Medicinal Plant Biology

ARTICLE

Open Access

https://doi.org/10.48130/mpb-0026-0022

Medicinal Plant Biology **2026**, 5: e027

The complete chloroplast genome of *Codonopsis pilosula* and comparative analysis with other *Codonopsis* plants: architecture, evolution, and candidate molecular markers

Xiaoli Tian[1,2], Jishuang Liu[2], Zhenxi Fang[3], Jiadong Hu[1], Wansheng Chen[1,2*] and Qing Li[1,2*]

[1] *Department of Pharmacy, Second Affiliated Hospital of Naval Medical University, Shanghai 200003, China*

[2] *Research and Development Center of Chinese Medicine Resources and Biotechnology, Institute of Chinese Materia Medica, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China*

[3] *Department of Pharmacy, Shanghai General Hospital Jiading Branch, Shanghai 201821, China*

* Correspondence: chenwansheng@smmu.edu.cn (Chen W); qli@smmu.edu.cn (Li Q)

Abstract

Codonopsis pilosula (Franch.) Nannf., one of the botanical origins of the Chinese medicine "Dangshen", belongs to the genus *Codonopsis* (Campanulaceae). Morphological convergence among congeners makes species identification within this genus particularly difficult. The chloroplast genome, with its conserved structure and moderate evolutionary rate, has emerged as a powerful tool for species identification and plant phylogenetic studies. The complete chloroplast genome of *C. pilosula* was sequenced, assembled, and annotated in this study, and a comparative genomic analysis with 10 other *Codonopsis* chloroplast genomes representing nine species was performed. All chloroplast genomes exhibited a typical quadripartite structure, with lengths ranging from 169,321 to 183,616 bp and GC contents between 37.97% and 38.72%. Structural variations were primarily attributed to expansions and contractions at the inverted repeat boundaries and variations in noncoding regions. Repeat analysis revealed a higher frequency of palindromic repeats, whereas simple sequence repeat analysis showed a predominance of mononucleotide repeats, particularly A/T repeats, which may be useful for species identification and evolutionary studies. Phylogenetic analysis supported the monophyly of the sampled *Codonopsis* species relative to outgroups, resolving several well-supported subclades. Evolutionary selection pressure analysis indicated purifying selection as the dominant force, with potential positive selection signals detected in *ycf2* and other genes. Nucleotide diversity analysis identified *ycf1* and *ndhF* as hypervariable regions, suggesting that the combined *ycf1–ndhF* fragment may serve as a candidate molecular marker for species identification. The findings of this study provide critical genomic resources for evolutionary studies and species authentication of this medicinally valuable genus.

Keywords: *Codonopsis pilosula*; Chloroplast genome; Comparative analysis; Phylogenetics; Identification

Citation: Tian X, Liu J, Fang Z, Hu J, Chen W, et al. 2026. The complete chloroplast genome of *Codonopsis pilosula* and comparative analysis with other *Codonopsis* plants: architecture, evolution, and candidate molecular markers. *Medicinal Plant Biology* 5: e027 https://doi.org/10.48130/mpb-0026-0022

Introduction

Codonopsis pilosula (Franch.) Nannf., a commonly used medicinal herb in traditional Chinese medicine, is a perennial plant belonging to the Campanulaceae family and serves as one of the original botanical resources of Codonopsis Radix (Dangshen) in the *Pharmacopoeia of the People's Republic of China*[1]. It is primarily distributed across South Asia, Central Asia and East Asia, with its roots typically harvested in autumn[2]. *C. pilosula* is rich in various bioactive components, including polysaccharides, lobetyolin, atractylenolide III, syringin, and amino acids. These components endow it with multiple pharmacological functions, such as promoting blood circulation, nourishing blood, and promoting fluid production. Clinically, it is used to treat syndromes associated with spleen and lung deficiency, qi deficiency, and ischemia. Additionally, it exhibits therapeutic effects in expelling phlegm to alleviate cough, generating fluids to relieve thirst, and regulating gastrointestinal dysfunction. As a result, *C. pilosula* is frequently used as a substitute for ginseng (*Panax* spp.) in medical applications and is extensively utilized in Chinese herbal health products[3,4]. To date, more than 40 species of *Codonopsis* have been identified[5]. However, accurate species identification within the genus remains challenging because of the high morphological similarity among congeners.

Chloroplasts are a class of organelles found in most plants and algae, typically exhibiting a genome size ranging from 120 to 200 kb and encoding approximately 110–140 genes[6]. Compared with the mitochondrial and nuclear genomes, the chloroplast genome possesses a relatively compact genomic structure and its own transcription/translation systems. Owing to its relatively stable genome organization, highly conserved gene structure, moderate evolutionary rate, and predominant maternal inheritance in most angiosperms, the chloroplast genome has emerged as a powerful tool and is widely used in species identification and phylogenetic studies[7,8]. The majority of plant chloroplast genomes are circular and exhibit a typical quadripartite structure consisting of a large single-copy (LSC) region, a small single-copy (SSC) region, and a pair of inverted repeat (IR) regions[9]. Each of these structural regions plays a distinct role in defining chloroplasts' functionality, facilitating the transcription of external genetic information, and ultimately influencing the overall physiological activities of the chloroplast.

At present, most research on *C. pilosula* is predominantly focused on techniques of extracting its chemical components and exploration of its pharmacological activities. Although the chloroplast genome sequence of *C. pilosula* is available in public databases (GenBank accession number MT982394.1 or NC_060312.1), a comparative analysis of chloroplast genomes among the currently available *Codonopsis* species, which is useful for understanding its

© The Author(s)

https://doi.org/10.48130/mpb-0026-0022

phylogenetic relationships and identifying candidate regions for genetic markers, remains limited. Moreover, the phylogenetic relationship between *C. pilosula* and its subspecies *C. pilosula* subsp. *tangshen* has not been adequately examined at the chloroplast genome level, despite their distinct medicinal applications in traditional Chinese medicine. This study therefore aims to (1) provide an independently assembled and annotated chloroplast genome of *C. pilosula* for in-depth characterization, and (2) perform a comparative genomic analysis of 11 *Codonopsis* chloroplast genomes (representing nine distinct species) to investigate their phylogenetic relationships, identify genomic variations, and determine highly divergent regions suitable for developing molecular markers.

Materials and methods

Plant material, DNA extraction, and sequencing

Fresh leaves of *C. pilosula* were collected from the medicinal botanical garden of Shanghai University of Traditional Chinese Medicine. A voucher specimen (specimen number: SHUTCM-DS202501) was deposited in the herbarium of Shanghai University of Traditional Chinese Medicine. Total genomic DNA was isolated from the leaves using the cetrimonium bromide (CTAB) method[10]. The concentration and purity of the extracted DNA were quantified using a Nanodrop 2000 spectrophotometer (Thermo Scientific), and its integrity was verified by electrophoresis on a 1% agarose gel stained with ethidium bromide. The chloroplast genome of *C. pilosula* was sequenced using the Illumina HiSeq 2500 platform (Biozeron, Shanghai, China) with the total genomic DNA, and 150-bp paired-end reads were generated.

Assembly and annotation of the chloroplast genome

The raw sequencing data were assessed for quality using FastQC (version 0.11.9, www.bioinformatics.babraham.ac.uk/projects/fastqc), followed by trimming and filtering with Trimmomatic v. 0.39 under the default settings[11]. The filtered reads were assembled *de novo* using GetOrganelle v. 1.7.5 with the default parameters (k-mer 21, 45, 65, 85, 105)[12]. The high-quality paired-end reads were then aligned to the assembled chloroplast genome using BWA v. 0.7.17[13], and SAMtools v. 1.10[14] was used to process the alignment results to determine the average read coverage depth. IR boundaries were validated using read mapping in IGV v. 2.19.2 with the assembled genome as a reference, followed by manual inspection of coverage consistency at the LSC–IRb, IRb–SSC, SSC–IRa, and IRa–LSC junctions. The assembled chloroplast genome of *C. pilosula* was annotated using CPGAVAS2[15] (http://47.96.249.172:16019/analyzer/home) with the default parameters and the 2,544-plastome reference dataset. For comparison, the chloroplast genomes of other *Codonopsis* plants retrieved from the National Center for Biotechnology Information (NCBI), with the original references provided in Table 1, were also annotated using the same pipeline. The start and stop codons of protein-coding genes were manually adjusted according to sequence alignment with homologous genes from closely related species available in NCBI to ensure the annotations' accuracy. The physical map of the *C. pilosula* chloroplast genome was generated by the Organellar Genome Draw (OGDRAW) v. 1.3.1[16].

Repeat sequence analysis

Simple sequence repeats (SSRs) in the chloroplast genomes of *Codonopsis* plants were identified using the MISA software (https://webblast.ipk-gatersleben.de/misa/index.php?action=1). The minimum repeat thresholds were defined as follows: 10 for mononucleotides, 4 for dinucleotides, and 3 for trinucleotides, tetranucleotides, pentanucleotides, and hexanucleotides. Long repeat sequences in the chloroplast genomes were quantified using the REPuter software (https://bibiserv.cebitec.uni-bielefeld.de/reputer). This analysis included the identification of palindromic repeats, forward repeats, reverse repeats, and complementary repeats. The parameters for REPuter were configured with a maximum of 5,000 repeats, a minimum of 30 repeats, and a Hamming distance of 3. Heatmap visualization was ultimately performed using TBtools 2.029[17].

Comparative analysis of chloroplast genomes

To investigate the structural conservation and divergence among *Codonopsis* plants, a comparative genomic analysis was conducted using 11 complete chloroplast genomes. Among these, 10 sequences were obtained from the NCBI database, and 1 chloroplast genome was newly assembled in this study. The whole-genome alignment and visualization were performed with mVISTA (https://genome.lbl.gov/vista/mvista/submit.shtml) via the Shuffle-LAGAN model, with the newly assembled chloroplast genome serving as the reference. Structural dynamics at the IR boundaries were examined using CPJSdraw (http://cloud.genepioneer.com:9929/#/tool/alltool/detail/335) with the default parameters.

Table 1. Summary of the 11 *Codonopsis* chloroplast genomes analyzed in this study.

Species	NCBI accession number	Length (bp)	GC (%)	LSC		IRs		SSC		Coding regions		Reference
				Size (bp)	GC (%)	Size (bp)	GC (%)	Size (bp)	GC (%)	Size (bp)	GC (%)	
C. pilosula	PQ541237	170,614	38.14	86,117	36.74	38,425	40.31	7,647	32.13	85,992	38.68	This study
C. pilosula	MT982394.1/NC_060312.1[a]	170,667	38.13	86,434	36.73	38,293	40.30	7,647	32.13	76,092	39.16	Unpublished
Codonopsis bhutanica	MW149120.1/NC_063738.1	169,718	38.38	84,771	37.13	38,621	40.36	7,705	32.29	84,258	38.83	Unpublished
Codonopsis tsinlingensis	MN122102.1/NC_056284.1	170,253	38.34	85,408	37.02	38,334	40.46	8,177	32.24	84,012	38.94	[27]
Codonopsis minima	KY587457.1/NC_036311.1	169,321	38.28	85,506	36.86	37,875	40.50	8,065	32.44	88,746	38.53	[28]
Codonopsis lanceolata	MH251613.1	170,075	38.24	85,241	36.87	38,387	40.37	8,060	32.34	88,659	38.52	Unpublished
Codonopsis lanceolata	MH018574.1	169,447	38.24	85,253	36.88	38,067	40.39	8,060	32.37	83,331	38.67	[29]
Codonopsis javanica	MW149125.1/NC_063743.1	172,874	37.97	87,389	36.76	38,698	39.99	8,089	31.66	69,894	39.23	Unpublished
Codonopsis hongii	MN849357.1/NC_049093.1	169,524	38.24	85,326	36.88	38,143	40.40	7,912	32.03	72,346	39.01	[30]
C. pilosula subsp. *tangshen*	MW415426.1	170,672	38.15	86,108	36.75	38,455	40.31	7,654	32.07	90,030	38.55	[31]
Codonopsis convolvulacea	NC_060685.1	183,616	38.72	96,234	37.59	38,454	40.74	10,474	34.27	75,537	39.41	[32]

[a] In this table, accession numbers separated by a slash (e.g., MT982394.1/NC_060312.1) represent identical sequences confirmed by sequence alignment.

Codon usage bias and sequence polymorphism analysis

Codon usage bias analysis was performed using CPStools[18]. To minimize potential biases, stringent filtering criteria were implemented: (1) Only coding sequences exceeding 300 bp in length were retained, (2) paralogous genes were removed, and (3) only genes containing canonical ATG initiation codons were selected. To investigate hypervariable regions in the chloroplast genomes of *Codonopsis* and establish a scientific identification system, multiple sequence alignment of conserved genes from 11 *Codonopsis* plants was conducted using MAFFT v7.505[19]. The aligned sequences were subsequently refined with Gblocks for quality control and concatenated into a supermatrix. Sliding window analysis was then performed using DnaSP v6[20] to calculate nucleotide diversity (Pi) across the concatenated alignment of conserved chloroplast genes, using a 600-bp window size with incremental steps of 200 bp[21]. For each window, the Pi was calculated, and the midpoint position was determined on the basis of the coordinates in the concatenated alignment.

Molecular evolution in *Codonopsis* chloroplast genomes

To investigate the evolutionary dynamics of chloroplast genomes in *Codonopsis* plants, a comprehensive analysis of the natural selection pressures acting on protein-coding sequences was conducted. A conserved set of orthologous genes was identified across all 11 *Codonopsis* chloroplast genomes, including the newly sequenced genome. For the analysis of selection pressure, the coding sequences of the chloroplast protein-coding genes were first aligned using MAFFT with the default parameters to ensure accurate codon-based alignment. The aligned sequences were then used as input for the yn00 program (implemented in the PAML package) according to Yang and Nielsen's method[22] to estimate nonsynonymous substitution rates (*dN*), synonymous substitution rates (*dS*), and their ratio ($\omega = dN/dS$), with the newly assembled *C. pilosula* chloroplast genome designated as the reference. The ω ratio is a commonly used indicator of selective pressure in protein-coding genes, where $\omega < 1$ generally reflects purifying selection, $\omega = 1$ indicates neutral evolution, and $\omega > 1$ may suggest positive selection[23].

Phylogenetic analysis

To infer the primary interspecific phylogenetic relationships, a maximum likelihood (ML) tree was constructed using the complete chloroplast genome sequences from 11 *Codonopsis* plants, with *Lobelia chinensis* and *Platycodon grandiflorus* as outgroup taxa. The ML approach was selected because it explicitly incorporates an evolutionary model and generally provides a more reliable topology for species-level phylogenetic inference. Sequence alignment was performed with MAFFT, followed by data refinement using trimAl v1.4 to trim ambiguous regions[24]. The ML tree was constructed using IQ-TREE 2[25], with the best-fit nucleotide substitution model automatically selected by the integrated ModelFinder Plus module on the basis of the Bayesian information criterion (BIC). Branch support was assessed using 5,000 ultrafast bootstrap replicates. The resulting ML tree was visualized and annotated using the iTOL online platform[26].

To evaluate the identification potential of candidate molecular markers, the complete *ycf1* and *ndhF* genes were extracted from each of the 11 *Codonopsis* chloroplast genomes. The average lengths of *ycf1* and *ndhF* were 6,794 and 2,246 bp, respectively. For each taxon, the two fragments were manually concatenated in the order *ycf1–ndhF* without a spacer, yielding an average concatenated length of 9,040 bp. Because these sequences were obtained directly from the assembled chloroplast genomes, no polymerase chain reaction (PCR) primers were used; for future experimental validation, primers can be designed from the conserved flanking regions. Three separate neighbor-joining (NJ) trees based on the *ycf1* genes, the *ndhF* genes, and the combined *ycf1–ndhF* fragments were constructed using MEGA v 5.05 with the Kimura 2-parameter (K2P) model with pairwise deletion of gaps. Nodal stability was assessed with 1,000 bootstrap replicates. The NJ analysis was used as a distance-based exploratory method to evaluate the discriminatory power of the candidate markers, whereas the ML analysis was used as the primary model-based method for phylogenetic inference. The resulting NJ topologies were compared with the ML phylogeny to further validate the utility of these hypervariable regions for species identification within *Codonopsis*.

Results

General characteristics of the 11 *Codonopsis* chloroplast genomes

In this study, the complete chloroplast genome of *C. pilosula* was sequenced, assembled and compared with those of 10 other *Codonopsis* plants. After removing the adapter sequences and low-quality reads, approximately 8.08 Gb of clean reads was obtained from the total sequencing data for assembly. The newly assembled complete chloroplast genome of *C. pilosula* was 170,614 bp in length with a GC content of 38.14% (Fig. 1 and Table 1). The assembly was supported by a high average read coverage depth of 7,750× (Supplementary Fig. S1), indicating high confidence in the assembled sequence. The chloroplast genome exhibited a typical quadripartite structure, consisting of two IR regions of 38,425 bp each (40.31% GC content), separated by an SSC region of 7,647 bp (32.13% GC content) and an LSC region of 86,117 bp (36.74% GC content).

All the 11 chloroplast genomes of *Codonopsis* plants exhibited a typical quadripartite structure, consisting of an LSC region (84,771–96,234 bp), an SSC region (7,647–10,474 bp), and a pair of IR regions (37,875–38,698 bp). No large-scale deletions or rearrangements were observed. The chloroplast genomes spanned a length of 169,321 to 183,616 bp, and their GC contents varied between 37.97% and 38.72%. Notably, the GC contents of the IR regions (39.99%–40.74%) were consistently higher than those of the LSC (36.73%–37.59%) and SSC (31.66%–34.27%) regions (Table 1).

Functional annotation of the *Codonopsis* chloroplast genomes

To avoid redundancy, repeated genes were counted only once. As summarized in Table 2, a total of 104 genes were annotated from the assembled chloroplast genome of *C. pilosula*, comprising 74 protein-coding genes, 26 tRNA genes, and 4 rRNA genes. Furthermore, 16 genes were identified to harbor introns, with 2 of them containing two introns and the others containing one intron. All core chloroplast genes were present in the assembled *C. pilosula* chloroplast genome.

Comparative analysis of chloroplast genome annotations across 11 *Codonopsis* plants revealed 103–106 genes per species, comprising 72–76 protein-coding genes, 26–27 tRNA genes, and 4 rRNA

https://doi.org/10.48130/mpb-0026-0022

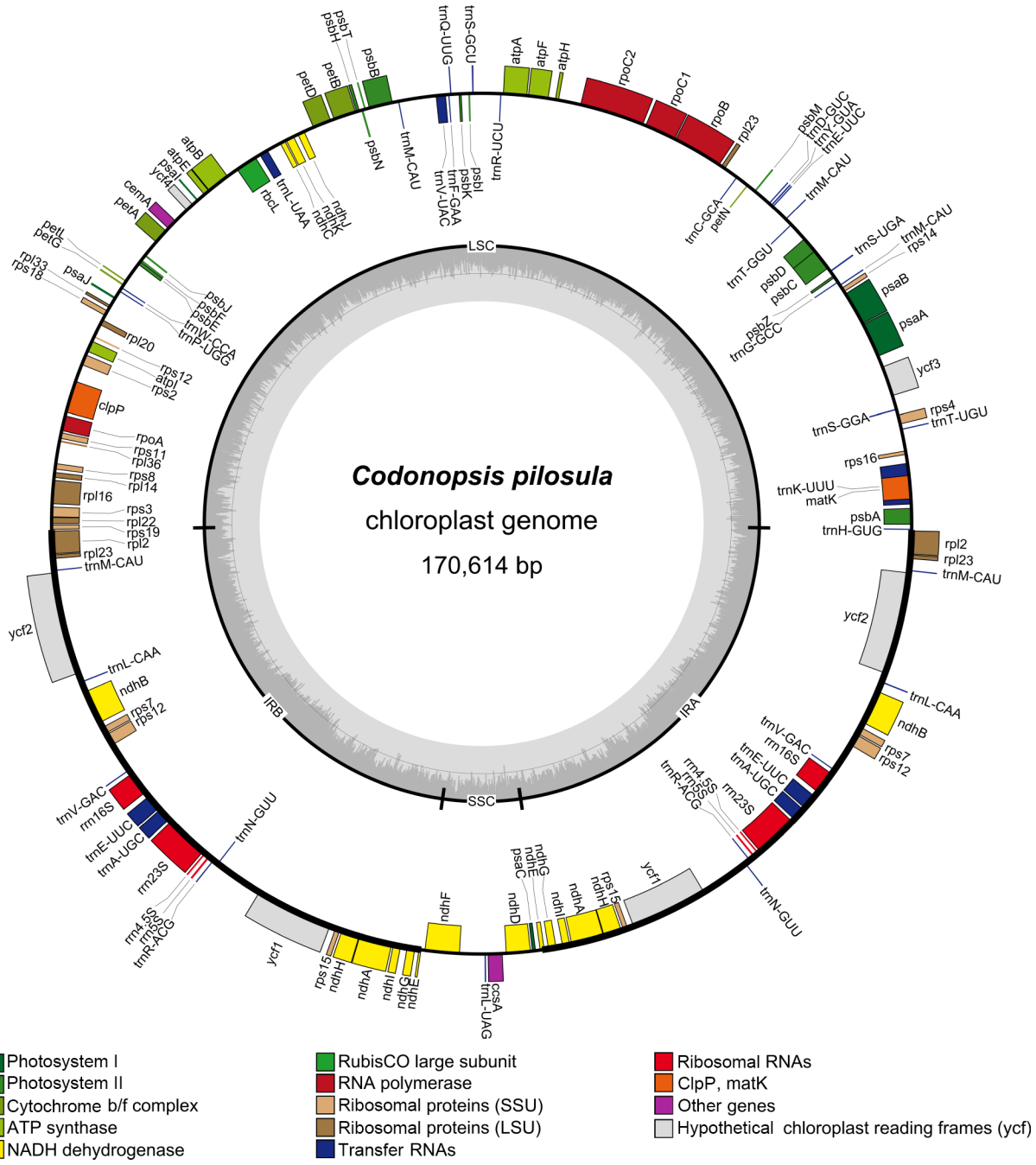


Fig. 1 Chloroplast genome map of *Codonopsis pilosula*. Genes shown outside the circle are transcribed counterclockwise, while those inside are transcribed clockwise. Functional gene categories are color-coded. The inner gray circle represents the GC content.

genes (Supplementary Table S1). The protein-coding genes were classified into four functional categories: Photosynthesis-related genes (42–44 genes), self-replication-related genes (53 or 54 genes), genes of unknown function (4 genes), and other genes (4 genes). Notably, certain genes, including *ndhK*, *rpl23*, and *rpl32*, were selectively annotated in specific species (Supplementary Table S1). Introns' presence/absence variation was also observed among the examined *Codonopsis* chloroplast genomes (Supplementary Table S1).

Codon usage bias

Codon preference analysis was conducted on the shared genes of the 11 *Codonopsis* chloroplast genomes, encompassing 21 amino acids and 64 codons, with 62 codons exhibiting usage bias. Among these, leucine (Leu), isoleucine (Ile), and serine (Ser) were the most abundant amino acids. Notably, analysis of the 11 chloroplast genomes revealed that nine codons displayed relative synonymous codon usage (RSCU) values > 1.5. Ser showed the strongest codon usage bias, with its TCT codon demonstrating the highest RSCU value of 1.91. In contrast, methionine (Met) and tryptophan (Trp) showed no codon preference, maintaining conserved RSCU values of 1.00 (Fig. 2 and Supplementary Table S2). In particular, codons ending with C or G had RSCU values < 1, accounting for 43.75% of the total, suggesting their relative rarity among the chloroplast genes of the 11 species.

Table 2. Annotation of the *C. pilosula* chloroplast genome.

Category of genes	Group of genes	Name of genes
Genes for photosynthesis	Subunits of ATP synthase	*atpA*, *atpB*, *atpE*, *atpF**, *atpH*, *atpI*
	Subunits of Photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbI*, *psbJ*, *psbK*, *psbM*, *psbN*, *psbT*, *psbZ*
	Subunits of NADH-dehydrogenase	*ndhA**(×2), *ndhB**(×2), *ndhC*, *ndhD*, *ndhE*(×2), *ndhF*, *ndhG*(×2), *ndhH*(×2), *ndhI*(×2), *ndhJ*, *ndhK*
	Subunits of cytochrome b/f complex	*petA*, *petB**, *petD**, *petG*, *petL*, *petN*
	Subunits of Photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*
	Subunit of rubisco	*rbcL*
Self-replication	Large subunit of ribosome	*rpl14*, *rpl16**, *rpl2**(×2), *rpl20*, *rpl22*, *rpl23*(×3), *rpl33*, *rpl36*
	DNA dependent RNA polymerase	*rpoA*, *rpoB*, *rpoC1*, *rpoC2*
	Small subunit of ribosome	*rps11*, *rps12*(×2), *rps14*, *rps15*(×2), *rps16*, *rps18*, *rps19*, *rps2*, *rps3*, *rps4*, *rps7*(×2), *rps8*
	Ribosomal RNAs	*rrn16S*(×2), *rrn23S*(×2), *rrn5S*(×2), *rrn4.5S*(×2)
	Transfer RNAs	*trnH*-GUG, *trnK*-UUU*, *trnT*-UGU, *trnS*-GGA, *trnM*-CAU(×5), *trnG*-GCC, *trnS*-UGA, *trnT*-GGU, *trnE*-UUC*(×3), *trnY*-GUA, *trnD*-GUC, *trnC*-GCA, *trnR*-UCU, *trnS*-GCU, *trnQ*-UUG, *trnF*-GAA, *trnV*-UAC*, *trnL*-UAA*, *trnW*-CCA, *trnP*-UGG, *trnL*-CAA(×2), *trnV*-GAC(×2), *trnA*-UGC*(×2), *trnR*-ACG(×2), *trnN*-GUU(×2), *trnL*-UAG
Other genes	C-type cytochrome synthesis gene	*ccsA*
	Envelope membrane protein	*cemA*
	Protease	*clpP***
	Maturase	*matK*
Unknown genes	Conserved open reading frames	*ycf1**(×2), *ycf2**(×2), *ycf3***, *ycf4*

*: Gene contains one intron; **: Gene contains two introns; (×2): The number of repetitive units was 2; (×3): The number of repetitive units was 3; (×5): The number of repetitive units was 5. NADH: reduced nicotinamide adenine dinucleotide.

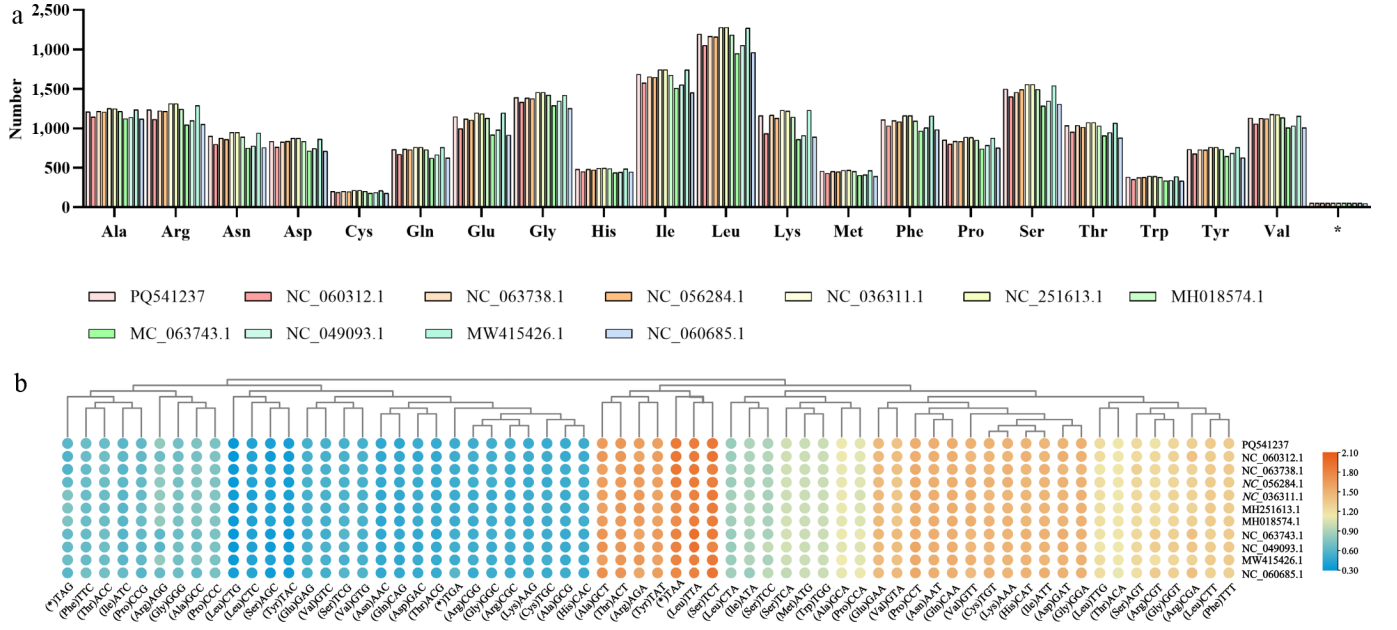


Fig. 2 Codon usage analysis of 11 *Codonopsis* chloroplast genomes. (a) Number of each amino acid. (b) Heatmap of relative synonymous codon usage (RSCU) values across 11 *Codonopsis* genomes. Rows represent species and columns represent codons. The color scale indicates RSCU values, with red representing higher bias (maximum 2.10) and blue representing lower bias (minimum 0.30).

Contraction or expansion of IR regions' boundaries

Through a comparative analysis of IR–SC (single-copy) junctions across 11 *Codonopsis* plants, consistent patterns of gene localization at the boundary regions were identified (Fig. 3). In the newly assembled chloroplast genome of *C. pilosula*, the genes *rps19* and *rpl2* were located at the LSC–IRb junction, with *rps19* positioned 19 bp from the LSC–IRb boundary and *rpl2* was 38 bp away from the same boundary. The *ndhE* genes spanned across the IRb–SSC and SSC–IRa boundaries, with extensions of 10 and 145 bp into the SSC region, respectively. Additionally, *ndhF* and *psaC* were situated 211 and 394 bp away from the IRb–SSC and SSC–IRa boundaries, respectively. Moreover, *rpl2* and *trnH* were positioned on both sides of the IRa–LSC boundary, with distances of 38 and 3 bp from the boundaries, respectively.

In the other 10 chloroplast genomes, the boundaries of the IRs remained stable without significant expansion or contraction. The LSC–IRb junction was consistently located between the genes *rps19* and *rpl2*. Specifically, the IRb region extended 19–169 bp into the *rps19* gene on the LSC side and 37 or 38 bp beyond *rpl2* on the IRb side. Furthermore, the *ndhE* genes were positioned at the IRb–SSC and SSC–IRa boundaries, with extensions of 10–145 and 3–156 bp into the SSC region, respectively. However, in *C. lanceolata* and *C. minima*, *ndhE* was absent at the IRb–SSC boundary and replaced by *ndhG*. Similarly, *ndhE* was replaced by *ndhG* at the SSC–IRa region of *C. bhutanica*. The *ndhF* gene was located 211–276 bp downstream

https://doi.org/10.48130/mpb-0026-0022

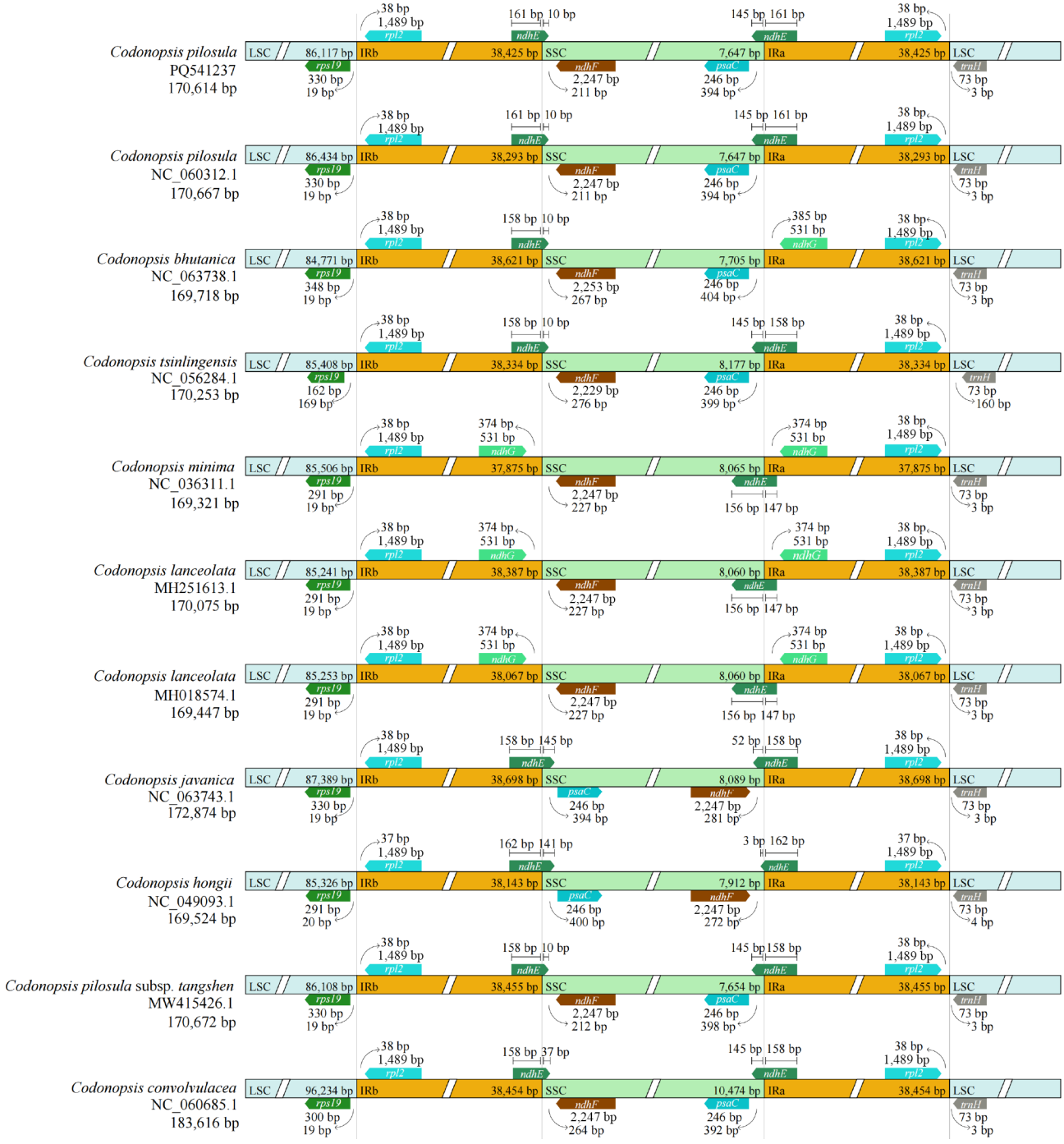


Fig. 3 Comparison of the LSC, SSC, and IR boundary regions across 11 *Codonopsis* chloroplast genomes. Genes near the boundaries are shown with their distances (in bp) from the junctions.

of the IRb–SSC boundary in nine species, whereas *psaC* was consistently positioned 392–404 bp left of the SSC–IRa junction in six species. Conversely, in *C. javanica* and *C. hongii*, the positions of *ndhF* and *psaC* were reversed. Additionally, the IRa–LSC boundaries of the chloroplast genomes in 11 *Codonopsis* plants were situated between *rpl2* and *trnH*, with *rpl2* positioned 37 or 38 bp into the IRa and *trnH* located 3–160 bp into the LSC region. This variation in boundary regions may contribute to the observed differences in chloroplast genomes' sizes. A comparison of the proximity of *rpl2* to the LSC–IRb and IRa–LSC boundaries revealed consistent positioning of *rpl2*, indicating a high degree of conservation (Fig. 3).

SSRs in 11 *Codonopsis* chloroplast genomes

SSRs, known as simple sequence repeats, consist of tandem repeats one to six nucleotides long. These sequences are widely utilized as genetic markers for identifying species. In this study, the newly assembled sequence of *C. pilosula* chloroplast genome contained 400 SSRs (Fig. 4). Among these, 231 were mononucleotide repeats, with A/T repeats being the most prevalent, representing 28.5% and 29.25% of the total SSRs, respectively. Notably, no G/C SSR repeats were observed. Dinucleotide repeats (86, 21.50%), trinucleotide repeats (48, 12.00%), and tetranucleotide repeats (35, 8.75%) were also detected in *C. pilosula* chloroplast

genome, but pentanucleotide and hexanucleotide repeats were absent. The majority of these SSRs were located in the LSC region, constituting 54.00% of the total SSRs.

Comparative analysis of SSRs across 11 *Codonopsis* plants revealed that mononucleotide repeats were the most abundant (209–342 per genome), followed by dinucleotides (43–129 per genome), trinucleotides (32–60 per genome), and tetranucleotides (17–35 per genome). Pentanucleotides and hexanucleotides were observed in *C. convolvulacea*, *C. tsinlingensis*, and *C. bhutanica*. A/T bias persisted in mononucleotides and dinucleotides (Fig. 4a, c and Supplementary Table S3), and SSRs were primarily enriched in LSC regions (46.94%–66.53%), followed by IR regions (27.12%–50.27%) (Fig. 4b and Supplementary Table S4).

Dispersed repeat sequence analysis

Dispersed repeat sequences in the 11 *Codonopsis* chloroplast genomes were analyzed using the online tool REPuter, with a minimum length threshold of 30 bp for four repeat types: Palindromic repeats, forward repeats, reverse repeats, and complementary repeats. In total, 762 repeat sequences were identified in the newly assembled *C. pilosula* chloroplast genome, consisting of 377 palindromic repeats (49.48%), 362 forward repeats (47.51%), 17 reverse repeats (2.23%), and 6 complementary repeats (0.79%), with most ranging between 30 and 39 bp in length (Fig. 5 and Supplementary Table S5). A comparative analysis of dispersed repeats across the 11 *Codonopsis* plants revealed variations in the repeat types and lengths among the different species. Palindromic repeats ranged from 301 to 671, whereas forward repeats ranged from 296 to 624. Distinctively, complementary repeats were absent in *C. minima* and *C. javanica*, but reverse repeats were undetected in *C. hongii* and *C. javanica* (Fig. 5a and Supplementary Table S6).

Comparative chloroplast genome analysis of 11 *Codonopsis* plants

To assess the overall similarity of the chloroplast genomes among *Codonopsis* plants, genetic variants across 11 chloroplast genomes were compared and depicted using mVISTA (Supplementary Fig. S2). Notable sequence divergence was observed in six noncoding regions: *rps16-trnT*-UGU, *psbD-rpl23*, *rpoC2-atpH*, *trnV*-UAC-*trnM*-CAU, *rps2-rpoA*, and *rpl36-rps8*. These regions exhibit varying degrees of diversity and can be utilized for phylogenetic analysis of different species. Moreover, the variability in the IR regions was markedly lower compared with the LSC region, consistent with their evolutionary conservation in angiosperms. Furthermore, in contrast to the reference genome (the newly assembled chloroplast genome in this study), significant distinctions were observed in the coding regions of the *ycf1* and *ycf2* genes.

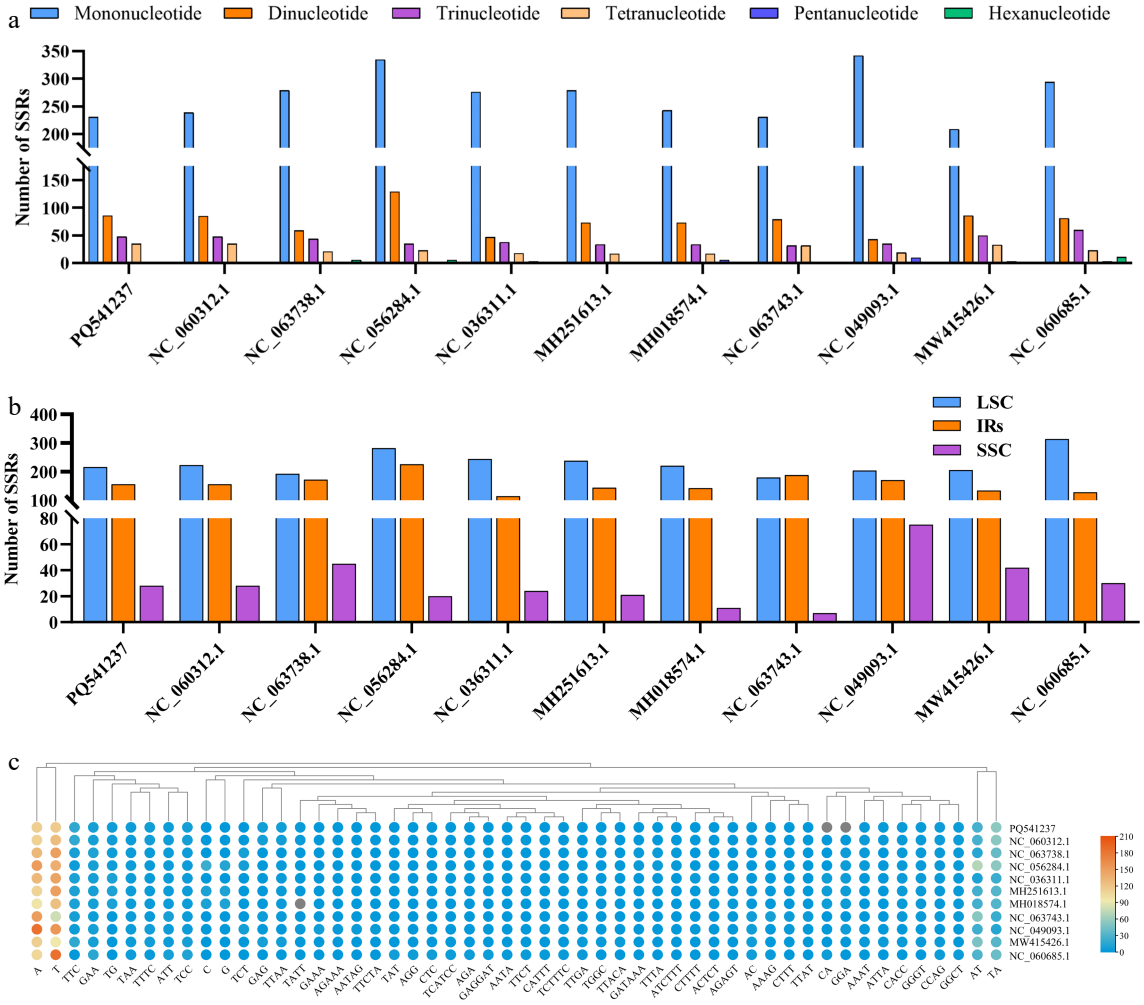


Fig. 4 Simple sequence repeat (SSR) analysis of 11 *Codonopsis* chloroplast genomes. (a) Number of SSRs by repeat type (mono-, di-, tri-, tetra-, penta-, hexanucleotide). (b) Distribution of SSRs across LSC, SSC, and IR regions. (c) The number of SSRs in each repeat type for each species.

https://doi.org/10.48130/mpb-0026-0022

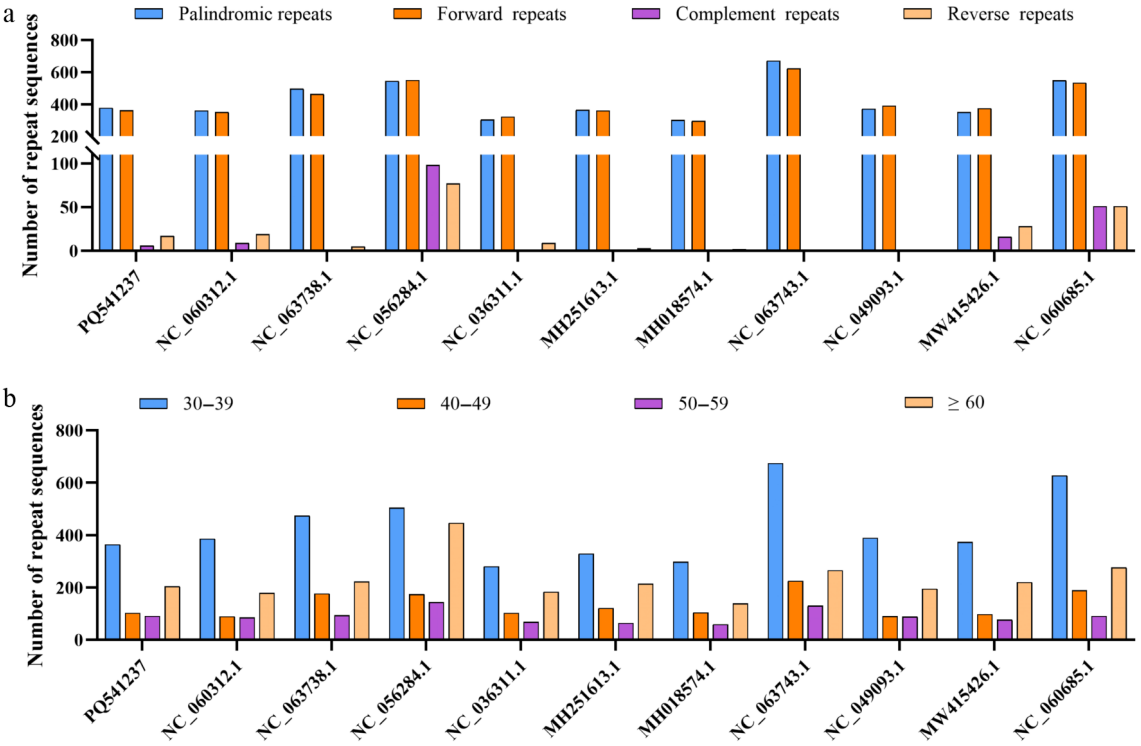


Fig. 5 Dispersed repeat analysis of 11 *Codonopsis* chloroplast genomes. (a). Classification and distribution of palindromic, forward, reverse, and complementary repeats. (b). Frequency distribution of dispersed repeats by length categories (30–39, 40–49, 50–59, and ≥ 60 bp).

Evolutionary selection pressure analysis

An analysis of selection pressure was performed on 71 functionally important chloroplast protein-coding genes. The coding sequences were aligned using MAFFT with the default parameters to ensure accurate alignment. The aligned sequences were then used as input for the yn00 program to estimate *dN*, *dS*, and ω = *dN*/*dS*, with the newly assembled *C. pilosula* chloroplast genome as the reference. The complete dataset for all 71 genes is provided in Supplementary Table S7. Among the genes, *rpl2* exhibited the highest nonsynonymous substitution rate (*dN* = 0.3681) and the most elevated synonymous substitution rate (*dS* = 2.729). This analysis demonstrates the predominance of synonymous substitutions over nonsynonymous substitutions across most genes, resulting in generally low ω values that are characteristic of purifying selection. However, potential positive selection signals ($\omega > 1$) were detected in several genes: *ycf1* and *ycf2* showed elevated ω values across multiple species. These findings indicate that genes with signals of potential positive selection include both functionally uncharacterized genes and functionally characterized genes such as *atpH* and *clpP*. The remaining genes showed consistently low ω values, suggesting they are predominantly under purifying selection. Together, these findings provide insights into the molecular evolution of chloroplast genomes in *Codonopsis* plants.

Nucleotide diversity analysis and validation of hypervariable regions

A comparative analysis of the polymorphisms across 11 *Codonopsis* chloroplast genomes revealed 71 evolutionarily conserved genes. The *ycf1* gene (6,944 bp), encoding a putative protein of unknown function, exhibited the highest nucleotide diversity (Pi = 0.06855), showing higher variability than *ndhF* (Pi = 0.03909, the second highest nucleotide diversity) (Supplementary Tables S8 and S9). These hypervariable regions showed approximately fivefold greater polymorphism than the average Pi value of conserved genes, suggesting their utility as potential phylogenetic markers for species discrimination within *Codonopsis*.

To validate the identification potential of these candidate regions, the hypervariable regions (*ycf1* and *ndhF*) were extracted from the 11 *Codonopsis* chloroplast genomes and combined to generate the *ycf1–ndhF* fragment. Three NJ trees were built from the *ycf1*, *ndhF*, and *ycf1–ndhF* combined fragments (Supplementary Fig. S3). The three phylogenetic trees exhibited similar topological structures. The single-gene tree of *ycf1* showed high bootstrap support for most clades and successfully resolved all sampled species into distinct monophyletic branches. The *ndhF* tree showed relatively low support at several nodes, indicating weaker identification potential than *ycf1*. The combined *ycf1–ndhF* fragment greatly increased the bootstrap values and achieved better phylogenetic resolution than each single locus. Across all trees, *C. convolvulacea*, *C. javanica*, and *C. hongii* formed well-separated independent clades. *C. bhutanica* and *C. tsinlingensis* were recovered as sister species with very close phylogenetic affinity, and *C. minima* was closely related to *C. lanceolata*. All accessions of *C. pilosula* (including the subspecies *C. pilosula* subsp. *tangshen*) constituted a monophyletic lineage.

Phylogenetic analysis of *Codonopsis* plants and their relatives

To resolve the phylogenetic position and evolutionary relationships of *C. pilosula* within the genus *Codonopsis*, a ML phylogenetic tree was reconstructed using all chloroplast genomes under the TVM + F + I + R8 nucleotide substitution model with ultrafast bootstrap support (5,000 replicates) (Fig. 6). The resulting phylogeny supported the monophyly of the sampled *Codonopsis* species relative to the outgroup taxa (*L. chinensis* and *P. grandiflorus*). Within the

core *Codonopsis* clade, the tree resolved several well-supported subclades (Fig. 6). *C. javanica* and *C. hongii* formed an early-diverging lineage, whereas the remaining *Codonopsis* species were further divided into additional subclades. Interestingly, *C. pilosula* and its subspecies *C. pilosula* subsp. *tangshen* formed separate branches. The remaining taxa formed a topology similar to the NJ trees constructed from *ycf1*, *ndhF* and the combined *ycf1–ndhF* fragment. *C. pilosula* formed a distinct subclade, *C. bhutanica* was sister to *C. tsinlingensis*, *C. minima* clustered with *C. lanceolata*, and the two accessions of *C. lanceolata* grouped tightly.

Discussion

C. pilosula is valued both as a traditional medicine and as an edible herb, and its market demand has steadily increased in recent years[33]. However, current research on this species has predominantly focused on its bioactive components, although its genetic resources remain underexplored. The chloroplast genome, owing to its conserved structure and moderate evolutionary rate, serves as an effective tool for investigating plants' phylogeny and genetic diversity[34]. In this study, we conducted a comparative analysis of chloroplast genomes across 11 *Codonopsis* plants, integrating a newly assembled genome of *C. pilosula* with publicly available data to elucidate their genomic architecture, phylogenetic relationships, and evolutionary dynamics. To ensure comparability, all genomes were uniformly reannotated using the same pipeline (CPGAVAS2), minimizing potential biases arising from inconsistent annotation standards.

Consistent with findings in other angiosperms, all 11 *Codonopsis* chloroplast genomes exhibited a typical quadripartite structure. Although the overall structural organization was conserved, the lengths and sequence composition of individual regions varied among species, with the LSC ranging from 84,771 to 96,234 bp, the SSC from 7,647 to 10,474 bp, and the IRs from 37,875 to 38,698 bp. The total GC contents across *Codonopsis* plants ranged from 37.97% to 38.72%, falling within the characteristic 30%–40% GC content range observed in angiosperm chloroplast genomes. Notably, the IR regions exhibited consistently higher GC contents (39.99%–40.74%) compared with the LSC (36.73%–37.59%) and SSC (31.66%–34.27%) regions, a phenomenon likely attributable to the presence of rRNA genes with an inherently high GC content in the IR regions[23]. These findings indicate that the overall quadripartite structure is conserved across the 11 *Codonopsis* plants, whereas the lengths and sequence composition of individual regions vary among species.

This variation in genome size (169,321–183,616 bp) was primarily attributed to length of the polymorphisms in the LSC region, particularly in the intergenic spacers, rather than in the coding regions. A key source of structural divergence was the expansion and contraction of the IR regions. Our analysis revealed that the LSC–IRb and IRa–LSC boundaries were highly conserved across most species. At the LSC–IRb boundary, *rps19* was consistently located without extending into the IR region. At the IRa–LSC boundary, *rpl2* and *trnH* were positioned on opposite sides. This configuration was nearly identical across all 11 species, underscoring the evolutionary constraint at these junctions. However, the distance between *trnH* and the IRa–LSC border varied from 3 bp in most species to 160 bp in *C. tsinlingensis*, indicating that shifts of the IR boundary represent an ongoing, albeit slow, evolutionary process within the genus. In contrast, the IRb–SSC boundary exhibited notable lineage-specific variations. In most *Codonopsis* species, the *ndhE* gene spanned this boundary. However, in *C. lanceolata* and *C. minima*, the *ndhE* copy was lost at this boundary and replaced by *ndhG* (Fig. 3). Consistent with this structural rearrangement, a reduction in the *ndhE* copy number was also observed. Two copies of *ndhE* are present in most species, whereas only a single copy was found in *C. lanceolata*, *C. minima*, and *C. bhutanica*. This lineage-specific loss of *ndhE* is particularly noteworthy from an evolutionary perspective and may reflect a functional redundancy, potentially related to their specific ecological adaptations. The pattern of gene loss and rearrangement suggests a shared evolutionary event within this clade, a hypothesis supported by the phylogenetic analysis (Fig. 6). Further structural polymorphism was observed in the early-diverging species *C. javanica* and *C. hongii*, where *ndhF* and *psaC* exhibited an inverted orientation at the IR–SSC junction relative to the typical configuration found in other *Codonopsis* species. Such rearrangements, though rare, contribute to diversity in the genome size and may influence gene expression by relocating regulatory elements near the borders. Similar IR boundary dynamics have been documented in other angiosperm lineages, including Orchidaceae, where they were associated with contraction of the SSC region and adaptive evolution[35].

The analysis of SSRs and long repeats provided additional insights into the genomes' plasticity and potential marker development. The predominance of A/T mononucleotide repeats across all species aligns with the overall AT-rich nature of chloroplast genomes and the observed base composition bias in the LSC region. The presence of species-specific repeat types, such as pentanucleotide and hexanucleotide repeats in *C. convolvulacea*, *C. tsinlingensis*, and *C. bhutanica*, suggests their potential utility as molecular markers.

Codon usage bias reflects evolutionary constraints on gene expression and has been widely documented in plant chloroplast genomes[36]. The RSCU value is a fundamental metric for measuring this bias, where values above 1 signify significant codon favoritism[37]. Analysis of the 11 *Codonopsis* chloroplast genomes revealed that Leu, Ser, and Ile were the most abundantly encoded

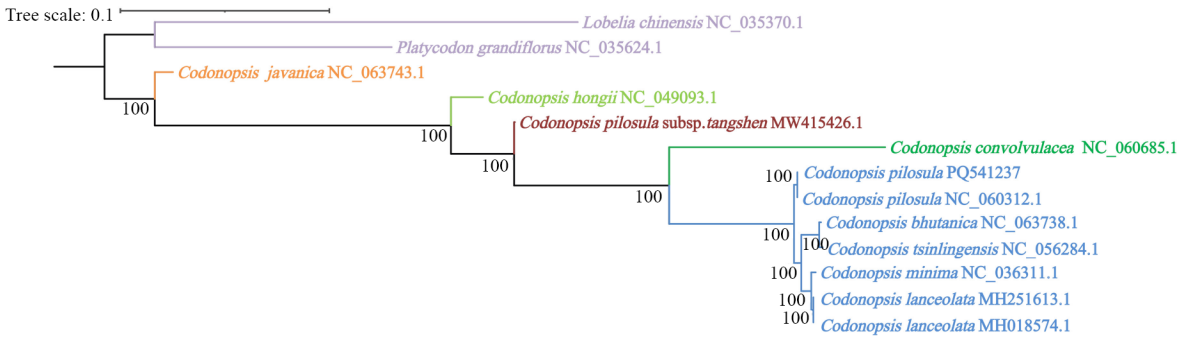


Fig. 6 The ML phylogenetic tree of 11 *Codonopsis* plants and allied plants. The tree was constructed using IQ-TREE 2 under the TVM + F + I + R8 model with 5,000 bootstrap replicates. Bootstrap support values (≥ 70%) are shown at the nodes.

https://doi.org/10.48130/mpb-0026-0022

amino acids. Among the 30 high-frequency codons (RSCU > 1), 13 terminated with A, 16 with T, and only 1 with G, reinforcing a pronounced A/T bias at the third codon position. This pattern is consistent with the findings of Yan et al.[38], who reported that 96.55% of high-frequency codons in *Codonopsis* chloroplast genomes ended with A/T, and aligns with observations in other medicinal plant chloroplast genomes, such as *Aconitum* and *Delphinium grandiflorum*[39,40], reflecting the overall AT-rich nature of the *Codonopsis* chloroplast genomes.

Identifying hypervariable regions with high nucleotide diversity is essential for developing molecular markers to differentiate closely related *Codonopsis* species. Nucleotide diversity analysis revealed that *ycf1* exhibited the highest nucleotide diversity among all examined regions. Previous research has indicated that *ycf1* demonstrates more variability than the *matK* locus in Orchidaceae, making it a promising plastid molecular marker for phylogeny at lower taxonomic levels[35]. Recent functional investigations have shown that *ycf1* encodes Tic214, a core component of the translocon complex at the inner envelope membrane of chloroplasts, and plays a crucial role in regulating the synthesis of chloroplast-encoded photosynthetic complex subunits[41]. This functional importance, coupled with its high sequence variability, implies that *ycf1* may serve as a candidate molecular marker for species identification in *Codonopsis*. Given the reported role of *ycf1*/Tic214 in chloroplast protein import and photosynthesis-related processes, future studies could further investigate whether sequence variation in *ycf1* has functional or adaptive significance. Additionally, the combined *ycf1–ndhF* region has also proven to be effective for species identification in *Prunus persica*[42]. In this study, the NJ trees constructed from the combined *ycf1–ndhF* fragment showed improved bootstrap support for most clades compared with the single-locus *ycf1* or *ndhF* trees (Supplementary Fig. S3). These findings suggest that *ycf1*, particularly when combined with *ndhF*, has the potential to serve as a candidate molecular marker for species identification in *Codonopsis*.

Investigating the evolutionary history of species requires an analysis of their adaptive evolution[43]. The ω ratio is a common indicator for evaluating the influence of selective pressures on protein-coding genes, and it is widely utilized to gauge species' adaptive evolutionary rates and the effects of positive selection. Most genes undergo negative selection, indicating a trend towards conservation in evolution[44]. Within the chloroplast genomes of *Codonopsis*, *ycf2* consistently showed elevated ω values across multiple species, suggesting a potential signal of positive selection. Elevated ω values for *ycf2* have also been documented in *Berberis*[45], although the functional basis remains unclear. The recurrent observation of elevated ω values for *ycf2* across diverse plant groups suggests that this gene may warrant further functional characterization.

Phylogenetic analysis based on the complete chloroplast genome sequences resolved the interspecific relationships within the sampled *Codonopsis* species and supported their monophyly relative to the outgroups (Fig. 6). Notably, the overall phylogenetic framework recovered by complete chloroplast genomes was similar to the topology derived from the *ycf1* genes, the *ndhF* genes, and the combined *ycf1–ndhF* fragments (Supplementary Fig. S3). These consistent phylogenetic patterns further confirm that the chloroplast genes *ycf1*, *ndhF* and their combined fragment may be suitable molecular markers for species identification within *Codonopsis*.

An intriguing finding of the phylogenetic analysis is the marked divergence between *C. pilosula* (PQ541237 and MT982394.1) and *C. pilosula* subsp. *tangshen* (MW415426.1). Despite their close taxonomic relationship, the two taxa were placed in distinct branches within the phylogenetic tree, suggesting that significant genetic differentiation has occurred between the typical variety and its subspecies. Notably, *C. pilosula* subsp. *tangshen* is recorded as *C. tangshen* in the *Chinese Pharmacopoeia*[1], where it is treated as an independent species rather than a subspecies, a treatment that is consistent with the phylogenetic divergence observed in this study.

The chloroplast genome-based ML phylogenetic tree also provides insights into the taxonomic status of two species with historical nomenclatural changes. *Codonopsis hongii* (originally described as *Leptocodon hirsutus*) constitutes an early-diverging lineage nested within the *Codonopsis* clade. This finding lends support to the taxonomic treatment transferring the former *L. hirsutus* into *Codonopsis*. By contrast, *C. convolvulacea* (recently renamed *Pseudocodon convolvulaceus* following taxonomic revision) forms a distinctly divergent sublineage in the phylogenetic tree. This deep genetic divergence supports its segregation into the newly recognized genus *Pseudocodon*. Nevertheless, both *C. hongii* and *C. convolvulacea* remain embedded within the broader *Codonopsis* clade in the plastid tree, implying that *Codonopsis* sensu lato may be paraphyletic.

Conclusions

This study presents a comparison of the chloroplast genomes of *Codonopsis*, encompassing 11 genomes (representing nine distinct species) and providing a newly assembled reference genome for *C. pilosula*. The analyses reveal both structural conservation and dynamics in the chloroplast genomes' architecture, identify the *ycf1–ndhF* combined region as a candidate molecular marker for species identification, and provide insights into the phylogenetic relationships among the sampled species, including the early divergence of *C. javanica* and *C. hongii* and the notable differentiation between *C. pilosula* and its subspecies *C. pilosula* subsp. *tangshen*. Collectively, these findings contribute to the growing genomic resources for *Codonopsis* and provide a foundation for future molecular marker development and phylogenetic studies with expanded sampling.

Author contributions

The authors confirm their contributions to the paper as follows: study conception and design: Li Q, Chen W; data collection: Liu J, Fang Z, Hu J; analysis and interpretation of results: Tian X; draft manuscript preparation: Tian X, Li Q. All authors reviewed the results and approved the final version of the manuscript.

Data availability

The raw Illumina paired-end sequencing reads generated in this study are available in the NCBI Sequence Read Archive (SRA) (www.ncbi.nlm.nih.gov/sra) under the accession number SRP737974 (www.ncbi.nlm.nih.gov/sra/?term=SRP737974). The complete chloroplast genome of *C. pilosula* has been deposited in GenBank (www.ncbi.nlm.nih.gov/genbank) with the accession number PQ541237 (www.ncbi.nlm.nih.gov/nuccore/PQ541237).

Acknowledgments

This work was financially supported by the National Key Research and Development Program of China (2022YFC3501700), the National Natural Science Foundation of China (32070327, 31770329 and 32400301), and China Postdoctoral Science Foundation (2023M744290).

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary information accompanies this paper online at: https://doi.org/10.48130/mpb-0026-0022.

Dates

Received 16 August 2026; Revised 25 September 2026; Accepted 19 October 2026; Published online 10 November 2026

References

[1] Chinese Pharmacopoeia Commission. 2020. *Pharmacopoeia of the People's Republic of China*. Beijing: China Medical Science Press. Volume I. https://2025.chp.org.cn/#/main

[2] Dar AA, Abrol V, Singh N, Gashash EA, Dar SA. 2023. Recent bioanalytical methods for the isolation of bioactive natural products from genus *Codonopsis*. *Phytochemical Analysis* 34:491–506

[3] Luan F, Ji Y, Peng L, Liu Q, Cao H, et al. 2021. Extraction, purification, structural characteristics and biological properties of the polysaccharides from *Codonopsis pilosula*: a review. *Carbohydrate Polymers* 261:117863

[4] Cao L, Du C, Zhai X, Li J, Meng J, et al. 2022. *Codonopsis pilosula* polysaccharide improved spleen deficiency in mice by modulating gut microbiota and energy related metabolisms. *Frontiers in Pharmacology* 13:862763

[5] Choi HB, Shim S, Wang MH, Choi YE. 2023. De novo transcriptome sequencing of *Codonopsis lanceolata* for identification of triterpene synthase and triterpene acetyltransferase. *International Journal of Molecular Sciences* 24:5769

[6] Zhang Y, Tian L, Lu C. 2023. Chloroplast gene expression: recent advances and perspectives. *Plant Communications* 4:100611

[7] Dobrogojski J, Adamiec M, Luciński R. 2020. The chloroplast genome: a review. *Acta Physiologiae Plantarum* 42:98

[8] Chung KP, Gonzalez-Duran E, Ruf S, Endries P, Bock R. 2023. Control of plastid inheritance by environmental and genetic factors. *Nature Plants* 9:68–80

[9] Li L, Hu Y, He M, Zhang B, Wu W, et al. 2021. Comparative chloroplast genomes: insights into the evolution of the chloroplast genome of *Camellia sinensis* and the phylogeny of *Camellia*. *BMC Genomics* 22:138

[10] Mavrodiev EV, Dervinis C, Whitten WM, Gitzendanner MA, Kirst M, et al. 2021. A new, simple, highly scalable, and efficient protocol for genomic DNA extraction from diverse plant taxa. *Applications in Plant Sciences* 9:e11413

[11] Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120

[12] Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, et al. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* 21:241

[13] Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760

[14] Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008

[15] Shi LC, Chen HM, Jiang M, Wang LQ, Wu X, et al. 2019. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Research* 47:W65–W73

[16] Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* 47:W59–W64

[17] Chen C, Wu Y, Li J, Wang X, Zeng Z, et al. 2023. TBtools-II: a "one for all, all for one" bioinformatics platform for biological big-data mining. *Molecular Plant* 16:1733–1742

[18] Huang L, Yu H, Wang Z, Xu W. 2024. CPStools: a package for analyzing chloroplast genome sequences. *iMetaOmics* 1:e25

[19] Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* 20:1160–1166

[20] Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, et al. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution* 34:3299–3302

[21] Chen C, Miao Y, Luo D, Li J, Wang Z, et al. 2022. Sequence characteristics and phylogenetic analysis of the *Artemisia argyi* chloroplast genome. *Frontiers in Plant Science* 13:906725

[22] Wang Y, Wei Q, Xue T, He S, Fang J, et al. 2024. Comparative and phylogenetic analysis of the complete chloroplast genomes of 10 *Artemisia selengensis* resources based on high-throughput sequencing. *BMC Genomics* 25:561

[23] Li E, Liu K, Deng R, Gao Y, Liu X, et al. 2023. Insights into the phylogeny and chloroplast genome evolution of *Eriocaulon* (Eriocaulaceae). *BMC Plant Biology* 23:32

[24] Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973

[25] Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, et al. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37:1530–1534

[26] Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* 49:W293–W296

[27] Zhou H, She R, Zhao P, Zhang S. 2019. The complete chloroplast genome of *Codonopsis tsinglingensis* (Campanulaceae), an endemic Chinese medicine species in Qinling mountains. *Mitochondrial DNA Part B, Resources* 4:3498–3499

[28] Cheon KS, Kim KA, Han JS, Yoo KO. 2017. The complete chloroplast genome sequence of *Codonopsis minima* (Campanulaceae), an endemic to Korea. *Conservation Genetics Resources* 9:541–543

[29] Lee J, Kim SY, Lim JS, Sook Chun H, Kwon K, et al. 2018. The complete chloroplast genome of *Codonopsis lanceolata* (Campanulaceae). *Mitochondrial DNA Part B, Resources* 3:1075–1076

[30] Li CJ, Wang RN, Li DZ. 2020. Comparative analysis of plastid genomes within the Campanulaceae and phylogenetic implications. *PLoS One* 15:e0233167

[31] Yue J, Ni Y, Jiang M, Chen H, Chen P, et al. 2022. Characterization of *Codonopsis pilosula* subsp. *tangshen* plastome and comparative analysis of *Codonopsis* species. *PLoS One* 17:e0271813

[32] Liu L, Li Q, Du J, Yuan W, Sun R, et al. 2025. Complete chloroplast genome sequencing of *Pseudocodon convolvulaceus*, a medicinal herb from Qinghai-Tibet Plateau in China. *PLoS One* 20:e0328307

[33] Mao RJ, Zhang GZ, Xu Y, Deng X, Yan Y, et al. 2024. Effects of Streptomyces pactum Act12 on root morphogenesis and medicinal quality of *Codonopsis pilosula*. *The Journal of Applied Ecology* 35:1608–1614 (in Chinese)

[34] Kassem MA. 2025. Comparative analysis of chloroplast genomes across 20 plant species reveals evolutionary patterns in gene content, codon usage, and genome structure. *International Journal of Plant Biology* 16:105

[35] Neubig KM, Whitten WM, Carlsward BS, Blanco MA, Endara L, et al. 2009. Phylogenetic utility of *ycf*1 in orchids: a plastid gene more variable than *mat*K. *Plant Systematics and Evolution* 277:75–84

[36] Tian T, Zhang Y, Du W, Wang Z. 2025. Comparative genomics of *Triticum*, *Secale*, and *Triticale*: codon usage bias in chloroplast genomes and its implications for evolution and genetic engineering. *International Journal of Molecular Sciences* 26:10266

[37] Fan Y, Guo D, Zhao S, Wei Q, Li Y, et al. 2022. Human genes with relative synonymous codon usage analogous to that of polyomaviruses are involved in the mechanism of polyomavirus nephropathy. *Frontiers in Cellular and Infection Microbiology* 12:992201

[38] Yan F, Wu S, Zhou Y, Wei S. 2026. Natural selection-driven A/U-ending preference in codon usage patterns of *Codonopsis* chloroplast genomes. *Genetic Resources and Crop Evolution* 73:49

https://doi.org/10.48130/mpb-0026-0022

[39] Xia C, Wang M, Guan Y, Li J. 2022. Comparative analysis of the chloroplast genome for *Aconitum* species: genome structure and phylogenetic relationships. *Frontiers in Genetics* 13:878182

[40] Duan H, Zhang Q, Wang C, Li F, Tian F, et al. 2021. Analysis of codon usage patterns of the chloroplast genome in *Delphinium grandiflorum* L. reveals a preference for AT-ending codons as a result of major selection constraints. *PeerJ* 9:e10787

[41] Wang Y, Gao LL, Yuan YJ, Scharff LB, Armarego-Marriott T, et al. 2025. Overexpression of protein import machinery subunit *ycf1*/*Tic214* represses photosynthetic complex biogenesis by inhibiting the synthesis of plastid-encoded core subunits in tobacco. *Journal of Experimental Botany* 76:5982−6004

[42] Amar MH. 2020. *ycf*1-*ndh*F genes, the most promising plastid genomic barcode, sheds light on phylogeny at low taxonomic levels in *Prunus persica*. *Journal of Genetic Engineering and Biotechnology* 18:42

[43] Hong Z, Wu Z, Zhao K, Yang Z, Zhang N, et al. 2020. Comparative analyses of five complete chloroplast genomes from the genus *Pterocarpus* (Fabaceae). *International Journal of Molecular Sciences* 21:3758

[44] Yan M, Dong S, Gong Q, Xu Q, Ge Y. 2023. Comparative chloroplast genome analysis of four *Polygonatum* species insights into DNA barcoding, evolution, and phylogeny. *Scientific Reports* 13:16495

[45] Li H, Abdullah, Yang H, Guo H, Yuan Y, et al. 2025. Chloroplast genome evolution of *Berberis* (Berberidaceae): implications for phylogeny and metabarcoding. *Gene* 933:148959

Copyright: © 2026 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit https://creativecommons.org/licenses/by/4.0/.