. (Ranunculales). *Syst Botany*. 2015;40(2):474–88.
- Hoot SB, Kadereit JW, Blattner FR, Jork KB, Schwarzbach AE, Crane PR. Data Congruence and Phylogeny of the Papaveraceae s.l. Based on Four Data Sets: atpB and rbcL Sequences, trnK Restriction Sites, and Morphological Characters. *Syst Botany*. 1997;22(3):575–90.
- Dong W, Liu Y, Xu C, Gao Y, Yuan Q, Suo Z, Zhang Z, Sun J. Chloroplast phylogenomic insights into the evolution of *Distylium* (Hamamelidaceae). *BMC Genomics*. 2021;22(1):293.
- Dong W, Xu C, Liu Y, Shi J, Li W, Suo Z. Chloroplast phylogenomics and divergence times of *Lagerstroemia* (Lythraceae). *BMC Genom*. 2021;22:434.
- Dong W, Li E, Liu Y, Xu C, Wang Y, Liu K, Cui X, Sun J, Suo Z, Zhang Z, et al. Phylogenomic approaches untangle early divergences and complex diversifications of the olive plant family. *BMC Biol*. 2022;20(1):92.
- Li L, Hu Y, He M, Zhang B, Wu W, Cai P, Huo D, Hong Y. Comparative chloroplast genomes: insights into the evolution of the chloroplast genome of *Camellia sinensis* and the phylogeny of *Camellia*. *BMC Genomics*. 2021;22(1):138.
- Zhang J, Wang Y, Chen T, Chen Q, Wang L, Liu Z, Wang H, Xie R, He W, Li M, et al. Evolution of Rosaceae Plastomes Highlights Unique *Cerasus* Diversification and Independent Origins of Fruiting Cherry. *Front Plant Sc*. 2012;12(2562):736053.
- Chen Q, Hu H, Zhang D. DNA Barcoding and Phylogenomic Analysis of the Genus *Fritillaria* in China Based on Complete Chloroplast Genomes. *Front Plant Sc*. 2022;13:764255.
- Li J, Wang S, Jing Y, Wang L, Zhou S. A modified CTAB protocol for plant DNA extraction. *Chin Bull Bot*. 2013;48(1):72–8.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinform*. 2014;30(15):2114–20.
- Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. 2020;21(1):241.
- Qu XJ, Moore MJ, Li DZ, Yi TS. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods*. 2019;15(1):50.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28(12):1647–9.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt KD, Sayers EW. GenBank. *Nucleic Acids Res*. 2018;46(D1):D41–7.
- Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res*. 2019;47(W1):W59–64.
- Kumar S, Stecher G, Li M, Nknyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol*. 2018;35(6):1547–9.
- Amiryousefi A, Hyvönen J, Poczai P. IRScope: an online program to visualize the junction sites of chloroplast genomes. *Bioinform*. 2018;34(17):3030–1.
- Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res*. 2001;29(22):4633–42.
- Beier S, Thiel T, Munch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics*. 2017;33(16):2583–5.
- Xu B, Yang Z. PAMLX: a graphical user interface for PAML. *Mol Biol Evol*. 2013;30(12):2723–4.
- Dong W, Xu C, Cheng T, Zhou S. Complete chloroplast genome of *Sedum sarmentosum* and chloroplast genome evolution in Saxifragales. *PLoS ONE*. 2013;8(10):e77965.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80.
- Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*. 2019;35(21):4453–5.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61(3):539–42.
- Zhou J, Cui Y, Chen X, Li Y, Xu Z, Duan B, Li Y, Song J, Yao H. Complete chloroplast genomes of *Papaver Rhoeas* and *Papaver Orientale*: molecular structures, comparative analysis, and phylogenetic analysis. *Molecules*. 2018;23(2):437.
- Li X, Tan W, Sun J, Du J, Zheng C, Tian X, Zheng M, Xiang B, Wang Y. Comparison of Four Complete Chloroplast Genomes of Medicinal and

- Ornamental *Meconopsis* Species: Genome Organization and Species Discrimination. *Sci Rep.* 2019;9(1):10567.
35. Wicke S, Schneeweiss GM, Depamphilis CW, Muller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol.* 2011;76(3–5):273–97.
 36. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 2016;17(1):1–29.
 37. Xiong Q, Hu Y, Lv W, Wang Q, Liu G, Hu Z. Chloroplast genomes of five *Oedogonium* species: genome structure, phylogenetic analysis and adaptive evolution. *BMC Genom.* 2021;22(1):707.
 38. Wen F, Wu X, Li T, Jia M, Liu X, Liao L. The complete chloroplast genome of *Stauntonia chinensis* and compared analysis revealed adaptive evolution of subfamily Lardizabaloideae species in China. *BMC Genom.* 2021;22(1):161.
 39. Dong W, Liu J, Yu J, Wang L, Zhou S. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE.* 2012;7(4):e35071.
 40. Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, Cheng T, Guo J, Zhou S. ycf1, the most promising plastid DNA barcode of land plants. *Sci Rep.* 2015;5:8348.
 41. Wu S, Chen J, Li Y, Liu A, Li A, Yin M, Shrestha N, Liu J, Ren G. Extensive genomic rearrangements mediated by repetitive sequences in plastomes of *Medicago* and its relatives. *BMC Plant Biol.* 2021;21(1):421.
 42. Waminal NE, Pellerin RJ, Kang SH, Kim HH. Chromosomal mapping of tandem repeats revealed massive chromosomal rearrangements and insights into *Senna Tora* dysploidy. *Front Plant Sc.* 2021;12:629898.
 43. Xiao S, Xu P, Deng Y, Dai X, Zhao L, Heider B, Zhang A, Zhou Z, Cao Q. Comparative analysis of chloroplast genomes of cultivars and wild species of sweetpotato (*Ipomoea batatas* [L.] Lam). *BMC Genom.* 2021;22(1):262.
 44. Xue C, Geng FD, Li JJ, Zhang DQ, Gao F, Huang L, Zhang XH, Kang JQ, Zhang JQ, Ren Y. Divergence in the *Aquilegia ecalcarata* complex is correlated with geography and climate oscillations: Evidence from plastid genome data. *Mol Ecol.* 2021;30(22):5796–813.
 45. Ebert D, Peakall R. Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. *Mol Ecol Resour.* 2009;9(3):673–90.
 46. Zhang Y, Wang J, Yang L, Niu J, Huang R, Yuan F, Liang Q. Development of SSR and SNP markers for identifying opium poppy. *Int J Legal Med.* 2022;136:1261–71.
 47. Wikström N, Bremer B, Rydin C. Conflicting phylogenetic signals in genomic data of the coffee family (Rubiaceae). *J Syst Evol.* 2020;58(4):440–60.
 48. Thode VA, Lohmann LG, Sanmartín I. Evaluating character partitioning and molecular models in plastid phylogenomics at low taxonomic levels: A case study using *Amphilophium* (Bignoniaceae, Bignoniaceae). *J Syst Evol.* 2020;58(6):1071–89.
 49. Yao X, Song Y, Yang JB, Tan YH, Corlett RT. Phylogeny and biogeography of the hollies (*Ilex* L., Aquifoliaceae). *J Syst Evol.* 2020;59(1):73–82.
 50. Zhao F, Chen Y-P, Salmaki Y, Drew BT, Wilson TC, Scheen AC, Celep F, Bräuchler C, Bendiksby M, Wang Q, et al. An updated tribal classification of Lamiaceae based on plastome phylogenomics. *BMC Biol.* 2021;19(1):2.
 51. Goremykin VV, Nikiforova SV, Biggs PJ, Zhong B, Delange P, Martin W, Woetzel S, Atherton RA, McLenachan PA, Lockhart PJ. The evolutionary root of flowering plants. *Syst Biol.* 2013;62(1):50–61.
 52. Goremykin VV, Nikiforova SV, Cavalieri D, Pindo M, Lockhart P. The root of flowering plants and total evidence. *Syst Biol.* 2015;64(5):879–91.
 53. Xi Z, Liu L, Rest JS, Davis CC. Coalescent versus concatenation methods and the placement of *Amborella* as sister to water lilies. *Syst Biol.* 2014;63(6):919–32.
 54. Hoot SB, Wefferling KM, Wulff JA. Phylogeny and character evolution of papaveraceae s. l. (Ranunculales). *Syst Bot.* 2015;40(2):474–88.

Publisher's note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

