BMC Plant Biology

RESEARCH

Open Access

Comparative and phylogenetic analysis of the chloroplast genomes of four commonly used medicinal cultivars of *Chrysanthemums morifolium*

Yingying Duan[1†], Yuqing Wang[1†], Wanyue Ding[1,2], Chun Wang[2], Ling Meng[2], Jie Meng[3], Na Chen[4,5], Yaowu Liu[5*] and Shihai Xing[1,4,6*]

Abstract

'Boju' and 'Huaiju' are cultivars of the *Chrysanthemum* (*Chrysanthemum morifolium* Ramat.) in the family Asteraceae, valued for their medicinal, tea, and ornamental properties, and valued by individuals. However, the yield and quality of medicinal *chrysanthemums* are limited by the characteristics of the germplasm resources, including the identification at the varieties and cultivation levels. Currently, research characterizing the chloroplast genomes of medicinal *Chrysanthemum* flowers is relatively limited. This study conducted chloroplast whole-genome sequencing on two cultivars of *Chrysanthemum*, 'Boju' and 'Huaiju', and compared them with the previously published chloroplast genomes of 'Hangbaiju' and 'Gongju'. The study analyzed the chloroplast genome structures of these four medicinal *Chrysanthemum*s, identifying mutation hotspots and clarifying their phylogenetic relationships. The chloroplast genome sizes of four medicinal *Chrysanthemum* cultivation products ranged from 151,057 to 151,109 bp, with GC content ranging from 37.45% to 37.76%. A total of 134 genes were identified, including 89 protein-coding genes, 37 ribosomal RNA genes, and 8 transfer RNA genes. Comparative genomic analysis revealed 159 large repeat sequences, 276 simple sequence repeats, 1 gene, and 8 intergenic regions identified as highly variable regions. Nucleotide diversity (Pi) values were high (≥ 0.004) for the *petN-psbM*, *trnR-UCU-trnT-GGU*, *trnT-GGU-psbD*, *ndhC-trnV-UCA*, *ycf1*, *ndhI-ndhG*, *trnL-UGA-rpl32*, *rpl32-ndhF*, and *ndhF-ycf1* fragments, aiding in variety identification. Phylogenetic analysis revealed consistent results between maximum likelihood and Bayesian inference trees, showing that the four medicinal *Chrysanthemum* cultivars, along with their wild counterparts and related species, evolved as a monophyletic group, forming a sister clade to *Artemisia* and *Ajania*. Among the six *Chrysanthemum* species, the wild *Chrysanthemum* diverged first (Posterior probability = 1, bootstrap = 1,000), followed by *Ajania*, while *C. indicum* and *C. morifolium* clustered together (Bootstrap = 100), indicating their closest genetic relationship. The chloroplast whole-genome data and characteristic information provided in this study can be used for variety identification, genetic conservation, and phylogenetic analysis within the family Asteraceae.

Keywords Asteraceae, *Chrysanthemum*, Chloroplast genome, Comparative genomics, Phylogenetic analysis

†Yingying Duan and Yuqing Wang contributed equally to this work.

*Correspondence:
Yaowu Liu
lywahljy@163.com
Shihai Xing
xshshihai@163.com
Full list of author information is available at the end of the article

Introduction

Chrysanthemum belongs to the genus Chrysanthemum of the family *Asteraceae*, and its dry flower heads are a commonly used traditional Chinese herbal medicine. Originating in China, *Chrysanthemum* has undergone over 3,000 years of development from wild to medicinal and ornamental use [1]. Throughout this extensive history, the medicinal, tea, culinary, and ornamental values of *Chrysanthemum* have gradually been explored, leading to a wide range of applications [2, 3]. For instance, ornamental *Chrysanthemum*s have become the leading cut flower among the four major types and the second most popular cut flower in the global flower market [4]. In comparison to ornamental *Chrysanthemum*s, medicinal *Chrysanthemum*s have a more ancient history. Since the Han Dynasty, medicinal *Chrysanthemum*s have gained widespread popularity. "Shennong Ben Cao Jing" categorizes them as a superior herb, while "Mingyi Bie Lu" records details such as their place of origin, harvesting time, and processing methods, thus confirming the medicinal value of *Chrysanthemum* [2]. Today, cultivated *Chrysanthemum* varieties such as 'Boju', 'Huaiju', 'Hangju', and 'Gongju' are all included in the latest edition of the Chinese Pharmacopoeia, representing typical medicinal *Chrysanthemum*s. *Chrysanthemum* contains active ingredients such as volatile oil [5], flavonoids [6], sugars, and hydroxycinnamic acid derivatives [7], which exhibit pharmacological effects including antibacterial, anti-inflammatory, antioxidant, antiviral, anti-tumor, and prevention of chronic diseases [3, 8]. In folk pharmacology, *Chrysanthemum* is widely believed to have the effects of soothing the liver, improving vision, and clearing heat and toxins [3], making *Chrysanthemum* highly valuable for scientific research. However, *C. morifolium* face taxonomic ambiguity, which significantly complicates the identification of medicinal varieties. The primary causes of this confusion stem from the vast diversity of chrysanthemum species, their similar morphological features, the prevalence of hybrids, and the wide range of cultivated varieties. Additionally, traditional identification methods are subjective and have inherent limitations [7, 9]. These challenges contribute to the widespread issue of adulteration, making quality control difficult and severely hindering scientific research and the effective utilization of medicinal *C. morifolium*. Therefore, molecular-level studies are essential for the accurate identification and authentication of medicinal *Chrysanthemum* species, as well as for revealing their molecular evolution and diversification processes.

There exists a vast array of plant species worldwide, yet the number that can be identified using traditional plant identification methods is relatively limited. The emergence of DNA barcoding technology has had a positive impact on this situation [10]. DNA barcoding technology is a method that uses short DNA sequences from either the nuclear genome or organelle genomes to identify biological samples. Compared to the mitochondrial genome, which evolves slowly and has a very low substitution rate in plants, and the nuclear genome, which exhibits higher heterozygosity and greater complexity in variation, the chloroplast genome offers several significant advantages in species identification. These advantages include higher stability, ease of sequencing, an intermediate mutation rate, and rich interspecies variation [11]. Therefore, in recent years, DNA barcoding has been widely used for the species identification, authentication, and phylogenetic analysis of medicinal plants, particularly in distinguishing morphologically similar or closely related species [12, 13]. While single locus DNA barcodes may face difficulties in identifying closely related species due to insufficient variation [14], whole chloroplast genome (cp genome) sequencing (super-barcoding) is more reliable when distinguishing between closely related varieties or even cultivated varieties compared to shorter chloroplast markers, owing to its broader sequence coverage and higher resolution [15]. For example, Wu et al. [16] utilized the whole chloroplast genome as a super barcode to successfully distinguish *Fritillaria* species listed in the Chinese Pharmacopoeia from their closely related species and counterfeit materials. This approach has also been widely applied in various other medicinal plants, yielding positive results [17–19]. Furthermore, the cp genome, characterized by highly conserved gene structure and sequence, as well as maternal inheritance, serves as an ideal molecular data resource widely employed in research fields such as phylogenetics, genetic engineering, breeding, and evolutionary biology [20, 21]. Chloroplasts are found in plants, algae, and some protists. They are double-membraned organelles and serve as the primary site of photosynthesis, converting light energy into chemical energy to provide energy for plant growth and development, as well as supplying carbon intermediates for many key metabolic reactions [22]. The cp genome of land plants exhibits highly conserved structure and content organization, typically presenting a typical circular tetrad structure composed of a large single-copy region (LSC) and a small single-copy region (SSC), separated by two inverted repeat sequences (IR) of equal length but in opposite directions [23–25]. Recent years have witnessed a spurt of progress in next-generation sequencing technologies and phylogenomics, cp genome sequencing has been widely applied in molecular evolution and phylogenetic analysis studies of many plant species [23, 26]. Through the integration of cp genomics and phylogenomics, more accurate classification and phylogenetic relationships of *Chrysanthemum*s can be obtained.

Despite extensive research has been conducted on the botanical ecology, active ingredients, clinical pharmacology, and folk pharmacology of medicinal chrysanthemums, issues such as taxonomic ambiguity, difficulties in quality control, and the lack of phylogenetic analysis persist. Therefore, chloroplast genome research is highly necessary. The whole chloroplast genome, as a super barcode, has shown great promise in distinguishing closely related species and counterfeit medicinal materials, especially in tightly related taxa [10]. Additionally, specific genes or non-coding regions in the chloroplast genome are relatively conserved but exhibit sufficient variation, making them suitable as candidate DNA barcodes to provide a foundation for future phylogenetic and evolutionary studies. This study, based on the Chinese Pharmacopoeia, utilized the chloroplast genomes of two newly sequenced chrysanthemum cultivars, 'Boju' and 'Huaiju,' along with the publicly available chloroplast genome data of two medicinal chrysanthemums, 'Hangbaiju' and 'Gongju.' The sequences were assembled, annotated, and compared to provide a detailed description of their basic characteristics. Nucleotide diversity was assessed, and highly variable regions were identified. Additionally, the evolutionary relationships were systematically studied at the cp genome sequence level. These four chrysanthemum cultivars, included in the 2020 edition of the "Chinese Pharmacopoeia", have a long history of medicinal use and are widely recognized as representative medicinal varieties [27]. The findings of this study can serve as a reference for the classification of medicinal *Chrysanthemum*s, identification of *Chrysanthemum* varieties, quality assessment, phylogenetic analysis, and development of molecular markers in the future.

Materials and methods

Plant Material

Fresh and well-grown specimens of 'Boju' and 'Huaiju' were collected from the Medicinal Plant Garden of Anhui University of Chinese Medicine (latitude N31°56′58″, longitude E117°23′38″). Voucher specimens of both cultivars were collected and stored at the Herbarium Centre of Anhui University of Traditional Chinese Medicine, Hefei, China. (AhtcmH, yxy.ahtcm.edu.cn/info/1006/6713.htm, Shi-hai Xing, xshshihai@163.com, under the voucher numbers 20231005 and 20,231,004). The complete cp genome data of the medicinal *Chrysanthemum* cultivars 'Hangbaiju' and 'Gongju' were downloaded from NCBI (Supplemental Table S1). In the phylogenetic analysis section, cp complete genome sequences of species such as *Arabidopsis thaliana*, *Salvia miltiorrhiza*, *Scutellaria baicalensis*, *Stylidium debile*, *Stylidium petiolare*, *Artemisia princeps*, *Artemisia stechmanniana*, *Ajania nematoloba*, and *Ajania tenuifolia* were all downloaded from NCBI.

DNA extraction and genome sequencing

Total genomic DNA from leaves of *C. morifolium* was extracted using a modified CTAB method with a plant genomic DNA extraction kit. The DNA concentration and quality were evaluated using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and 1% agarose gel electrophoresis. The extracted DNA underwent further fragmentation and end repair to prepare the genomic libraries. The constructed libraries were subsequently sequenced on the Illumina HiSeq×high-throughput platform (Illumina, San Diego, CA, USA). Library preparation and sequencing were performed by Genewiz (Soochow, China). The raw reads obtained were subjected to quality control using the Cutadapt (V1.9.1) software for second-generation sequencing data. This process involved removing reads containing adapter sequences, contaminants, and low-quality reads to obtain clean reads.

Cp genome assembly and annotation

Based on the clean reads obtained from sequencing, NOVOPlasty (version 2.7.2) [28] was employed for cp genome assembly. To enhance the accuracy of assembly, additional software was utilized to optimize the assembly results. The contig sequences were further assembled into scaffold sequences using SSPACE software (version 3.0) [29]. Subsequently, GapFiller software (version 1–10) [30] was used to fill gaps in the scaffold sequences. After alignment and correction, the complete cp genome sequence was obtained. After assembly, we initially annotated the cp genome using DOGMA and GeSeq [31, 32], with the *Chrysanthemum* whole genome as the reference sequence. During the annotation process, special attention was given to the trans-spliced gene rps12, short exon genes, pseudogenes, and RNA editing genes to prevent annotation errors [33]. Manual corrections to the annotation were performed using Geneious R11. The tRNA annotation was conducted using tRNAscan-SE (v2.0.7) [34]. Finally, the physical map of the cp genome was constructed using the OGDraw program [35]. The annotated sequences of the cp genomes of *C. morifolium* cultivar 'Boju' and *C. morifolium* cultivar 'Huaiju' were deposited in GenBank with accession numbers PP979455 and PP908960, respectively.

Comparative analysis of the Cp genomes of the four cultivars

Comparison analysis of the CPGs

Using the 'Boju' cp genome as a reference with annotations, we conducted a visual comparative analysis of the overall similarity of cp genomes among four *Chrysanthemum* cultivars using the mVISTA software [36] in Shuffle-LAGAN mode. To evaluate the variation of the inverted repetition (IR) regions more comprehensively and the contraction and expansion of boundaries in the cp genomes of four cultivars of *Chrysanthemum*, this study compared the LSC/IRb, IRb/SSC, SSC/IRa, IRa/LSC junctions, and adjacent gene arrangements in their cp genomes. Visualization of the four major region boundaries in the cp genomes of the four cultivars was achieved using IRscope [37].

SSRs and repeat sequence analysis

The REPuter [38] (https://bibiserv.techfak.uni-bielefeld.de/reputer) online tool was utilized to visualize the quantity and types of scattered repeat sequences in each cp genome, including forward, reverse, complement, and palindromic sequences. A Hamming distance of 3 was set, with a minimum repeat unit size of 30 bp and a maximum repeat count of 50 bp, to identify the types of repeat sequences in the cp genomes of the four *Chrysanthemum* cultivars. Simple sequence repeats (SSRs), formed by tandem repeats of 1-6 bp-long repeating units, serve as informative loci for molecular marker analysis and are frequently employed in population genetic studies of species. The cp genomes of four medicinal *Chrysanthemum* cultivars were analyzed for simple sequence repeats using MISA-web [39]. The parameters for minimum repeat counts were set at 10, 5, 4, 3, and 3, corresponding to mononucleotide to hexanucleotide repeats.

Analysis of codon preference

We utilized the CodonW program (https://codonw.sourceforge.net/) to calculate the frequency of synonymous codon usage for each encoded amino acid in four *Chrysanthemum* cultivars. Subsequently, we performed Relative Synonymous Codon Usage (RSCU) cluster analysis using TBtools v1.0987663 software (https://github.com/CJ-Chen/TBtools/releases) [40]. Relative Synonymous Codon Usage (RSCU) is calculated by comparing the observed frequency of specific codons within the coding sequence to the expected frequency within their synonymous codon family. An RSCU value close to 1.0 indicates no significant bias [41]. RSCU values greater than 1.0 denote high-frequency codons, suggesting a preference for those codons due to their higher usage frequency.

Analysis of highly variable areas

To identify highly variable regions, we aligned the extracted conserved genes and intergenic regions using MAFFT, followed by sliding window analysis using DnaSP software [42]. We calculated nucleotide diversity and polymorphic sites in the cp genomes of four medicinal *Chrysanthemum* cultivars to assess the variability of variable sites in the highly variable regions. In the nucleotide polymorphism analysis, a search window length of 600 bp and a step length of 200 bp were set. Hotspot variable regions were selected based on the magnitude of the Pi value, and R programming was used for visualization.

Phylogenomics analysis

To gain a deeper understanding of the phylogenetic relationships among medicinal *Chrysanthemum*s (*Chrysanthemum* spp.) and their related genera and closely related species, we downloaded cp genome sequences of other species from the Asteraceae family and closely related families from the NCBI database as supplementary data for phylogenetic analysis (Supplementary Table S1). After aligning all sequences using MAFFT v7.50 [43], we trimmed and aligned the sequences using trimAI v1.41 [44]. *A. thaliana* was used as an outgroup. We conducted maximum likelihood (ML) analysis using RaxML [45] software under the GTR + GAMMA [46] model and performed 1000 bootstrap replicates to assess the robustness of the tree topology. Additionally, a Bayesian tree was constructed using MrBayes software through Bayesian inference. The posterior probabilities of the phylogenetic model were estimated using the Markov Chain Monte Carlo (MCMC) method, running for 100,000 generations, with one tree saved every 100 generations for posterior analysis. The MCMC analysis started with a random tree and then constructed a consensus tree after discarding burn-in samples.

Results and discussions

General characteristics of the Cp genome

In this study, clean reads of the cp genomes of 'Boju' and 'Huaiju' were obtained through Illumina sequencing. The cp genomes of four medicinal *Chrysanthemum* cultivars range in size from 151,057 bp (Boju) to 151,109 bp (Hangbaiju) and exhibit a typical circular structure (Fig. 1). Among them, the shortest sequence is 151,057_bp for 'Boju,' and the longest sequence is 151,109 bp for 'Hangbaiju.' After optimizing the raw data, the effectiveness rates of Q20 and Q30 surpassed 92%. The sequencing quality is deemed to be excellent. The cp genomes of the four *Chrysanthemum* cultivars exhibit a typical quadruple structure, with a large single-copy region (LSC) (82,806 bp-82,858 bp) and a small single-copy

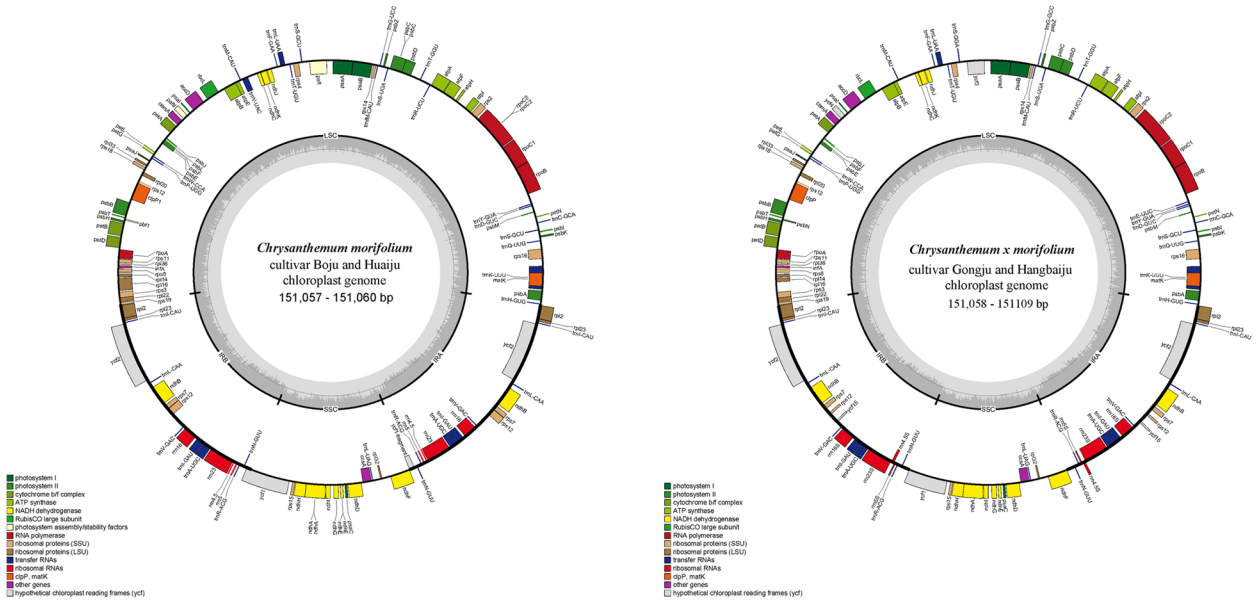

Fig. 1 Chloroplast gene map of four medicinal *Chrysanthemum* cultivars. Genes are color-coded by function. GC and AT content are represented by shades of gray. Thick lines denote boundaries between small single-copy (SSC) and large single-copy (LSC) regions, marked by reverse repeat sequences (IRa and IRb)

region (SSC) (18,293 bp-18,350 bp) separated by a pair of inverted repeat regions (IR) (24,953 bp-24,957 bp). The GC content of the cp genomes of 'Boju', 'Huaiju', 'Hangbaiju', and 'Gongju' is 37.76%, 37.75%, 37.46%, and 37.45%, respectively. This is significantly lower than the AT content, consistent with the characteristic abundance of AT in cp genomes. The GC content of the IR regions is 43.07%, which is higher compared to the LSC (35.54–35.57%) and SSC (30.77–30.82%) regions. This indicates that the IR regions possess higher stability than the SSC and LSC regions (Table 1).

Annotation was performed on the cp genomes of 'Hangbaiju', 'Gongju', and the cultivars sequenced in this study, using data downloaded from the NCBI database. It was found that based on gene count, the four medicinal *Chrysanthemum* cultivars can be divided into two categories. One category includes 'Hangbaiju' and 'Gongju', both of which contain 127 genes in their cp genomes, comprising 85 CDS genes, 34 tRNA genes, and 8 rRNA genes. The other category includes 'Boju' and 'Huaiju', which contain 125 genes, including 83 CDS genes, 34 tRNA genes, and 8 rRNA genes (Table 1). The differential genes identified among the four medicinal *Chrysanthemum* cultivars include *clpP1*, *pbf1*, *trnS-GGA*, *psbL*, *pafII*, *psaI*, *trnV-UAC*, *trnfM-CAU*, *trnG-UCC*, *psbN*, *clpP*, *ycf4*, *ycf3*, *trnG-GCC*, *trnE-UUC*, and *ycf15*. Except for *ycf15*, which is located in the IR region of 'Hangbaiju' and 'Gongju', the remaining differential genes are situated in the LSC region. The functional analysis of the cp genome genes can be categorized into four groups: self-replication group, photosynthesis group, other functions, and unknown function group (Table 2).

Table 1 Summary of the CPG features of four medicinal *chrysanthemum*

Items		Boju*	Huaiju*	Hangbaiju	Gongju
Total Size(bp)		151,057	151,060	151,109	151,058
LSC (bp)		82,806	82,858	82,851	82,858
SSC (bp)		18,334	18,293	18,350	18,292
IR (bp)		24,957	24,953	24,954	24,954
GC (%)	Total (%)	37.47	37.46	37.46	37.45
	LSC (%)	35.57	35.55	35.54	35.54
	SSC (%)	30.82	30.78	30.78	30.77
	IR (%)	43.07	43.07	43.07	43.07
Gene number		125	125	127	127
PCG		83	83	85	85
tRNA		34	34	34	34
rRNA		8	8	8	8

Chrysanthemum cultivars with Newly Sequenced Chloroplast Genomes Annotated with Asterisks (*)

LSC for Large Single Copy, SSC for Small Single Copy, IR for Inverted Repeat Sequences, GC for Guanine-Cytosine Content, and PCG for Protein-Coding Genes

Table 2 Genes present in the chloroplast genomes of the four medicinal *Chrysanthemum*

Genes Category	Gene Groups	Gene Name
Self-replication	Ribosomal RNAs	*rrn16(2), rrn23(2), rrn4.5(2), rrn5(2)*
	Transfer RNAs	*trnP-UGG, trnW-CCA, trnM-CAU, trnV-UAC*[a], *trnF-GAA, trnL-UAA*[a], *trnT-UGU, trnS-GCU, trnfM-CAU, trnG-UCC, trnS-UGA, trnT-GGU, trnR-UCU, trnY-GUA, trnD-GUC, trnC-GCA, trnS-GCU, trnQ-UUG, trnK-UUU*[a], *trnH-GUG, trnL-UAG, trnG-GCC, trnE-UUC, trnI-CAU*[b], *trnL-CAA*[b], *trnV-GAC*[b], *trnI-GAU*[ab], *trnA-UGC*[ab], *trnR-ACG*[b], *trnN-GUU*[b]
	Small ribosomal subunit	*rps19, rps3, rps8, rps11, rps18, rps4, rps14, rps2, rps16*[a], *rps15, rps7*[b], *rps12*[ab]
	Large ribosomal subunit	*rpl16*[a], *rpl14, rpl36, rpl20, rpl33, rpl32, rpl2*[ab], *rpl23(2)*[b]
	RNA polymerase	*rpoA, rpoC2, rpoC1, rpoB*
Photosynthesis	Photosystem I	*psaJ, psaI, psaA, psaB, psaC*
	Photosystem II	*psbH, psbT, psbB, psbE, psbF, psbL, psbJ, psbZ, psbC, psbD, psbM, psbI, psbK, psbA, psbN*
	RuBisCO	*rbcL*
	photosystem assembly/stability factors	*pafI*[a],*pafII*
	Cytochrome b/f complex	*petD*[a],*petB*[a],*petG, petL, petA, petN*
	ATP synthase	*atpB, atpE, atpA, atpF*[a],*atpH, atpI,*
	DNADH-dehydrogenase	*ndhC, ndhK, ndhJ, ndhF, ndhD, ndhE, ndhG, ndhI, ndhA*[a],*ndhH, ndhB*[ab]
Other genes	Subunit of acetyl-CoA-carboxylase	*accD*
	cytochrome c biogenesis	*ccsA*
	Maturase	*matK*
	Envelope membrance protein	*cemA*
	ATP-dependent Clp protease	*clpP1*[a], *clpP*[a]
	translation initiation factor	*infA*
Genes of unknown function	Conserved hypothetical chloroplast ORF	*Ycf2*[b], *ycf15*[b], *ycf1, ycf4, ycf3*[a]

Gene[a]: Intron-containing gene, Gene[b]: gene located in IR regions, Gene[ab]: The gene contains introns and is located in the IR region

Comparison analysis of the Cp genomes

Structural analysis of the Cp genomes

The complete cp genome data alignment enables a more accurate exploration of sequence differences among *Chrysanthemum* cultivars. The IR region is the most conserved region in cp genomes, and the size of the cp genome in angiosperms is closely related to the expansion and contraction of the IR and SC boundaries. We conducted sequence alignment using the complete cp genomes of the four medicinal *Chrysanthemum* cultivars and visualized it with mVISTA. From the results, it can be observed that the cp genome sequences of the four medicinal *Chrysanthemum* cultivars exhibit overall high similarity. No significant genome rearrangements or insertions were observed, indicating that the genome structure is relatively conservative at the molecular level, and the four *Chrysanthemum* cultivars have close phylogenetic relationships. It is noteworthy that the degree of differentiation in the IR region is smaller than that in the LSC/SSC regions, but the non-coding region sequences show more obvious differences than the coding regions. Variations mainly occur in intergenic regions, such as *psbA-trnH-GUG*, *psaA-pafI*, *ndhC-trnV-UCA*, *rbcL-accD*, *psbE-petL*, *rps3-rpl22*, and *ndhI-ndhG* (Fig. 2). Among them, the most significant changes in coding regions were observed in the ndhI-ndhG intergenic region. These highly variable genes and intergenic spacer regions can serve as molecular markers for phylogenetic and population genetic studies of the four *Chrysanthemum* cultivars.

Meanwhile, we used IRscope to observe the expansion and contraction of the cp genomes of the four medicinal *Chrysanthemum* cultivars (Fig. 3). Analysis of the cp genome boundaries of the four *Chrysanthemum* cultivars reveals that, compared to the 52-58 bp length variations observed in the SSC and LSC regions, their IR lengths are relatively conservative, with no significant contraction or expansion, and only minor differences in a few base pairs (4_bp). The LSC-IRb, SSC-IRb, SSC-IRa, and LSC-IRa boundaries in the cp genomes of the four *Chrysanthemum* cultivars are located within six genes: *rps19*, *ycf1*, *ndhF, trnN*, *rpl12*, and *trnH*. Both *rps19* and *ycf1* span the LSC-IRb and SSC-IRb regions in all four *Chrysanthemum* cultivars. Of particular note is the *ycf1* gene located at the IRb-SSC boundary, primarily situated in the SSC region, with a total length ranging from 4,457 bp to 4,458 bp. In 'Boju' and 'Huaiju', the *ycf1* gene has a copy at the IRb/SSC junction, overlapping with the

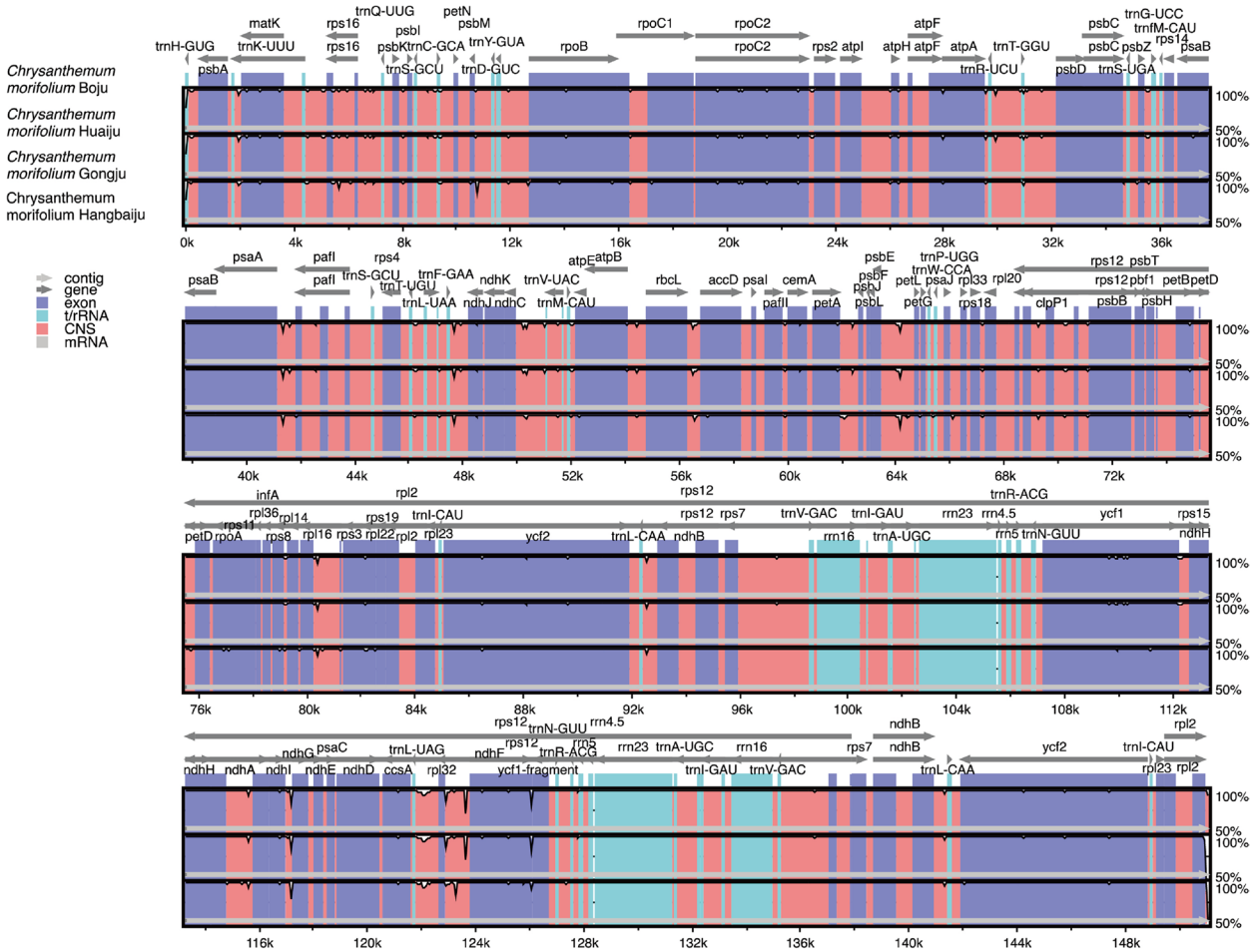

Fig. 2 Sequence alignment of four medicinal *chrysanthemum* chloroplast genomes (CPGs). Alignment was performed using the mVISTA program with *C. morifolium* Boju chloroplast genome as reference. Coding and non-coding regions are indicated in blue and red, respectively. The y-axis represents the degree of consistency from 50 to 100%. Black arrows denote the position and direction of each gene. CNS: Conserved non-coding sequences

IRa boundary, resulting in a pseudogene fragment of *ycf1* entirely located within the IRa region, with a length of 558 bp, shorter than the *ycf1* gene.

Codon usage bias

Codon bias is formed during the long-term evolution of organisms and involves a complex set of mechanisms. There are 61 synonymous codons encoding 20 amino acids in the cp genomes of the four cultivars (Fig. 4). Among the 20 amino acids, leucine is the most abundant in the cp genomes of the four *Chrysanthemum* cultivars (2122–2162 codons, accounting for 5.26–5.33%). Next is isoleucine, which is also encoded by six codons (1671–1715 codons, 4.20–4.23%). Cysteine has the lowest proportion, accounting for only 0.53–0.54% (211–218 codons) (Supplemental Table S2) The number of high-frequency codons (RSCU > 1) is 29, with 28 of them ending with (A/U) bases, accounting for 96.5%. This indicates that the high-frequency codons in the cp genomes of these four *Chrysanthemum* cultivars tend to end with (A/U) bases. There are 30 low-frequency codons with RSCU values less than 1, and 28 of them end with G/C as the third base, accounting for 93.3%. This suggests that low-frequency codons in the four cp genomes tend to end with G/C. The codon usage preferences are relatively similar among the four cultivars, with small differences observed between them.

Analysis of the SSRs and long repeat

Simple sequence repeats, also known as microsatellite sequences or tandem repeats, are characterized by their abundance and the presence of multiple alleles. They are widely used in DNA molecular markers and plant genotyping due to their abundance and polymorphic nature. Upon detection, a total of 276 simple sequence repeats (SSRs) were identified in the cp genomes of the four

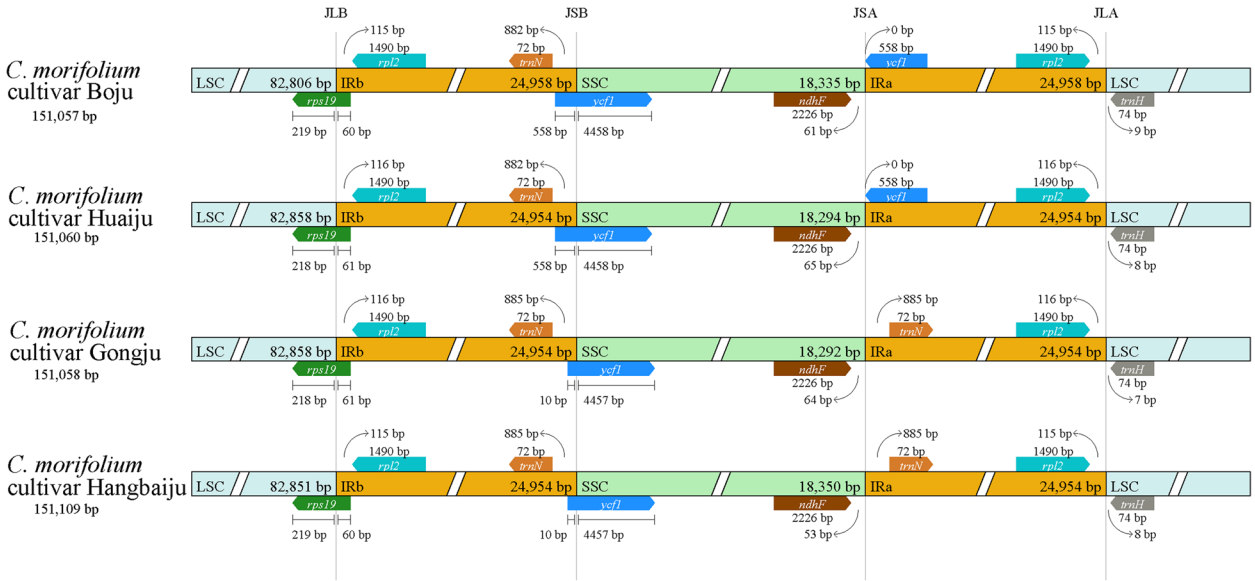

Fig. 3 Comparison of the boundaries of the large single copy (LSC), small single copy (SSC), and inverted repeat (IR) regions among the chloroplast genomes (CPGs) of four medicinal *chrysanthemum* cultivars. Gaps between the end of each boundary and the adjacent genes are depicted in base pairs (bps) above the main line

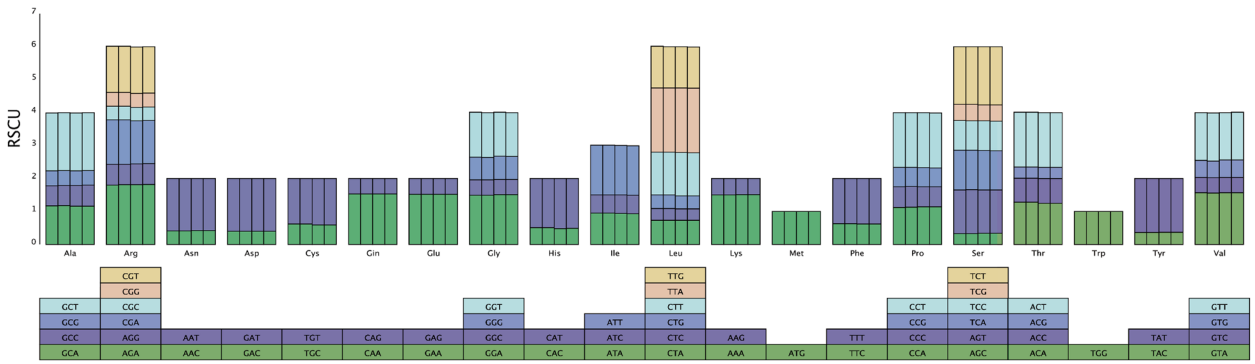

Fig. 4 Relative synonymous codon usage (RSCU) histogram of four medicinal *chrysanthemum* cultivars. Each block below represents a different codon encoding an amino acid. The columns at the top depict the total RSCU values for the 20 amino acids

Chrysanthemum cultivars. Mononucleotide repeats were the most abundant, comprising 156 repeats (56.52%), followed by tetranucleotide repeats (60 repeats, 21.74%) and dinucleotide repeats (34 repeats, 12.32%). Trinucleotide and pentanucleotide repeats were each detected in 15 and 11 instances, respectively, while no hexanucleotide repeats were detected (Fig. 5A) (Supplemental Table S3). The distribution of SSRs in the four repetitive regions of the four *Chrysanthemum*s is generally consistent, with the highest number of SSRs observed in the LSC region (73.24–77.94%), followed by the SSC region (11.59–13.24%). The distribution of SSRs in the IRa and IRb regions is relatively low and equal (4.41–7.04%) (Supplemental Figure S1). The SSRs identified in the cp genomes of the four *Chrysanthemum* cultivars are primarily composed of A and T nucleotides, with over half of the SSR sequences consisting solely of A or T bases. Among the four *Chrysanthemum* cultivars, 'Hangbaiju'has the highest proportion of mononucleotide repeats (57.75%), followed by 'Boju' (56.52%), while 'Huaiju' and 'Gongju' have equal proportions of mononucleotide repeats (55.88%). In mononucleotide repeats, the most abundant type in 'Hangbaiju' is T (26 repeats), followed by 'Boju' (23 repeats), 'Huaiju' (21 repeats), and 'Gongju' (21 repeats) (Fig. 5B).

Using REPuter to analyze the large repeat sequences in the cp genomes of the four medicinal *Chrysanthemum* cultivars, a total of 161 long sequence repeats were obtained, including 75 forward repeats and 86

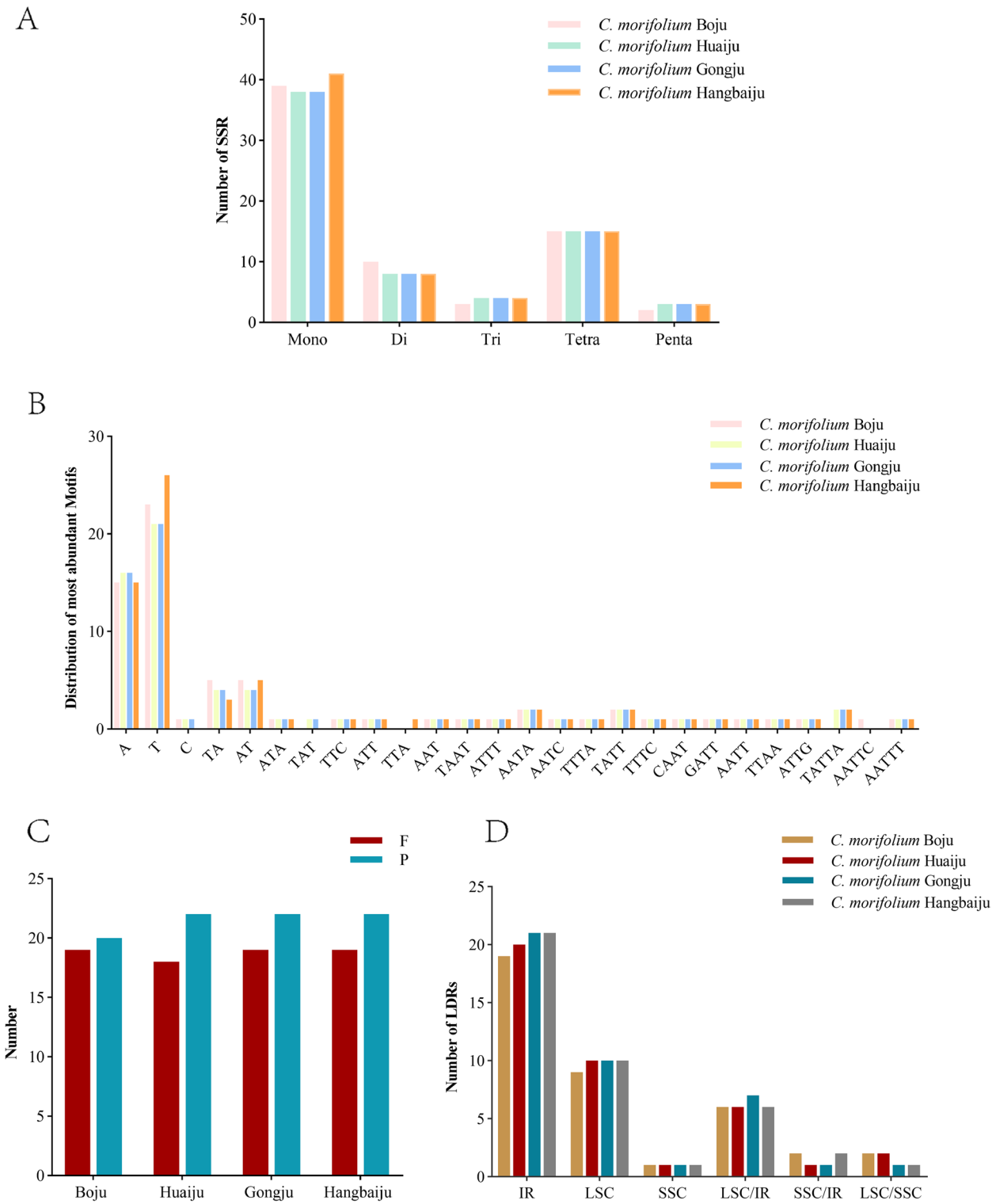

Fig. 5 The simple sequence repeats and long sequence repeats detected in four medicinal chrysanthemum cultivars. **A** Types of simple sequence repeats, **B** Repeating units of simple sequence repeats, **C** Types of long sequence repeats, **D** Localization of long sequence repeats

palindromic repeats. No complement repeats or reverse repeats were found in any of the four *Chrysanthemum* cultivars. Among the four *Chrysanthemum* cultivars, the fewest forward repeat sequences were detected in 'Huaiju' (18 repeats), while 'Boju', 'Gongju', and 'Hangbaiju' each had 19 forward repeat sequences. 'Boju' had the fewest palindromic repeat sequences (20 repeats), while the other three *Chrysanthemum* cultivars each had 22 palindromic sequences (Fig. 5C). Among these long repeat sequences, those distributed in the IR region had the highest proportion, followed by the LSC region. Additionally, some repeat sequences were shared among the SSC/IR/LSC regions (Fig. 5D).

Identification of highly variable in Cp genome

Nucleotide polymorphism is an important indicator for assessing diversity and genetic variation among populations or related species. In this study, we analyzed the complete sequences of four medicinal *Chrysanthemum* cultivars using DNAsp, calculated nucleotide polymorphism values, and performed statistical analysis of the results (Fig. 6). The results indicate that the Pi values of the four medicinal *Chrysanthemum* cultivars range from 0 to 0.01194, with an average value of 0.000721. The sequence with the highest level of variation is the ndhI-ndhG intergenic region, with a Pi value of 0.01194, and it contains 11 variable sites (Supplemental Table S4). The average Pi value of the SSC region is 0.001567, higher than that of the LSC region (0.000861) and the IR region (0.000176). In addition, nine highly variable regions (Pi > 0.04) were identified among these cultivars, including one gene, *ycf1*, and eight intergenic regions: *petN-psbM*, *trnR-UCU-trnT-GGU*, *trnT-GGU-psbD*, *ndhC-trnV-UCA*, *ndhI-ndhG*, *trnL-UGA-rpl32*, *rpl32-ndhF*, and *ndhF-ycf1* fragments. The high Pi values indicate that these regions have a high degree of variation, providing informative loci for high variation. These highly variable regions can be used to study the evolutionary characteristics of cp genome fragments in medicinal *Chrysanthemum* cultivars and develop molecular markers. They can provide valuable insights into understanding the genetic diversity and evolutionary processes of *Chrysanthemum* species.

Phylogenetic analysis

The cp genome, characterized by its conservative sequence structure and moderate nucleotide substitution rate, possesses significant advantages in phylogenetic

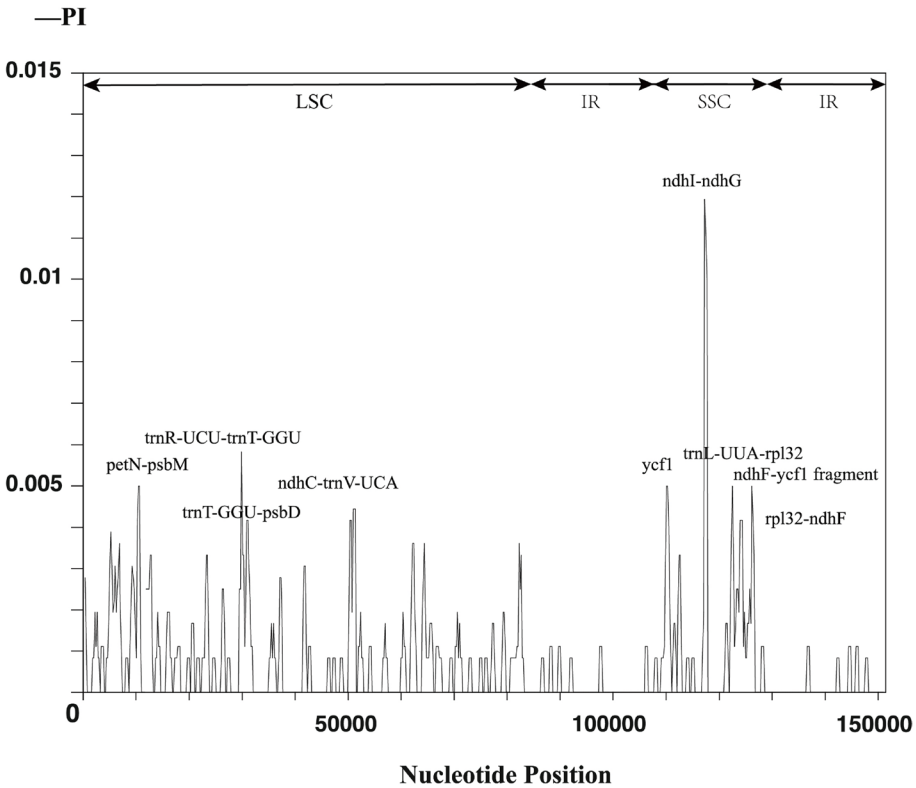

Fig. 6 Sliding window analysis of nucleotide diversity (Pi) of the chloroplast genomes of four medicinal *chrysanthemum* cultivars. The X-axis represents the position of the midpoint of each window, while the Y-axis represents the nucleotide diversity of each window

analysis research [47]. To determine the phylogenetic position of the four medicinal *Chrysanthemum* cultivars, this study used *A. thaliana* chloroplast sequences as outgroups. Combined with the cp genome sequences of the four medicinal *Chrysanthemum* cultivars and ten other species from the order Asterales and the Asteranae, Bayesian inference (BI) and maximum likelihood (ML) methods were employed to construct phylogenetic trees (Fig. 7). The results indicate that the phylogenetic trees constructed by these two methods are consistent, demonstrating the good stability of the phylogenetic tree (Supplemental Figure S2). The evolutionary tree is classified into three branches. The four medicinal *Chrysanthemum* cultivars, along with the congeneric *C. dichroum* and *C. indicum*, form one clade. They are sister groups to Artemisia species from the same family, including *A. princeps* and *A. stechmanniana*, as well as to the genus Ajania species, *A. nematoloba* and *A. tenuifolia*, from the same subfamily, forming a monophyletic branch that significantly differs from other families and outgroup species, with a support rate of 100%. There are significant differences in branching patterns compared to species from the family Stylidiaceae (Asterales), the family Lamiaceae (Asteranae), and the outgroup *A. thaliana*. These results indicate a clear understanding of inter-family relationships in plants and further confirm the accuracy of our clustering.

Discuss

In this study, comparative and phylogenetic analyses were conducted on four cp genomes. The four medicinal *Chrysanthemum* cultivars belong to the genus *Chrysanthemum* and are closely related in evolutionary terms. Comparative cp genomic results indicate that the cp genomes of the four *Chrysanthemum* cultivars are similar in length and structure, exhibiting a typical quadripartite structure. The genome lengths range from 151,057 to 151,109 bp, with a gene count ranging from 125 to 127, which is consistent with previously reported cp genomes of cultivated *Chrysanthemum* species [48, 49]. The overall GC content of the four cp genomes is similar, ranging from 37.45% to 37.76%. The GC content in the inverted repeat (IR) regions (43.07%) is significantly higher than that in the large single-copy (LSC) region (35.54–35.57%) and small single-copy (SSC) region (30.77–30.82%). This difference may be attributed to the presence of RNA genes with higher GC content in the IR region.

The evolutionary rate of plant cp genomes (CPGs) is only half that of nuclear DNA sequences, possibly due to lower mutation rates, and the mutation rate of genes in the inverted repeat (IR) region is significantly lower than that of genes in the single-copy regions [50]. Comparative analysis of the CPG sequences of the four medicinal *Chrysanthemum* cultivars reveals that the main differentiation occurs in the large single-copy (LSC) and small single-copy (SSC) regions, with non-coding regions showing significantly more differences than coding regions. A total of seven intergenic regions have been identified, and these regions exhibit significant differences among the CPGs of the four *Chrysanthemum* cultivars (Fig. 2). Shrinkage and expansion of the inverted repeat (IR) region are important factors contributing to length variations in plant cp genomes, with reports of IR copy loss during the evolutionary process [51]. In boundary analysis, expansion of the IR region was observed at the IR/single copy (SC) boundary, resulting in the presence of a pseudogene, *ycf1*, at the IR/small single-copy (SSC) boundary in *Ajania* and *Artemisia*, consistent with findings in *Ricinus* [52] and Gossypium [53]. Such subtle

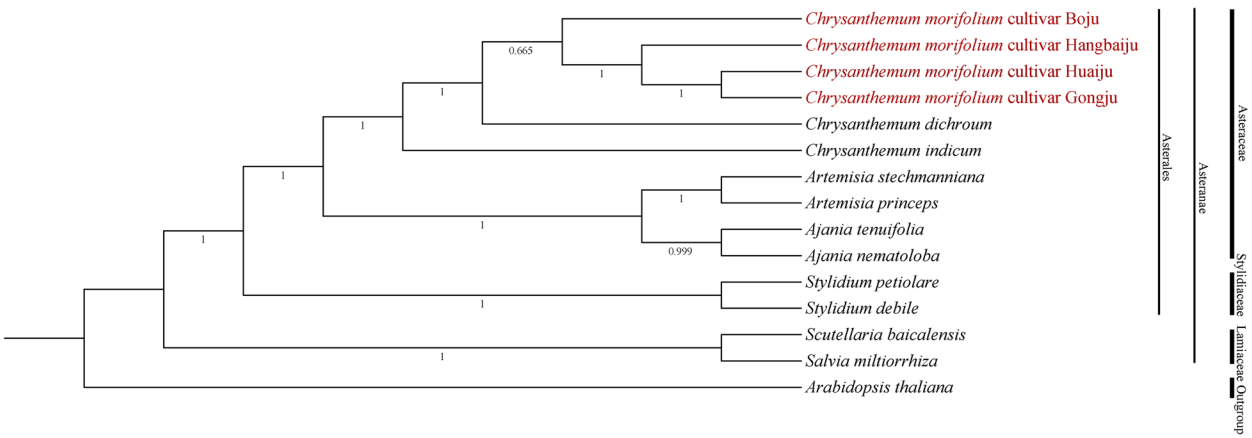

Fig. 7 Phylogenetic relationships based on four cultivars of *Chrysanthemums* and ten species from the order Asterales and *Asteranae*, with *A. thaliana* as an outgroup

boundary differences can also serve as features for cultivar identification.

Among the 20 amino acids encoded in plant genomes, all except methionine (MET) and tryptophan (TRP) are encoded by two or more codons, leading to high variability in codon usage. Current reports suggest that codon usage bias evolves through mutation and natural selection, reflecting the origin, mutation patterns, and evolution of species or genes. Codon usage bias is closely associated with genome GC content, gene expression levels and patterns, gene length, and the position of codons within genes [54, 55]. In this study, the codon usage patterns of the cp genomes of the four *Chrysanthemum* cultivars are similar. This similarity may be attributed to their close evolutionary relationship as cultivated *Chrysanthemum* species. Additionally, the four cp genomes exhibit high similarity in GC content, the number and position of introns and exons, gene types, and lengths.

Currently, DNA barcoding is becoming increasingly popular in various fields such as medicine, tea, and food. Since the advent of DNA barcoding technology, multiple loci have been proposed as DNA barcodes for plants, including coding genes and non-coding intergenic regions from both nuclear and plastid genomes. Initially, the combination of *rbcL* and *matK* was proposed as the core barcode for land plants. However, this recommendation is only applicable to a relatively small number of species [56]. For flowering plants, the *trnH-psbA* intergenic spacer is often regarded as a potential DNA barcode [57]. W. John Kress [58] et al. demonstrated, by validating eight presumptive plant barcodes (*trnH-psbA, rbcL-a, accD, matK, ndhJ, rpoB2, rpoC1*, and *ycf5*), that multi-locus combinations are more effective for species differentiation than single loci. Techen N [12]. et al. also supported this view, which suggests a potential trend towards the development of multi-locus combinations in plant DNA barcoding to achieve higher species identification accuracy. In addition, in the *Asteraceae* family, regions such as *ycf1*, *rbcL*, *psbA-trnH*, *trnC-ycf6*, *ycf6-psbM*, *rpl16*, and *trnL-F* have been utilized as DNA markers [59–61]. Highly variable regions, which demonstrate strong discriminatory power at the species level, are considered potential candidates for DNA barcodes. In the four medicinal *Chrysanthemum* cultivars, we have precisely identified nine highly variable regions with high pi values, including one gene, *ycf1*, and eight intergenic regions: *petN-psbM*, *trnR-UCU-trnT-GGU*, *trnT-GGU-psbD*, *ndhC-trnV-UCA*, *ndhI-ndhG*, *trnL-UGA-rpl32*, *rpl32-ndhF*, and *ndhF-ycf1* fragment. However, these mutation hotspots have not yet been analyzed in the classification group of the *Chrysanthemum* genus. Among these regions, the *ycf1* gene deserves particular attention. Reports have already confirmed the high resolution of *ycf1* at the species level, affirming its pivotal role in plant plastid barcoding [62]. Additionally, Kim et al. have suggested that *accD* and *ycf1* have the potential to become core molecular markers for the Asteraceae family [63]. In this study, ycf1 is the only gene found in the highly variable region rather than intergenic regions. In our research, we observed that in the cp genomes of 'Boju' and 'Huaiju', *ycf1* spans the small single-copy (SSC) and inverted repeat (IR) regions. Moreover, *ycf1* exhibits high variability among the four *Chrysanthemum* cultivars, which warrants further exploration of its potential as a DNA barcode.

In addition to the highly variable regions, we also detected two types of repetitive sequences (SSRs and LSRs) in the four cp sequences. SSRs are significant tools in evolutionary population biology due to their polymorphic and widespread nature. They hold particular importance for population genetic analysis among closely related taxa [64]. SSR (Simple Sequence Repeats) and LSR (Long Sequence Repeats) play a crucial role in understanding evolution through their effects on genetic variation and genome structure. SSRs, as microsatellites, are key genetic markers for analyzing variation and evolutionary patterns, and are vital in estimating migration rates, population size, and kinship [65]. SSRs have also been instrumental in identifying new hybrids and verifying their parental species [66, 67]. In Chrysanthemum, studies utilizing SSR markers to evaluate genetic diversity have explored the clustering of varieties based on geographic origin, offering insights into species diversity and population structure [68]. LSRs, due to their larger repetitive segments, induce structural variations in chloroplast genomes, significantly contributing to the study of evolutionary relationships and genomic differences among populations [69]. LSRs, due to their larger repeat segments, can induce structural variations in chloroplast genomes and hold significant importance in studying evolutionary relationships between populations and differences in genome structure [69]. In this study, the majority of SSRs detected in the cp genomes of the four *Chrysanthemum* cultivars were mononucleotide repeats, with A and T being predominant. This could be attributed to the fact that plastid SSRs typically consist of poly-A and poly-T repeats and rarely contain guanine (G) and cytosine (C) repeat sequences. Additionally, A/T bases have fewer hydrogen bonds and lower energy consumption [70–72].

The widespread occurrence of hybridization and polyploidy complicates the interspecific relationships of *Chrysanthemum*s [73, 74]. Cp genomes, characterized by their relatively small size, are typically maternally inherited in most angiosperms and exhibit maternal inheritance, making them well-suited for molecular

systematics and phylogenetic analyses [75]. Phylogenetic analysis revealed that a total of fifteen species could be classified into four clades: Asteraceae, Stylidiaceae, Lamiaceae, and the outgroup *A. thaliana* from the Brassicaceae family. Each branch showed relatively high support, with the four medicinal *Chrysanthemum* cultivars clustered together with other *Chrysanthemum* species and forming a larger clade within the Asteraceae family. They are clearly distinguished from closely related families such as Poaceae in the Poales order and Lamiaceae in the Lamiales order. In this study, 'Boju', 'Huaiju', 'Gongju', and 'Hangbaiju' clustered together. These four medicinal *Chrysanthemum* cultivars exhibit similar morphological characteristics, with minor differences observed only in traits such as plant height and petal size, which is consistent with our clustering results [76]. It is noteworthy that from the phylogenetic tree, medicinal *Chrysanthemum*s cluster together with wild *Chrysanthemum*s and garden *Chrysanthemum*s. This may be attributed to the common practice of using wild *Chrysanthemum* (*C. indicum*) as a parental species in *Chrysanthemum* breeding, achieved through distant hybridization and germplasm aggregation. Additionally, some reports suggest that wild *Chrysanthemum* is widely considered to be the wild ancestor of cultivated *Chrysanthemum*s, which is consistent with the findings of Hao et al. [77].

Conclusion

In this study, we sequenced the complete cp genomes of two cultivars from the *Chrysanthemum* genus and conducted the comparative analysis with published data of two other medicinal *Chrysanthemum* cultivars. The results revealed that all cp genomes exhibited a typical quadripartite structure with lengths ranging from 151,057 to 151,109 bp. A total of 134 genes were identified, including 89 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. The cp genomes of the four *Chrysanthemum* cultivars showed conservation at the molecular level, with sequence variations primarily present in non-coding regions. Through boundary analysis, no significant contraction or expansion was detected in the cp genomes of the four cultivars, with only minor differences in a few nucleotides. High-frequency codons in the four *Chrysanthemum* cultivars tend to end with A/U, while low-frequency codons are more inclined to have G/C as the third base. Additionally, a total of 161 LSRs and 276 SSRs were identified across the four *Chrysanthemum* cultivars, which can be utilized as molecular markers for DNA typing and plant classification. In addition, nine highly variable regions were identified, that are of significant importance for species delineation within the *Chrysanthemum* genus. Furthermore, phylogenetic analysis confirmed the close relationship among the four *Chrysanthemum* cultivars. Currently, the identification of medicinal *Chrysanthemum* cultivars based on morphological characteristics poses certain challenges, and different cultivars may emphasize different pharmacological effects. Therefore, this study not only plays a crucial role in understanding the internal structure and phylogeny of the cp genomes of the four medicinal *Chrysanthemum* cultivars but also contributes to species and cultivar identification and further development and utilization.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-024-05679-0.

Supplementary Material 1.
Supplementary Material 2.
Supplementary Material 3.
Supplementary Material 4.
Supplementary Material 5.
Supplementary Material 6.

Acknowledgements
We would like to thank Genewiz Biotechnology (Suzhou) Co., Ltd in China for chloroplast genome sequencing and bioinformatics analysis. We also thank Dr. Hanif Khan from Anhui University of Traditional Chinese Medicine of China for revising the manuscript.

Authors' contributions
Conceived, designed, and implemented the study: SX, YD, YW; collected specimens and prepared samples for sequencing: YD, YW, WD, CW and SX, Statistics analysis and interpretation: YD, SX, YW, LM; Reagents/materials/analysis tools: NC, YL; Drafted the manuscript: SX, YD, JM and YW; All authors edited the manuscript and approved the final version.

Funding
This work was supported by Key Natural Science Research Projects in Anhui Universities (2022AH050461, 2022AH040328), Open Project of Provincial and Ministerial Scientific Research Platform, Fuyang Normal University (No. FSKFKT010D), Research Funds of Joint Research Center for Chinese Herbal Medicine of Anhui of IHM (yjzx2023001, yjzx2023011), and Scientific Research Team Program of Anhui Colleges and Universities (No. 2022AH010036).

Availability of data and materials
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable. No specific permits were required for the collection of specimens for this study. This research was carried out in compliance with the relevant laws of China.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]College of Pharmacy, Anhui University of Chinese Medicine, Hefei 230012, China. [2]Institute of Traditional Chinese Medicine Resources Protection and Development, Anhui Academy of Chinese Medicine, Hefei 230012, China. [3]Jiuzhou Fangyuan Pharmaceutical Co., Ltd., Anhui Modern Industry

Research Institute of Traditional Chinese Medicine, Bozhou 236821, China. [4]Joint Research Center for Chinese Herbal Medicine of Anhui of IHM, Anhui University of Chinese Medicine, Hefei 230012, China. [5]Joint Research Center for Chinese Herbal Medicine of Anhui of IHM, College of Pharmacy, Bozhou Vocational and Technical College, Bozhou 236800, China. [6]MOE-Anhui Joint Collaborative Innovation Center for Quality Improvement of Anhui Genuine Chinese Medicinal Materials, Hefei 230038, China.

Received: 8 August 2024 Accepted: 7 October 2024
Published online: 22 October 2024

References

1. Zhang LJ, Dai SL. Research advance on germplasm resources of chrysanthemum × morifolium. Chinese Bulletin of Botany. 2009;44(5):526–35.
2. Ma A, Zou F, Zhang R, Zhao X. The effects and underlying mechanisms of medicine and food homologous flowers on the prevention and treatment of related diseases. J Food Biochem. 2022;46(12):14430.
3. Liu YC, Lu CF, Zhou J, Zhou FF, Gui AJ, Chu HL, et al. Chrysanthemum morifolium as a traditional herb: A review of historical development, classification, phytochemistry, pharmacology and application. J Ethnopharmacol. 2024;10:330.
4. Mekapogu M, Kwon O-K, Song H-Y. Jung Jae-A Towards the improvement of ornamental attributes in Chrysanthemum: recent progress in biotechnological advances. Int J Mol Sci. 2022;23(20):12284.
5. Wang ZL, Yuan YX, Hong B, Zhao X, Gu ZY. Characteristic volatile fingerprints of four chrysanthemum teas determined by HS-GC-IMS. Molecules. 2021;26(23):7113.
6. Duan XX, Zhang WJ, Li JJ, Xu HY, Hu J, Zhao L. Comparative metabolomics analysis revealed biomarkers and distinct flavonoid biosynthesis regulation in Chrysanthemum mongolicum and C. rhombifolium. Phytochem Anal. 2021;33(3):373–85.
7. Yuang HW, Jiang S, Liu YK, Daniyal M, Jian YQ, Peng CY, et al. The flower head of Chrysanthemum morifolium Ramat. (Juhua): A paradigm of flowers serving as Chinese dietary herbal medicine. J Ethnopharmacol. 2020;28:261.
8. Shao YH, Sun YD, Li D, Chen YP. Chrysanthemum indicum L.: A comprehensive review of its botany, phytochemistry and pharmacology. Am J Chin Med. 2020;48(04):871–97.
9. Zhao JP, Khan SI, Wang M, Vasquez Y, Yang MH, Avula B, et al. Octulosonic acid derivatives from roman chamomile (Chamaemelum nobile) with activities against inflammation and metabolic disorder. J Nat Prod. 2014;77(3):509–15.
10. Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S. Plant DNA barcoding: from gene to genome. Biol Revi Camb Philos Soc. 2014;90(1):157–66.
11. Chen SL, Yin XM, Han J, Sun W, Yao H, Song JY, et al. DNA barcoding in herbal medicine: Retrospective and prospective. J Pharm Anal. 2023;13(5):431–41.
12. Techen N, Parveen I, Pan Z, Khan IA. DNA barcoding of medicinal plant material for identification. Curr Opin Biotechnol. 2014;25:103–10.
13. Al-qurainy F, Khan S, Tarroum M, Al-Hemaid FM, Ali MA. Molecular authentication of the medicinal herb Ruta graveolens (Rutaceae) and an adulterant using nuclear and chloroplast DNA markers. Genet Mol Res. 2011;10(4):2806–16.
14. Zhang P, Liu CS, Zheng XS, Wu L, Liu ZX, Liao BS, et al. Full-length multibarcoding: DNA barcoding from single ingredient to complex mixtures. Genes. 2019;10(5):343.
15. Teske D, Peters A, Möllers A, Fischer M. Genomic profiling: the strengths and limitations of chloroplast genome-based plant variety authentication. J Agric Food Chem. 2020;68(49):14323–33.
16. Wu L, Wu M, Cui N, Xiang L, Li Y, Li XW, et al. Plant super-barcode: a case study on genome-based identification for closely related species of Fritillaria. Chinese medicine. 2021;16(1):52.
17. Zhang Z, Zhang Y, Song M, Guan Y, Ma X. Species identification of Dracaena using the complete chloroplast genome as a super-barcode. Front Pharmacol. 2019;10:1441.
18. Jiang Y, Zhu CH, Wang ST, Wang FS, Sun ZY. Identification of three cultivated varieties of Scutellaria baicalensis using the complete chloroplast genome as a super-barcode. Sci Rep. 2023;13(1):5602.
19. Chen XL, Zhou J, Cui Y, Wang Y, Duan BZ, Yao H. Identification of Ligularia herbs using the complete chloroplast genome as a super-barcode. Front Pharmacol. 2018;9:695.
20. Wei X, Shen F, Zhang Q, Liu N, Zhang Y, Xu M, et al. Genetic diversity analysis of Chinese plum (Prunus salicina L.) based on whole-genome resequencing. Tree Genet Genomes. 2021;17(3):2.
21. Dong WP, Liu J, Yu J, Wang L, Zhou SL. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PLoS ONE. 2012;7(4):e35071.
22. Zhang Y, Tian L, Lu C. Chloroplast gene expression: Recent advances and perspectives. Plant Communications. 2023;4(5):100611.
23. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17(1):134.
24. Yang J, Yue M, Niu C, Ma XF, Li ZH. Comparative analysis of the complete chloroplast genome of four endangered herbals of notopterygium. Genes. 2017;8(4):124.
25. De Vries J, Archibald JM. Plastid genomes. Curr Biol. 2018;28(8):R336–7.
26. Zhu Y, Zhang X, Yan S, Feng C, Wang D, Yang W, et al. SSR identification and phylogenetic analysis in four plant species based on complete chloroplast genome sequences. Plasmid. 2023;125:102670.
27. Commission CP. Pharmacopoeia of the People's Republic of China [M]. Beijing, China: China Medical Science Press; 2020.
28. Dierkckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2016;45(4):e18.
29. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding preassembled contigs using SSPACE. Bioinformatics. 2011;27(4):578–9.
30. Nadalin F, Vezzi F, Policriti A. GapFiller: a de novo assembly approach to fill the gap within paired reads. BMC Bioinformatics. 2012;13(14):S8.
31. Wyman SK, Jansen RK, Bookre JL. Automatic annotation of organellar genomes with DOGMA. Bioinformatics. 2004;20(17):3252–5.
32. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fishcer A, Bock R, et al. GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45(W1):W6–11.
33. Qu Xj, Zou D, Zhang RY, Stull GW, Yi TS. Progress, challenge and prospect of plant plastome annotation. Front Plant Sci. 2023;14:1166140.
34. Chan PP, LIN BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Res. 2021;49(16):9077–96.
35. 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.
36. Frazer KA, Pachter L, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004;32:W273–9.
37. Amieyousefi A, Hyvönen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics. 2018;34(17):3030–1.
38. 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.
39. Beier S, Thiel T, Münch T, Scholz U, Martin M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33(16):2583–5.
40. Yang MH, Liu JH, Yang WQ, Li Z, Hai YL, Duan BZ, et al. Analysis of codon usage patterns in 48 Aconitum species. BMC Genomics. 2023;24(1):703.
41. Sharp PM, Tuohy TMF, Mosurski KR. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. 1986;14(13):5125–43.
42. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25(11):1451–2.
43. Katoh K, Standley DM. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
44. Capella-gutiérrez S, Silla-marktinez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972–3.
45. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22(21):2688–90.

46. Barba-montoya J, Tao Q, Kumar S. Using a GTR+Γ substitution model for datingsequence divergence when stationarity and time-reversibility assumptions are violated. Bioinformatics. 2020;36(Suppl_2):i884–94.
47. Twyford AD, Ness RW. Strategies for complete plastid genome sequencing. Mol Ecol Resour. 2016;17(5):858–68.
48. Xu YF, Liao BR, Ostevik KL, Zhou HG, Wang FL, Wang BS. The maternal donor of chrysanthemum cultivars revealed by comparative analysis of the chloroplast genome. Front Plant Sci. 2022;13: 923442.
49. Ma YP, Zhao L, Zhang WJ, Zhang YH, Xing X, Duan XX, et al. Origins of cultivars of Chrysanthemum-evidence from the chloroplast genome and nuclear LFY gene. J Syst Evol. 2020;58(6):925–44.
50. Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci. 1987;84(24):9054–8.
51. Wu CS, Chaw SM. Highly rearranged and size-variable chloroplast genomes in conifers II clade (cupressophytes): evolution towards shorter intergenic spacers. Plant Biotechnol J. 2013;12(3):344–53.
52. Muraguri S, Xu W, Charpman M, Muchugi A, Oluwaniyi A, Oyebanji O, et al. Intraspecific variation within Castor bean (Ricinus communis L.) based on chloroplast genomes. Industrial Crops and Products. 2020;155:10.
53. Yang YM, Jiao J, Fan XC, Zhang Y, Jiang JF, Li M, et al. Complete chloroplast genome sequence and characteristics analysis of Vitis ficifolia. Acta Horticulturae Sinica. 2019;46(4):635–48.
54. Parvathy ST, Udayasuriyan V, Bhadana V. Codon usage bias. Mol Biol Rep. 2022;49(1):539–65.
55. Iriarte A, Lamolle G, Musto H. Codon usage bias: An endless tale. J Mol Evol. 2021;89(9):589–93.
56. CBOL Plant Working Group. A DNA barcode for land plants. Proc Natl Acad Sci USA. 2009;106(31):12794–7.
57. Kress WJ, Wurdack KJ, Zimmer EA, Weight LA, Janzen DH. Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci USA. 2005;102(23):8369–74.
58. Kress WJ, Erickson DL. A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. PLoS ONE. 2007;2(6):e508.
59. Yu J, Han Y, Xu H, Han S, Li XP, Niu Y. Structural divergence and phylogenetic relationships of Ajania (Asteraceae) from plastomes and ETS. BMC Genomics. 2023;24(1):602.
60. Liu X, Zhou B, Yang H, et al. Sequencing and analysis of Chrysanthemum carinatum schousb and Kalimeris indica. The complete chloroplast genomes reveal two inversions and rbcL as barcoding of the vegetable. Molecules. 2018;23(6):1358.
61. Huang Y, An YM, Meng SY, Guo YP, Rao GY. Taxonomic status and phylogenetic position of Phaeostigma in the subtribe Artemisiinae (Asteraceae). J Syst Evol 2017;55.
62. Dong W, Xu C, Li C, Sun J, Zuo J, Shi S, et al. ycf1, the most promising plastid DNA barcode of land plants. Sci Rep. 2015;5(1):8348.
63. Kim G-B, Lim CE, Kim J-S, Kim K, Lee J-H, Mun J-H. Comparative chloroplast genome analysis of Artemisia (Asteraceae) in East Asia: insights into evolutionary divergence and phylogenomic implications. BMC Genomics. 2020;21(1):415.
64. Molina-cano JL, Russell JR, Moralejo MA, Escacena JL, Arias G, Powell W. Chloroplast DNA microsatellite analysis supports a polyphyletic origin for barley. Theor Appl Genet. 2005;110(4):613–9.
65. Selkoe KA, Tooner RJ. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol Lett. 2006;9(5):615–29.
66. Pooler M, Ma H. Interspecific hybridizations in ornamental flowering cherries validated by simple sequence repeat analysis. J Am Soc Hortic Sci. 2013;138(3):198–204.
67. Pollegioni P, Woeste KE, Major Á, Mugnozza GS, Malvolti ME. Characterization of Juglans nigra (L.), Juglans regia L. and Juglans x intermedia Carr. by SSR markers: a case study in Italy. Silvae Genetica. 2009;58:68–78.
68. Zhang Y, Wang C, Ma HZ, Dai SL. Assessing the Genetic Diversity of Chrysanthemum Cultivars with Microsatellites. J Am Soc Hortic Sci. 2013;138(6):479–86.
69. Xu S, Teng K, Zhang H, Gao K, Wu JY, Duan LS. Chloroplast genomes of four Carex species: Long repetitive sequences trigger dramatic changes in chloroplast genome structure. Front Plant Sci. 2023;14:1100876.
70. Gichira AW, Li Z, Saina JK, Long ZC, Hu GW, Gituru RW, et al. The complete chloroplast genome sequence of an endemic monotypic genus Hagenia (Rosaceae): structural comparative analysis, gene content and microsatellite detection. Peer J. 2017;5:e2846.
71. Kuang DY, Wu H, Wang YL, Gao LM, Zhang SZ, Lu L. Complete chloroplast genome sequence of Magnolia kwangsiensis (Magnoliaceae): implication for DNA barcoding and population genetics. Genome. 2011;54(8):663–73.
72. Niu ZT, Xue QY, Wang H, Xie XZ, Zhu SY, Liu W, et al. Mutational Biases and GC-biased gene conversion affect GC content in the plastomes of Dendrobium genus. Int J Mol Sci. 2017;18(11):2307.
73. Yang WH, Glover BJ, Rao GY, Rao GY, Yang J. Molecular evidence for multiple polyploidization and lineage recombination in the Chrysanthemum indicum polyploid complex (Asteraceae). New Phytol. 2006;171(4):8758–86.
74. Song AP, SU JS, Wang HB, Zhang ZR, Zhang XT, Van de Peer Y, et al. Analyses of a chromosome-scale genome assembly reveal the origin and evolution of cultivated chrysanthemum. Nat Commun. 2023;14(1):2021.
75. Yan LJ, Wang HL, Huang X, Li YC, Yue YH, Wang ZW, et al. Chloroplast genomes of genus Tilia: comparative genomics and molecular evolution. Front Genet. 2022;13:925726.
76. Wang DQ, Liang YM, Liu SJ. Species evolution of Chinese flos Dendranthematis. China J Chin Materia Med. 1999;24(10):584–637.
77. Hao DC, Song Y, Xiao P, Xiao PG, Zhong Y, Wu PL, et al. The genus chrysanthemum: phylogeny, biodiversity, phytometabolites, and chemodiversity. Front Plant Sci. 2022;13:973197.

Publisher's Note