

Original Research

Sequence Analysis of the Plastomes of Two Tibetan Medicinal Plants of the Family *Papaveraceae*

Danyang Wu¹, Qian Qin¹, Xiaoyun Wang¹, Yiming Niu¹, Shouwen Zhang^{1,*}, Zejing Mu^{1,*}

¹The Research Center for Traditional Chinese Medicine Resources and Ethnic Minority Medicine, Jiangxi University of Chinese Medicine, 330004 Nanchang, Jiangxi, China

*Correspondence: wztsw@163.com (Shouwen Zhang); 20131060@jxutcm.edu.cn (Zejing Mu)

Academic Editor: Naeem Khan

Submitted: 27 June 2023 Revised: 12 August 2023 Accepted: 5 September 2023 Published: 27 December 2023

Abstract

Background: With the rapid development of next-generation sequencing technology, more plants plastomes have been sequenced, further advancing species identification and phylogenetic studies. However, there are a few studies on the genetic and phylogenetic analysis of the plastomes of *Dicranostigma lactuoides* Hook. f. et Thoms. and *Hypecoum leptocarpum* Hook. f. et Thoms. **Methods:** In this study, we sequenced and analyzed the plastomes of *Dicranostigma lactuoides* Hook. f. et Thoms. and *Hypecoum leptocarpum* Hook. f. et Thoms., and conducted a phylogenetic analysis using 13 related species. **Results:** The results showed that the plastomes of both *D. lactuoides* and *H. leptocarpum* had a typical tetrad structure, with sizes of 166,819 bp and 163,282 bp, respectively. We annotated 133 genes for *D. lactuoides* and 120 genes for *H. leptocarpum*. A total of 72 and 43 simple repetitive sequences were detected in *D. lactuoides* and *H. leptocarpum*, respectively. Codon preference analysis showed that the relative usage frequency of codons and the relative abundance of synonymous codons used were the same for both plastomes. Nucleotide polymorphism analysis identified seven variant loci with high nucleotide diversity (Pi) values, all located in the large single copy (LSC) region. Inverted repeat (IR) boundary analysis revealed differences in gene types and locations on both sides of the boundary, except for the small single copy/inverted repeat (SSC/IRA) boundary. The phylogenetic analysis showed the species clustered into two major groups, one with five genera (*Hypecoum*, *Corydalis*, *Papaver*, *Meconopsis*, and *Dicranostigma*) and the other with two genera (*Coreanomecon*; and *Hylomecon*). **Conclusions:** Comparative analysis of the plastome genomic characteristics and phylogeny of *D. lactuoides* and *H. leptocarpum* laid the foundation for identifying the above two species and the phylogenetic study and comprehensive exploitation of the *Papaveraceae*.

Keywords: *Dicranostigma lactuoides* Hook. f. et Thoms.; *Hypecoum leptocarpum* Hook. f. et Thoms.; *Papaveraceae*; plastomes; phylogeny

1. Introduction

There are about 30 genera and 700 species of the *Papaveraceae* Juss. family worldwide, among which 18 genera and 362 species are native to China, mainly southwest China, with about 9 genera and 158 species in the Qinghai-Tibet Plateau [1]. *Dicranostigma lactuoides* Hook. f. et Thoms. is a herbaceous plant of the *Dicranostigma* Hook. f. & Thomson genus in the *Papaveraceae* family, which is distributed in the central Himalayas in Tibet, western Sichuan, northwestern India, and Nepal [2]. It is cold and drought-tolerant species growing on rocky slopes above 3000 meters above sea level [3]. *D. lactuoides*, which is considered a rare Chinese herbal medicine, is recorded in the “Shaanxi Chinese Herbal Medicine” and “National Compilation of Chinese Herbal Medicine” [4]. Its alkaloids have beneficial biological activities, such as antibacterial, antitumor, and hepatoprotective activities, excellent healing properties on tuberculosis, gastric ulcers, and surgical wounds and the ability to boost immunity [5]. *Hypecoum leptocarpum* Hook. f. et Thoms. is an annual herb of the *Hypecoum* L. genus in the family *Papaveraceae* and is found in grasslands, valleys, riverbanks, gravels, and

other parts of the slopes of northeast China and Tibet [6]. *H. leptocarpum* has a long history of medicinal applications in China. According to ancient materia medica, the whole herb is used as medicine; *H. leptocarpum* is bitter and cold and is mainly used for treating cold, fever, and sore throat due to its heat clearing, detoxification, blood cooling, and pain-relieving effects [7]. *H. leptocarpum* also has strong vitality, branching, and other characteristics and is common around river banks and wetlands, where it exhibits an ecological value in preventing soil erosion [8]. In recent years, the research on Tibetan *D. lactuoides* and *H. leptocarpum* has mainly focused on their chemical composition and pharmacological effects. However, there are fewer reports on the genetic information of Tibetan *D. lactuoides* and *H. leptocarpum*. Currently, there are no reports on the complete plastome sequences of *D. lactuoides* and *H. leptocarpum*.

Chloroplasts are the main photosynthesis sites in higher plants, some algae, and a few protists, which are important for energy conversion in nature [9]. As semi-autonomous organelles with relatively independent genetic material, chloroplasts can synthesize some of their own pro-



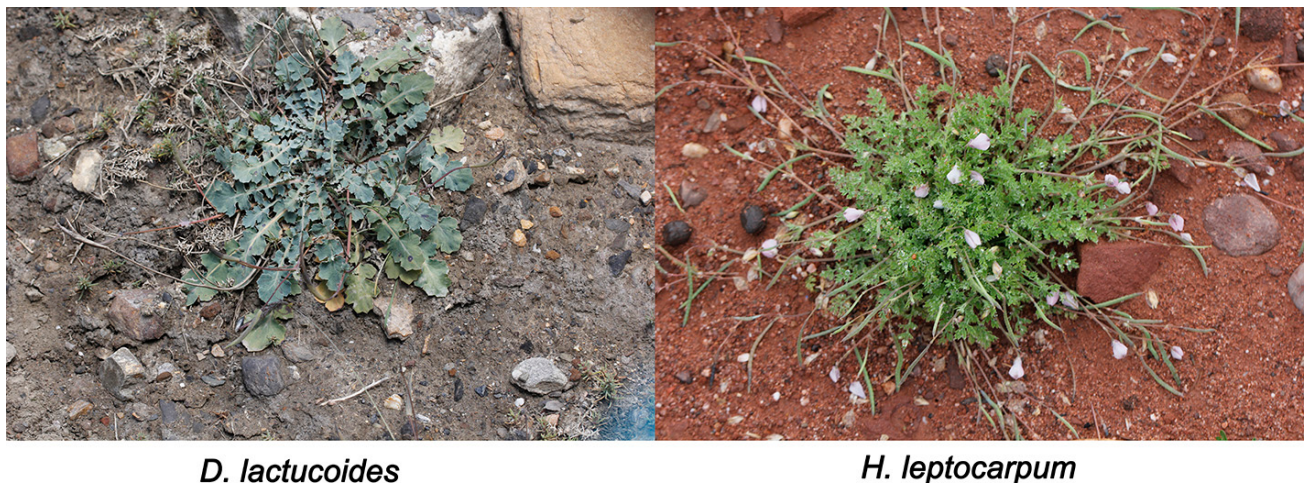


Fig. 1. Primitive plant pictures of *D. lactuoides* and *H. leptocarpum*.

Table 1. Table of collection information for 2 *Papaveraceae* species.

Species	Collection location	Longitude and Latitude	Altitude	Sample number	Accession number
<i>D. lactuoides</i>	Rikaze, Tibet Gyirong County Gompa Village	86°16'7.5"E 28°37'49.3"N	3722.3 m	H3050312	OR085636
<i>H. leptocarpum</i>	Lhasa, Tibet Daga Township, Qushui County	90°38'9.46"E 29°18'33.49"N	3596.0 m	H3050148	OR085637

teins [10]. The plastomes has a complete set of DNA sequences, typically 120–180 kb in size, containing 110–130 genes with a highly conserved structure and gene content [11]. The plastomes has a ring-like tetrameric structure that exists in multiple copies in the cell and mainly encodes photosynthetic light response-related proteins, which also regulate transcription and translation processes during gene expression [12]. Because of its high stability, small molecular weight, and easy sequencing, the plastomes has been widely used in plant classification, phylogenetic analysis, and evolutionary studies [13]. Therefore, sequencing and analyzing plastomes of the Tibetan *D. lactuoides* and *H. leptocarpum* are important for the molecular identification and population genetic diversity of the *Papaveraceae*.

In this study, the plastomes of two Tibetan poppy families were sequenced using next-generation sequencing (NGS) and were assembled and annotated. The obtained plastomes sequences of Tibetan *D. lactuoides* and *H. leptocarpum* were compared, and their sequence composition, sequence characteristics, and phylogenetic relationships were analyzed, aiming to provides a foundation for the rational classification and molecular identification of *Papaveraceae* species.

2. Material

2.1 Plant Materials

The plant materials of both *D. lactuoides* and *H. leptocarpum* were harvested from the Tibet Autonomous Region, China (Table 1) and dried in silica gel for analysis.

The leaves were identified by researcher Guoyue Zhong of Jiangxi University of Traditional Chinese Medicine (JJUCM) as *Papaveraceae* plants, and the voucher specimens were deposited in the herbarium of JJUCM (Fig. 1).

2.2 DNA Extraction and Sequencing

The genomic DNA was extracted from the leaves of *D. lactuoides* and *H. leptocarpum* using a plant genomic DNA kit (Tiangen Biochemical Technology (Beijing) Co., Ltd., Beijing, China). The quality of genomic DNA was detected by a spectrophotometer and 1% agarose gel electrophoresis. Samples with qualified DNA concentration greater than 20 ng·μL⁻¹ were lysed by an ultrasonic instrument, and subjected to fragment DNA purification and end repair. Thereafter, libraries with an insert size of 350 bp were prepared according to Illumina's standard procedure for genomic DNA library preparation, and then sequenced on the Illumina NovaSeq 6000 high-throughput sequencing platform (Illumina, San Diego, CA, USA). The sequencing was performed by Novogene Biotech Co., Ltd. (Nanjing, Jiangsu, China).

2.3 Plastomes Assembly and Annotation

The raw sequencing data were filtered using Trimmomatic software [14] (Aachen and Institute of Bio- and Geosciences: Plant Sciences, Forschungszentrum Jülich, Leo-Brandt-Straße, Jülich, Germany) to remove splice sequences and low-quality base sequences. The filtered plastomes were assembled from scratch using GetOrganelle

Table 2. Table of plastome information of 14 *Papaveraceae* species.

Species	GenBank accession number	Genome size/bp
<i>Corydalis saxicola</i> Bunting	MT077877.1	189,155
<i>Corydalis saxicola</i> Bunting	MT077879.1	189,029
<i>Corydalis tomentella</i> Franch.	MT093187.1	190,247
<i>Corydalis tomentella</i> Franch.	MT077878.1	190,198
<i>Corydalis conspersa</i> Mazim.	NC_047208.1	187,810
<i>Corydalis bungeana</i> Turcz.	MW538958.1	167,629
<i>Papaver somniferum</i> L.	NC_029434.1	152,931
<i>Papaver rhoeas</i> L.	NC_037831.1	152,905
<i>Papaver orientale</i> L.	NC_037832.1	152,799
<i>Meconopsis punicea</i> Maxim.	NC_050878.1	153,281
<i>Meconopsis henrici</i> Bur. et Franch.	NC_050877.1	153,788
<i>Coreanomecon hylomeconoides</i> Nakai	NC_031446.1	158,824
<i>Hylomecon japonica</i> (Thunb.) Prantl et Kundig	NC_045388.1	160,011
<i>Chelidonium majus</i> L.	NC_046829.1	159,734

1.7.5 software [15] (Chinese Academy of Sciences, Kunming, Yunnan, China) and were annotated using CP-GAVAS2 software [16] (Faculty of Technology, University of Bielefeld, Bielefeld, Germany) (<http://47.96.249.172:16019/analyzer/annotate/>). The plastome physical structure mapping was performed using OGDRAW software [17] (Max-Planck-Institut für Molekulare Pflanzenphysiologie, Am Mühlenberg, Germany) (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>).

2.4 Repeat Sequence Analysis

Simple sequence repeat (SSR) was screened using the online tool MISA v2.1 (Leibniz Institute of Plant Genetics and Crop Research, Gatersleben, Germany) (<https://webb.last.ipk-gatersleben.de/misa/>) software with parameters set to 10 for single nucleotide SSR repeats, 5 for dinucleotide SSR repeats, 4 for single and trinucleotide SSR repeats, and 3 for tetranucleotide and pentanucleotide SSR repeats. The REPuter software [18] (Faculty of Technology, Bielefeld, Germany) (<http://bibiserv.techfak.uni-bielefeld.de/reputer/>) was used to annotate the plastome repeats, including forward, reverse, complementary, and palindromic repeats.

2.5 Analysis of Codon Preference and Selection Pressure

CodonW 1.4.2 software [19] (Paul Sharp lab, Dept of Genetics, University of Nottingham, Nottingham, UK) was used to calculate the codon preference (Relative synonymous codon usage, RSCU), and PAML v4.973 software [20] (Department of Biology, Galton Laboratory, London, UK) was used to calculate the ratio between non-synonymous mutation frequency (Ka) and synonymous mutation frequency (Ks) for each gene.

2.6 Nucleotide Diversity Analysis

The MAFFT v7 software [21] (Bioinformatics Center, Institute for Chemical Research, Kyoto University Uji, Ky-

oto 611-0011, Japan) was used to compare the plastome sequences of *D. lactuoides* and *H. leptocarpum* (with default parameters), while DnaSP v5.1 software [22] (Departament de Genètica, Facultat de Biologia and Institut de Recerca de la Biodiversitat, Universitat de Barcelona, Barcelona, Spain) was used to assess the nucleotide diversity (Pi) of the plastomes with a step size of 200 bp and a window length of 600 bp.

2.7 Inverted Repeat (IR) Boundary Analysis

IR boundary analysis was conducted using the online tool IRSCOPE (<https://irscope.shinyapps.io/irapp/>) [23] to compare the gene types and the location in the border area. The contraction and expansion of the IR regions in the plastomes of the 2 species were also analyzed.

2.8 Covariance Analysis

Covariance analysis of the whole genome sequences of the *D. lactuoides* and *H. leptocarpum* plastomes was performed using the online mVISTA software (<http://genome.lbl.gov/vista/mvista/submit.shtml>) [24] to determine whether there were changes in the order of gene arrangement (Shuffle-Lagan was used for the alignment).

2.9 Phylogenetic Relationship Analysis

Phylogenetic analysis was performed using 14 spliced plastome sequences of the family *Papaveraceae* downloaded from the National Center for Biotechnology Information (NCBI) database (Table 2) and 2 newly sequenced plastomes of the *Papaveraceae* species in Tibet. *Chelidonium majus* L. was used as an outgroup. The phylogenetic tree was constructed based on the maximum likelihood (ML) method using PhyloSuite v1.2.2 [25] (Key Laboratory of Aquaculture Disease Control, Chinese Academy of Sciences, Wuhan, China) with bootstrap values of 10,000 replicates.

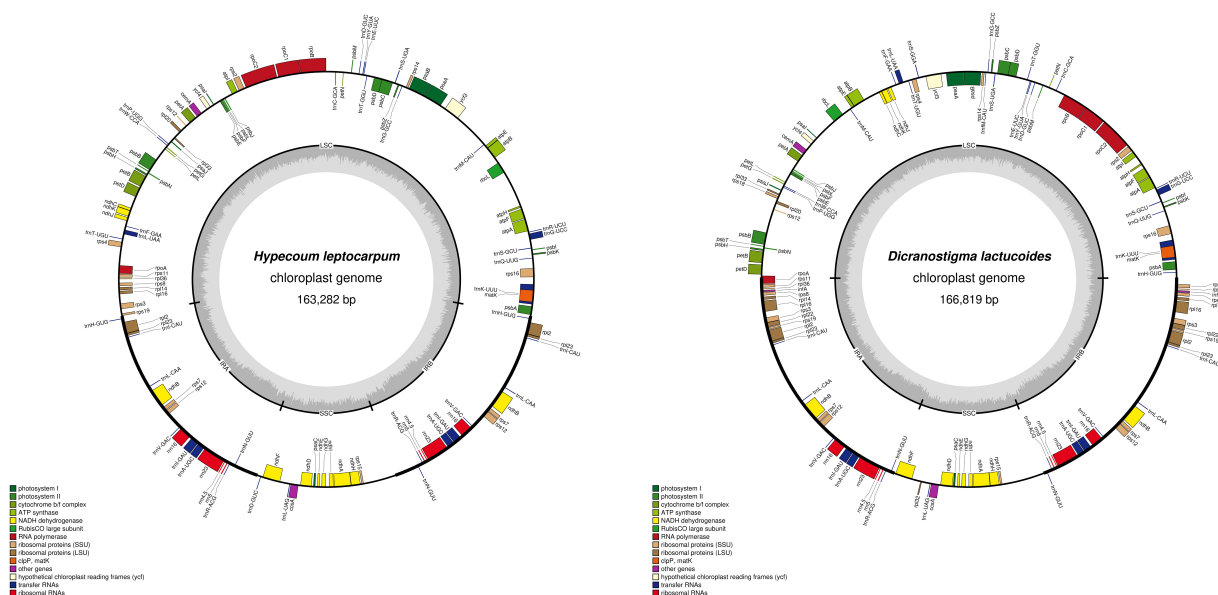


Fig. 2. Plastome structure of *D. lactuoides* and *H. leptocarpum*. Note: The thick black line indicates that the two reverse repeat regions IRa and IRb will be the large single copy region (LSC), the small single copy region (SSC) separated; dark part of inner circle indicates GC content.

Table 3. Plastome genomic characteristics of *D. lactuoides* and *H. leptocarpum*.

Species	Size	Length (LSC)	Length (SSC)	Length (IR)	GC%	PCGs	tRNAs	rRNAs	Genes
<i>D. lactuoides</i>	166,819	82,927	19,538	32,177	39.23	89	36	8	133
<i>H. leptocarpum</i>	163,282	90,852	17,600	27,415	38.57	76	36	8	120

3. Results and Analysis

3.1 Basic Characteristics of the Plastome Structure of *D. lactuoides* and *H. leptocarpum*

The full-length plastomes of *D. lactuoides* and *H. leptocarpum* are 166,819 and 163,282 bp, respectively, and exhibit a typical tetrad structure (Fig. 2, Table 3) with the GC contents of 39.23% and 38.57%, respectively. The large-single copy (LSC) regions of the plastomes of *D. lactuoides* and *H. leptocarpum* are 82,927 bp and 90,852 bp, while their small-single copy (SSC) regions are 19,538 bp and 17,600 bp, respectively. The LSC and SSC regions of *D. lactuoides* and *H. leptocarpum* are separated by two reversed IR regions of 32,177 bp and 27,415 bp, respectively.

For *D. lactuoides*, we annotated 133 genes (Table 4), containing 89 protein-coding genes (PCGs), 36 tRNA, and 8 rRNA genes. Among the 89 PCGs, 14 (*ndhB*, *rpl2*, *rpl14*, *rpl16*, *rpl22*, *rpl23*, *rpl36*, *rps3*, *rps7*, *rps8*, *rps11*, *rps12*, *rps19*, and *infA*) were double-copy genes. Similarly, seven tRNA genes (*trnA-UGC*, *trnI-CAU*, *trnA-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, and *trnV-GAC*) and four rRNA genes (*rrn4.5*, *rrn5*, *rrn6*, and *rrn23*) had double copies. For *H. leptocarpum*, we annotated 120 genes containing 76 PCGs, 36 tRNA, and 8 rRNA genes. Five PCGs (*ndhB*, *rpl2*, *rpl23*, *rps7*, and *rps12*), nine tRNA genes (*trnA-UGC*, *trnD-GUC*, *trnH-CUG*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, and *trnV-GAC*), and four

rRNA genes (*rrn4.5*, *rrn5*, *rrn6*, and *rrn23*) had double copies. *D. lactuoides* and *H. leptocarpum* had 14 identical genes containing one intron and 2 identical genes containing two introns. However, *D. lactuoides* had 13 more genes than *H. leptocarpum*, with *rpl16* containing one intron. *rpl22*, *rpl32*, *rpl18*, *infA*, *trnS-GGA*, and *trnM-CAU* were unique genes.

3.2 Repeat Sequence Analysis

We identified 72 simple repeat sequences (SSRs) from *D. lactuoides* (Fig. 3), of which 43 were single nucleotide, 14 were dinucleotide, 5 were trinucleotide, 7 were tetranucleotide, and 3 were hexanucleotide sequences. The SSR sequences were mostly distributed in the LSC region (49), followed by the SSC region (11), and the IRa and IRb regions which both had 6. *H. leptocarpum* had 43 SSRs, which contained 16 single nucleotides, 12 dinucleotide, 6 trinucleotide, and 9 tetranucleotide sequences. The distribution pattern was similar to that of *D. lactuoides*, with 24 SSRs located in the LSC, 7 in the SSC, 12 in the IRa and IRb regions (six each). Among the single nucleotide sequences, the A/T repeats were predominant. Thus, SSRs were widely distributed in the plastomes and are important for molecular marker development, which could aid hybrid breeding and variety identification in *Papaveraceae* plants [26].

Table 4. Categorization of annotated genes in the plastomes of *D. lactuoides* and *H. leptocarpum*.

Gene function classification	Gene type	Gene name
Photosynthetic	Optical System I	<i>psaA, psaB, psaC, psaI, psaJ</i>
	Optical Systems II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	NADH-dehydrogenase subunit	<i>ndhA*</i> , <i>ndhB*(×2)</i> , <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>
	Cytochrome b/f complex	<i>petA</i> , <i>petB*</i> , <i>petD*</i> , <i>petG</i> , <i>petL</i> , <i>petN</i>
	ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF*</i> , <i>atpH</i> , <i>atpI</i>
	Ribulose diphosphate carboxylase large subunit	<i>rbcL</i>
	Ribosomal large subunit (LSU)	<i>rpl14(×2)</i> , <i>rpl16*(×2)</i> , <i>rpl2*(×2)</i> , <i>rpl20</i> , <i>rpl22(×2)[#]</i> , <i>rpl23(×2)</i> , <i>rpl32[#]</i> , <i>rpl33</i> , <i>rpl36(×2)</i>
	Small subunit of ribosome (SSU)	<i>rps11(×2)</i> , <i>rps12***(×2)</i> , <i>rps14</i> , <i>rps15</i> , <i>rps16*</i> , <i>rps18[#]</i> , <i>rps19(×2)</i> , <i>rps2</i> , <i>rps3(×2)</i> , <i>rps4</i> , <i>rps7(×2)</i> , <i>rps8(×2)</i>
	RNA polymerase subunits	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1*</i> , <i>rpoC2</i>
	Ribosomal RNA genes	<i>rrn16(×2)</i> , <i>rrn23(×2)</i> , <i>rrn4.5(×2)</i> , <i>rrn5(×2)</i>
Self-replication	Transfer of RNA genes	<i>trnA-UGC*(×2)</i> , <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC</i> , <i>trnF-GAA</i> , <i>trnG-GCC</i> , <i>trnG-UCC*</i> , <i>trnH-GUG</i> , <i>trnI-CAU(×2)</i> , <i>trnI-GAU*(×2)</i> , <i>trnK-UUU*</i> , <i>trnL-CAA(×2)</i> , <i>trnL-UAA*</i> , <i>trnL-UAG</i> , <i>trnM-CAU</i> , <i>trnN-GUUU(×2)</i> , <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnR-ACG(×2)</i> , <i>trnR-UCU</i> , <i>trnS-GCU</i> , <i>trnS-GGA[#]</i> , <i>trnS-UGA</i> , <i>trnT-GGU</i> , <i>trnT-UGU</i> , <i>trnV-GAC(×2)</i> , <i>trnW-CCA</i> , <i>trnY-GUA</i> , <i>trnY-M-CAU[#]</i>
	Transcription initiation factor	<i>infA(×2)[#]</i>
	Mature enzymes	<i>matK</i>
Other genes	Periplasmic proteins	<i>cemA</i>
	Cytochrome C synthesis genes	<i>ccsA</i>
	Hypothetical conserved chloroplast open reading frame	<i>ycf3**</i> , <i>ycf4</i>

Note: The parentheses stand for multi-copy genes; * and ** refer to genes containing 1 intron and 2 introns, respectively; # represents the specific gene of *D. lactuoides*.

The tandem repeat sequence analysis showed 258 and 91 long repeats in *D. lactuoides* and *H. leptocarpum*, respectively, including forward, palindromic, reverse, and complementary sequences (Fig. 4). These long repetitive sequences can be used to study the evolutionary history of species [27].

3.3 Preference Analysis of Codons

The RSCU value is the ratio of the actual frequency of codon usage to the desired frequency of use and is often used to measure the codon use preference. The RSCU value >1 indicates a high codon usage frequency [28]. *D. lactuoides* and *H. leptocarpum* encoded 65 codons, and both had 32 high-frequency codons with RSCU values >1. The comparative analysis of the RSCU values showed that the highest RSCU value was 1.9962 and the lowest was 0.0038, and there was an insignificant difference in RSCU

values between the two species. The most frequently used codon was TTA encoding leucine (Leu) (Fig. 5), and the least used codon was CGC encoding arginine (Arg). All amino acids were encoded by multiple codons except tryptophan (Trp). Arg, Leu, and serine (Ser) have six synonymous codons, while alanine (Ala), glycine (Gly), proline (Pro), threonine (Thr), and valine (Val) have four synonymous codons. Isoleucine (Ile) has three synonymous codons, and the rest of the amino acids have two synonymous codons. The codons usually differ only at the third base.

3.4 Ka/Ks Analysis

Ka/Ks indicates the ratio between heterosense substitution (Ka) and synonymous substitution (Ks) and is often used to determine whether the coding gene is subject to selective pressure during evolution; thus, it is important to de-

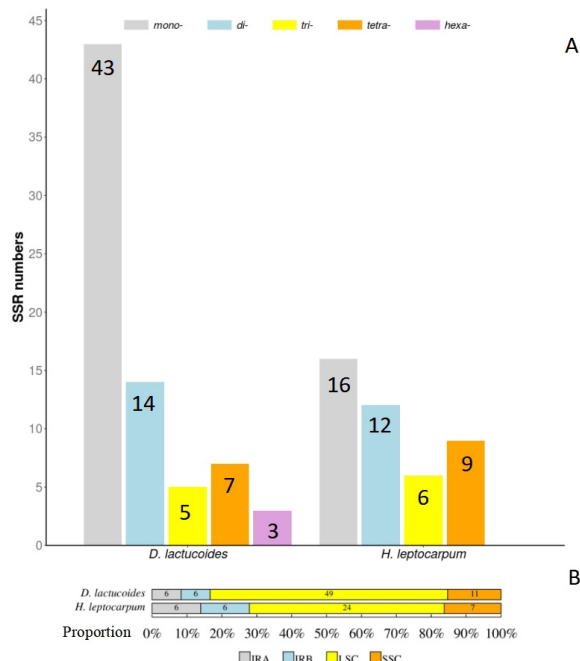


Fig. 3. Type and number of SSRs in the plastomes of *D. lactuoides* and *H. leptocarpum* (A); Number of SSRs in the LSC, SSC and IR regions (B). SSRs, simple repeat sequences.

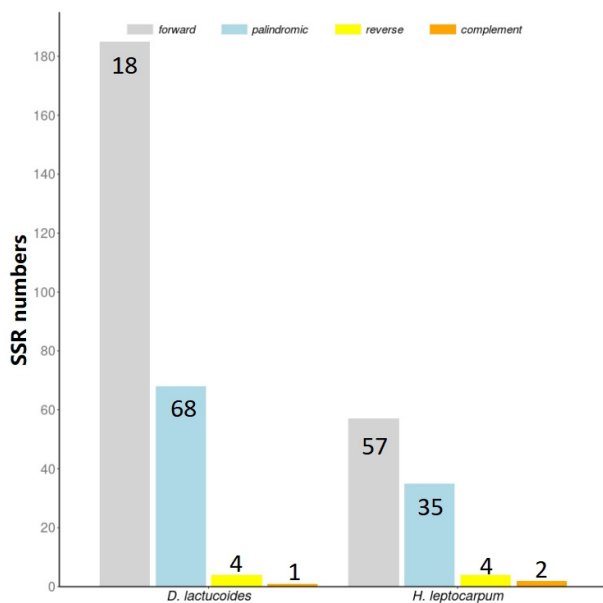


Fig. 4. Analysis of long repeat sequences of plastomes of *D. lactuoides* and *H. leptocarpum*.

termine the Ka/Ks ratio during evolutionary analysis [29]. If Ka/Ks is >1 , the coding gene is considered to be under positive selective pressure, and if Ka/Ks = 1, the gene is considered to be under neutral selective pressure. However, when Ka/Ks is <1 , the gene is considered to be under purifying selection pressure. In this study, two genes (*psJ* and *rsp7*) had Ka/Ks ratios >1 (Fig. 6), indicating a positive selection effect and that the genes would be retained dur-

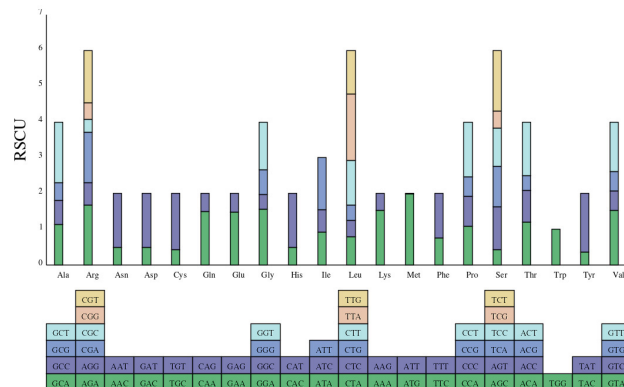


Fig. 5. Relative use abundance of synonymous codons in chloroplast genomes of *D. lactuoides* and *H. leptocarpum*.

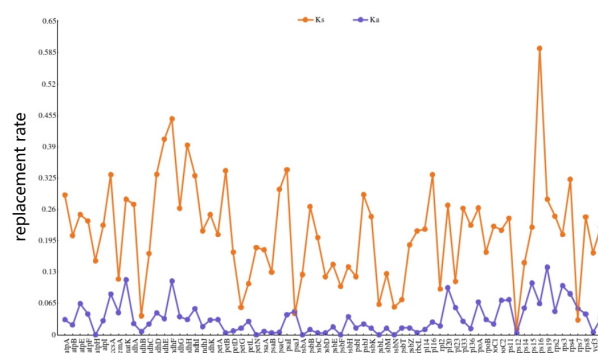


Fig. 6. Gene selection pressure analysis of plastomes of *D. lactuoides* and *H. leptocarpum*.

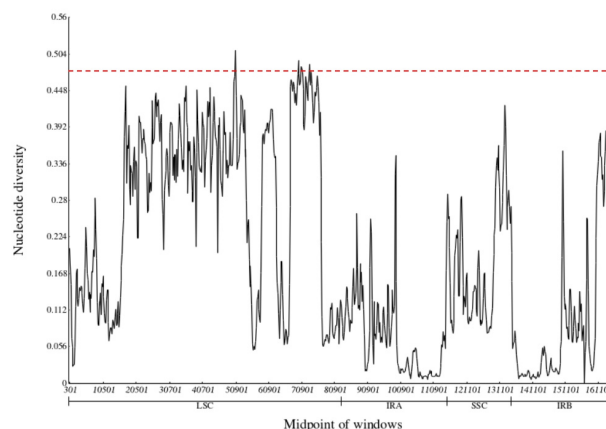
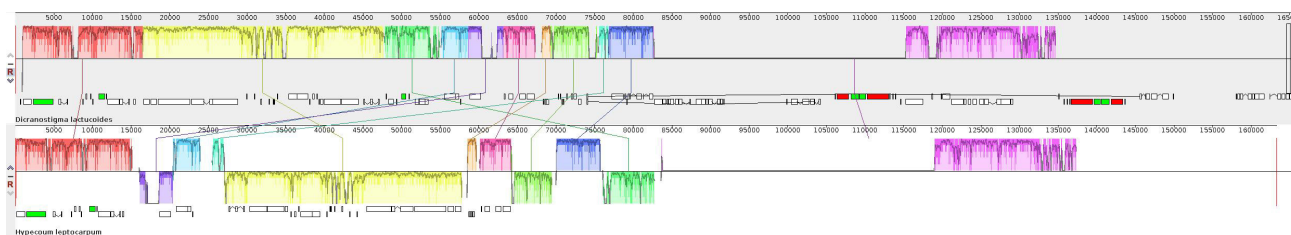
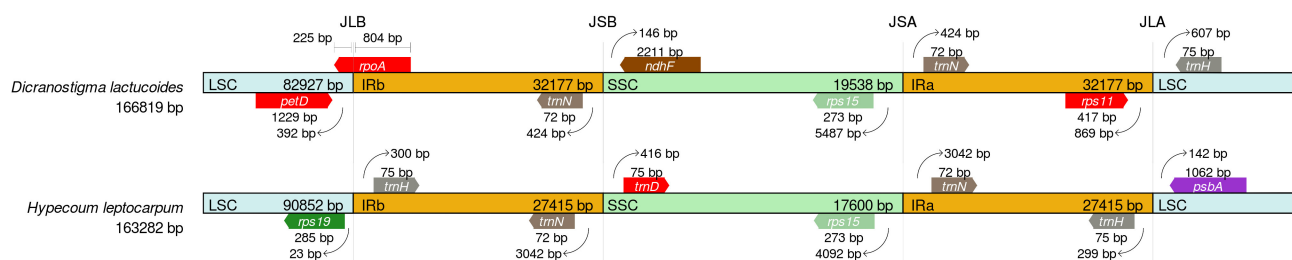


Fig. 7. Nucleotide diversity analysis of plastomes of *D. lactuoides* and *H. leptocarpum*.

ing evolution. The remaining genes had Ka/Ks ratios of 0 to 0.51, indicating a purifying selection effect and that the mutant genes would be eliminated during evolution. Six genes (*atpH*, *petN*, *psbA*, *psbF*, *psbL*, and *psbN*) had Ka/Ks ratios of 0 and were all photosynthesis-related genes with a synonymous mutation frequency of 0.



3.5 Nucleotide Diversity Analysis

A total of 828 polymorphic loci were detected in the plastomes of *D. lactuoides* and *H. leptocarpum* (Fig. 7), with nucleotide diversity (Pi) values ranging from 0 to 0.5096 with a mean value of 0.1947. The IR region was less variable and more conservative than the LSC and SSC regions in terms of sequence variation. In addition, there were seven polymorphic loci with Pi values >0.476, located between 40,701 and 80,901 bp, all distributed within the LSC region. These regions with very high variability provide a basis for evolution which could aid species identification and kinship analysis of the *Papaveraceae*.

3.6 IR Boundary Analysis

The JBL boundary (the boundary between the LCS and IRb regions) of *D. lactuoides* is located within the *rpoA* gene (Fig. 8), while that of *H. leptocarpum* is located between the *rsp19* and *trnH* genes. Moreover, the JSB boundary (the boundary between the IRb and SSC regions) of *D. lactuoides* is flanked by the *trnN* and *ndhF* genes, while that of *H. leptocarpum* is flanked by the *trnN* and *trnD* genes. Overall, the JSA boundary had the least difference variance, while the JLB boundary had the most variance, which may be caused by the expansion or contraction occurring in these regions.

3.7 Covariance Analysis

Covariance reveals interlocking homologous genes in the genomes of different species and their order of arrangement in the genetic map. The genetic distance between two species is positively correlated with the degree of covariance [30]. The genomes of different species may have some

regions of covariance, which may be interrupted by rearrangements and inversions during evolution, and such phenomena can be detected visually through covariance analysis. As shown in Fig. 9, rearrangements and inversions occurred roughly between 15,000 and 85,000 bp region. The covariance is good at both ends, indicating that the gene homology is higher in this region.

3.8 Phylogenetic Analysis

In this study, the plastomes of *D. lactuoides* and *H. leptocarpum* (obtained by sequencing) and 12 other published species of *Papaveraceae* were used to construct a maximum likelihood (ML) phylogenetic tree (Fig. 10), with the plastomes of *C. majus* as an outgroup. The results showed that the self-extension values of all branch nodes were greater than 70%, indicating the reliability of the constructed developmental tree. All species originated from seven genera, and species of the same genus clustered together. The *Hylomecon japonica* and the genus *Coreanomecon hylomeconoides* clustered as the first major branch, located at the base of the phylogenetic tree. The second major branch included four species of the genus *Corydalis* (*C. saxicola*, *C. tomentella*, *C. conspersa*, and *C. bungeana*), one species of *Hypecoum* (*H. leptocarpum*), three species of *Papaver*, (*P. somniferum*, *P. rhoeas*, and *P. orientale*), two species of *Meconopsis* (*M. punicea* and *M. henrici*), and one species of *Dicranostigma* (*D. lactuoides*). In the second major branch, the genus *Dicranostigma* clustered separately as a single group, and the remaining four genera clustered into a large group, within which *Meconopsis* and *Papaver* diverged into a separate group from that of *Hypecoum* and *Corydalis*. Thus, the genus

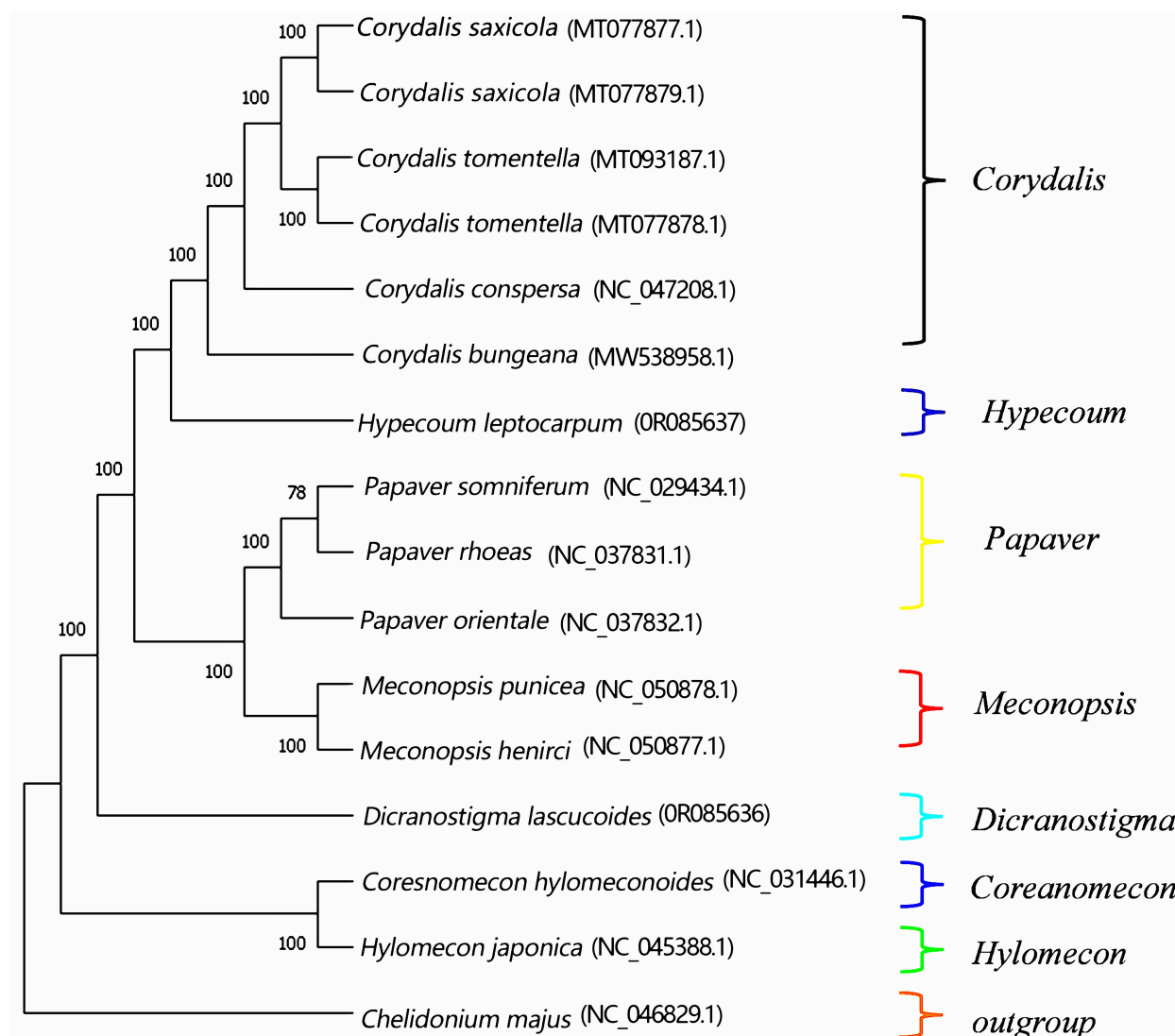


Fig. 10. Phylogenetic tree of plastomes of 14 species in *Papaveraceae*. Note: The parameter is, model = GTR+G, method = maximum likelihood, bootstrap = 10,000.

Meconopsis is closely related to *Papaver*, while *Hypecoum* is closely related to *Corydalis*. The results of this study are consistent with the taxonomic results reported by Ma Wenming [31].

4. Discussion

This study sequenced, assembled, and annotated the plastomes of the Tibetan *D. lactuoides* and *H. leptocarpum* for the first time. The results showed that the structure (tetrad structure), length (166,819 bp for *D. lactuoides* and 163,282 bp for *H. leptocarpum*), and GC content (39.23% for *D. lactuoides* and 38.57%, *H. leptocarpum*) of these two species were similar to those of other reported *Papaveraceae* species [32,33]. The annotation of *D. lactuoides* and *H. leptocarpum* plastomes yielded 133 and 120 functional genes, respectively. The comparison revealed that *H. leptocarpum* was missing two *rpl22* and *infA* genes, one *rpl14*, *rpl16*, *rpl32*, *rpl36*, *rps3*, *rps8*, *rps11*, *rps18*,

rps19, *trnS-GGA*, and *trnM-CAU* genes, and one *trnD-GUC* and *trnH-GUG* genes. Loss of the *infA* gene, encoding a translation initiation factor and regulates protein translation, occurred in *H. leptocarpum*. A high loss frequency of the *infA* gene was also reported in several angiosperms [34]. Since the morphological characteristics of *D. lactuoides* and *H. leptocarpum* are relatively similar, the existing methods could not distinguish them; thus, the variation analysis of the plastomes and gene loss may supplement their identification methods.

Simple repeat sequences (SSRs) are widely present in plant plastomes but vary in number and fragment length in different species; thus, they can be used to distinguish different species due to their high specificity [35]. Long repetitive sequences can influence gene expression and play an important role in biological evolution [36]. We identified 72 and 43 SSR loci in the genomes of *D. lactuoides* and *H. leptocarpum*, respectively, including single, dinu-

cleotide, trinucleotide, tetranucleotide, and hexanucleotide sequences. In addition, 258 and 98 long repeat sequences were detected in *D. lactuoides* and *H. leptocarpum*, respectively, including forward, palindrome, reverse, and complementary sequences. These SSR loci and long repeat sequences provide candidate molecular markers for species identification in the family *Papaveraceae*.

Codon preference helps to study the evolution and environmental adaptations of different species [37]. In this study, we found that the preference for codons ending in A/U and the weak codon numbering in the plastomes of chicory leaf vulture flower and fine-fruited hornbeam fennel were consistent with most of the species previously reported [38,39]. Natural selection pressures cause differences in base composition at the three positions of synonymous codons in the protein-coding sequences of organisms, and changes in the third base of the codon usually do not alter the coding amino acids [40]. The results of the present study indicate that the codon preferences of the plastomes of *D. lactuoides* and *H. leptocarpum* were mainly affected by selection, similar to previous studies on related plants such as *Rabdosia rubescens* [41]. In this study, seven variant loci with high nucleotide polymorphism ($P_i > 0.476$) were identified via nucleotide diversity analysis. These high mutation regions can provide a basis for DNA barcode screening and a marker for phylogenetic analysis for *D. lactuoides* and *H. leptocarpum* identification.

The contraction and expansion of IR regions during biological evolution can lead to variations in plastomes, and analyzing the IR region boundaries of the plastomes can reveal the direction of biological evolution [42]. Previous studies found that the IR/SSC boundary is relatively conserved in early-differentiated dicotyledons, whereas the frequency of the IR/LSC boundary variation is higher [43], consistent with the present study. In this study, we found that the genes on both sides of the SSC/IRa boundaries were identical in *D. lactuoides* and *H. leptocarpum*, while the boundary genes of LSC/IRb, SSC/IRb, and LSC/IRa differed significantly. These differences could provide new insights for identifying *Papaveraceae* species.

In this study, we constructed an ML phylogenetic tree using the plastomes of *D. lactuoides*, *H. leptocarpum*, and 13 other reported relatives, using *C. majus* as the outgroup. The tree had two main branches, one with five genera, including *Hypecoum*, *Corydalis*, *Papaver*, *Meconopsis*, and *Dicranostigma*, and the other containing two genera, *Coreanomecon* and *Hylomecon*, with 100% support values. Genera *Corydalis* and *Hypecoum* are more closely related, and the genera *Papaver* and *Meconopsis* are sister branches of the same clade.

In conclusion, the phylogenetic relationships among *Papaveraceae* genera obtained in this study using the plastomes of 16 poppy families are relatively accurate and reliable. However, additional plastomes of *Papaveraceae* plants need to be sequenced and annotated, and further phy-

logenetic studies of *Papaveraceae* plants should incorporate various data types, such as plant morphological characteristics and nuclear genomic and transcriptomic data.

5. Conclusions

The plastome genomic data obtained in this study can provide a basis for genetic diversity studies of *D. lactuoides* and *H. leptocarpum* and molecular identification of the *Papaveraceae* genus.

Availability of Data and Materials

The datasets used and/or analysed during the current study are available from the first author upon reasonable request. Raw data for the plastomes sequencing can be accessed by logging on to the website <https://www.ncbi.nlm.nih.gov/>.

Author Contributions

DW: Conceptualization, Methodology, Software, Investigation, Formal Analysis, Writing Original Draft; QQ: Data Curation, Writing Original Draft; XW: Designed the research study and provided help and advice on paper writing; YN: Visualization, Investigation; SZ and ZM: Conceptualization, Funding Acquisition, Resources, Supervision, Writing Review & Editing. All authors read and approved the final manuscript. All authors contributed to editorial changes in the manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

The plant materials of 2 species of the *Papaveraceae* in this study were collected from the Tibet Autonomous Region of China (Tibet) and identified by Professor Guoyue Zhong of Jiangxi University of Chinese Medicine (JUTCM). The certificate specimens were all stored in the herbarium of JUTCM.

Acknowledgment

Not applicable.

Funding

This research was funded by the Jiangxi University of Chinese Medicine Science and Technology Innovation Team Development Program (GrantCXTD22002).

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] Du X, Cao L, Mu Z, Gao Y, Zhong G. Basal origin and standardization of medicinal herbs of *Papaveraceae* for Tibetan medicine. *China Modern Traditional Chinese Medicine*. 2022; 24: 1187–1195. (In Chinese)

- [2] Editorial Committee of the Flora of China. Chinese Academy of Sciences. Flora of China. Science Press: Beijing. 1993.
- [3] Zhao Q. Resources and utilization of the genus Baldrops in China. Journal of Chinese Veterinary Medicine. 2019; 38: 28–31. (In Chinese)
- [4] National Compilation of Chinese Herbal Medicines Writing Group. National compilation of Chinese herbal medicines (pp. 266). Previous book. People's Health Publishing House: Beijing. 1975.
- [5] Long FL, Yue ZG, Yu D, Liu H, Hu PH, Jin M. Progress in the study of chemical composition, pharmacological effects and toxicity of Baldrops. Hebei Medicine. 2017; 39: 3492–3495. (In Chinese)
- [6] Ao L, Li J, Xu N, Lu B, Chen X. Progress of chemical composition and pharmacological research of the Mongolian medicinal herb Angular fennel. Journal of Inner Mongolia University for Nationalities (Natural Science Edition). 2019; 34: 197–200. (In Chinese)
- [7] Zhang R, Cha S, Yin X, Bai R, Li Y, Tu P, *et al.* Progress in the study of chemical composition and pharmacological activity of the Tibetan medicine horny fennel and its congeners. Chinese Herbal Medicine. 2016; 47: 1217–1224. (In Chinese)
- [8] Liu Q. Wildlife atlas of Arjinshan National Nature Reserve (pp. 103). China Environment Publishing Group: Beijing. 2018.
- [9] Daniell H, Jin S, Zhu XG, Gitzendanner MA, Soltis DE, Soltis PS. Green giant-a tiny chloroplast genome with mighty power to produce high-value proteins: history and phylogeny. Plant Biotechnology Journal. 2021; 19: 430–447.
- [10] Turdi R, Mu L, Tian X. Characteristics of the chloroplast genome of *Isopyrum anemonoides*. Sheng Wu Gong Cheng Xue Bao. 2022; 38: 2999–3013.
- [11] Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biology. 2016; 17: 134.
- [12] Yurina NP, Sharapova LS, Odintsova MS. Structure of Plastid Genomes of Photosynthetic Eukaryotes. Biochemistry. Biokhimiia. 2017; 82: 678–691.
- [13] Niu Y, Gao C, Liu J. Comparative analysis of the complete plastid genomes of *Mangifera* species and gene transfer between plastid and mitochondrial genomes. PeerJ. 2021; 9: e10774.
- [14] Govender KN, Eyre DW. Benchmarking taxonomic classifiers with Illumina and Nanopore sequence data for clinical metagenomic diagnostic applications. Microbial Genomics. 2022; 8: mgen000886.
- [15] Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, *et al.* GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biology. 2020; 21: 241.
- [16] Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, *et al.* CP-GAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Research. 2019; 47: W65–W73.
- [17] Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Research. 2019; 47: W59–W64.
- [18] 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 Research. 2001; 29: 4633–4642.
- [19] Peden JF. Analysis of codon usage [PhD thesis]. University of Nottingham: UK. 1999.
- [20] Shekhovtsov SV, Vasiliev GV, Latif R, Poluboyarova TV, Peltek SE, Rapoport IB. The mitochondrial genome of *Dendrobaena tellermanica* Perel, 1966 (Annelida: Lumbricidae) and its phylogenetic position. Vavilovskii Zhurnal Genetiki i Selektcii. 2023; 27: 146–152.
- [21] Katoh K, Kuma KI, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Research. 2005; 33: 511–518.
- [22] Volpe M, Das J. methylR: a graphical interface for comprehensive DNA methylation array data analysis. Bioinformatics (Oxford, England). 2023; 39: btad184.
- [23] Amiryousefi A, Hyvönen J, Pocai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics (Oxford, England). 2018; 34: 3030–3031.
- [24] Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, *et al.* VISTA: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics (Oxford, England). 2000; 16: 1046–1047.
- [25] Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, *et al.* PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Molecular Ecology Resources. 2020; 20: 348–355.
- [26] Zhang D-H, Gong L-Z, Zhuang J-X, Zou X, Wang XB, Zhan R, *et al.* Genomic SSR molecular marker development and germplasm genetic diversity in patchouli. Genomics and Applied Biology. 2023; 1–19. (In Chinese)
- [27] Zhang R, Zhang TL, Fan TT, Zhu B, Zhang LP, Xu LY, *et al.* Evolutionary relationships between transposons and tandem repeat sequences in bovine subfamily species. Chinese Agricultural Science. 2022; 55: 1859–1867. (In Chinese)
- [28] Wang T, Chen Y, Li Z, Li RM. Analysis of codon preferences in the chloroplast genome of *Citrus aurantium*. Genomics and Applied Biology. 2023; 1–13. (In Chinese)
- [29] Shulakov AA, Egorov AV, Onipchenko VG. A phylogenetic analysis of plant communities of the Teberda State Biosphere Reserve. Biology Bulletin Review. 2017; 7: 238–245.
- [30] Jansen RK, Raubeson LA, Boore JL, dePamphilis CW, Chumley TW, Haberle RC, *et al.* Methods for obtaining and analyzing whole chloroplast genome sequences. Methods in Enzymology. 2005; 395: 348–384.
- [31] Ma WN, Ma YL, Wang JC, Xie HC, Ma YG. Exploring the phylogenetic relationships of Poppyidae based on rbcL and matK sequences. Qinghai Science and Technology. 2019; 26: 30–36. (In Chinese)
- [32] Xiao W, Simpson BB. Phylogenetic Analysis of Meconopsis (Papaveraceae) and Evaluation of Two Controversial Taxonomic Species. Lundellia. 2015; 18: 14–27.
- [33] Ren FM, Wang YW, Xu ZC, Li Y, Xin TY, Zhou JG, *et al.* DNA barcoding of *Corydalis*, the most taxonomically complicated genus of Papaveraceae. Ecology and Evolution. 2019; 9: 1934–1945.
- [34] Wang L, Dong WP, Zhou SL. Progress in structural variation of angiosperm chloroplast genomes. Northwest Journal of Botany. 2012; 32: 1282–1288. (In Chinese)
- [35] Przybylski P. Assessment of Variability: Chloroplast Microsatellite DNA, Defoliation, and Regeneration Potential of Old Pine Stands of Different Origins in the Context of Assisted Genotype Migration. Forests. 2022; 13: 1829.
- [36] Asaf S, Khan AL, Khan MA, Shahzad R, Lubna, Kang SM, *et al.* Complete chloroplast genome sequence and comparative analysis of loblolly pine (*Pinus taeda* L.) with related species. PloS One. 2018; 13: e0192966.
- [37] Zhao Y, Liu Z, Yang P, Cheng Y, Yang Y. Methods of codon bias analysis and progress of codon bias research in tea. Tea Communication. 2016; 43: 3–7.
- [38] Liu YJ, Tian XP, Li Q. Analysis of codon bias in chloroplast genome of *Fraxinus pennsylvanica*. Jiangsu Nongye Kexue (Jiangsu Agricultural Sciences). 2020; 48: 83–88. (In Chinese)
- [39] Yuan XL, Kang HM, Wang Y. Codon usage bias analysis of chloroplast genome in *Nyssa yunnanensis*. Xibeilin Xueyuan Xuebao (Journal of Northwest Forestry University). 2020; 35: 26–

31, 124. (In Chinese)

- [40] Liu XY, He ZJ, Qiu YM. Codon bias in the chloroplast genome of four rosaceae fruit trees. *Fenzi Zhiwu Yuzhong* (Molecular Plant Breeding). 2022; 20: 5299–5308.
- [41] Cai C, Xiao S, Liang X, Guo S. Analysis of codon preferences in the chloroplast genome of *Cordyceps militaris*. *Molecular Plant Breeding*. 2023; 1–18. (In Chinese)
- [42] Liu LX, Li R, Worth JRP, Li X, Li P, Cameron KM, *et al.* The Complete Chloroplast Genome of Chinese Bayberry (*Morella rubra*, Myricaceae): Implications for Understanding the Evolution of Fagales. *Frontiers in Plant Science*. 2017; 8: 968.
- [43] Sun Y, Moore MJ, Zhang S, Soltis PS, Soltis DE, Zhao T, *et al.* 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. *Molecular Phylogenetics and Evolution*. 2016; 96: 93–101.