The First Chloroplast Genome of Rare Japanese Peony *Paeonia suffruticosa* 'Cun Song Ying': Comparative Genomics and Maternal Origin Insight

Wenzhen Cheng
Heze University
Conghao Hong
Beijing Forestry University
Mingyu Li
Shanghai Construction Management Vocational College
Changyong Gao
gaochangyong@hezeu.edu.cn

Heze University

Research Article

Keywords: Paeonia suffruticosa 'Cun Song Ying', chloroplast genome, phylogeny, codon usage

Posted Date: May 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9340796/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Genetic Resources and Crop Evolution on July 11th, 2026. See the published version at https://doi.org/10.1007/s10722-026-02862-0.

Abstract

• *Paeonia suffruticosa* ‘Cun Song Ying’ is a rare Japanese horticultural variety with significant ornamental value, but its genomic characteristics and phylogenetic status remain unreported. This study reports the first complete chloroplast (cp) genome of *P. suffruticosa* ‘Cun Song Ying’ to fill the gap in its genomic and evolutionary research, aiming to provide a theoretical basis for peony cultivar identification, origin research, and breeding protection.

• High-throughput sequencing combined with bioinformatics tools was used for genome assembly, annotation, and comparative analyses with 15 related Paeoniaceae species. The cp genome of ‘Cun Song Ying’ was systematically characterized in terms of structural features, repetitive sequences, codon usage bias, and phylogenetic relationships.

• The 152,704 bp cp genome has a typical quadripartite structure (LSC:84,365 bp; SSC:17,047 bp; IR:25,646 bp×2), encodes 138 genes, with 6 hypervariable regions, 43 tandem repeats, 73 SSRs, A/U-preferred codons, and is sister to *P. suffruticosa* ‘Luo Yang Hong’ (BS = 100, PP = 1.00).

• This study enriches the chloroplast genomic resources of Paeoniaceae, clarifies the evolutionary characteristics and phylogenetic position of ‘Cun Song Ying’, and provides important genomic data support for peony cultivar identification, origin tracing, and genetic breeding.

1 Introduction

Paeonia suffruticosa ‘Cun Song Ying’ is an important cultivated variety of the genus *Paeonia* in the *Paeoniaceae* family. It is regarded as a rare Japanese cultivated variety among peony horticultural varieties due to its strong growth vigor, high flowering rate and unique ornamental characteristics of multilayered pinkish-white petals. However, despite its horticultural value has a tracted much attention, there is still a lack of systematic research on the origin, genetic background and phylogenetic status of the *Paeonia suffruticosa* ‘Cun Song Ying’ within the Paeonia genus. The chloroplast genomic characteristics and evolutionary relationship of it remain blank.

Chloroplasts are key organelles in plant cells responsible for photosynthesis and various metabolic pathways. Their origin can be traced back to an ancient endosymbiotic event with cyanobacteria approximately 1.5 billion years ago [1, 2]. During the evolution, primitive cyanobacteria established a stable endosymbiotic relationship with host eukaryotic cells. Although their genomes underwent substantial reduction, resulting in only 110–130 genes remaining in the chloroplast genomes of most terrestrial plants [3], they retained independent replication and transcription systems and evolved into distinct circular quadripartite structure [4]. This semi-autonomous nature makes chloroplast genomes valuable molecular markers for phylogenetic reconstruction, species identification, and genetic improvement, particularly in distinguishing closely related taxa and detecting hybridization events [5].

Due to their high copy number, maternal inheritance, and conserved gene content and organization, chloroplast genomes are powerful tools for studying phylogenetic relationships among closely related species [6, 7, 8]. In angiosperms, the typical chloroplast genome displays a highly conserved quadripartite structure in which a pair of 26 kb inverted repeats (IRs) divide the circular DNA molecule into large and small single copy regions (LSC and SSC). This structure is important in maintaining genome stability and facilitating homologous recombination [9]. However, recent studies have revealed structural variations across plant lineages, including IR boundary expansion or contraction, gene loss, and sequence inversion, which provide new perspectives for investigating plant adaptive evolution [10, 11, 12].

Many species in the *Paeoniaceae* family hold not only exceptional ornamental value but also significant medicinal importance. Among them, *Paeonia suffruticosa* 'Cun Song Ying' is recognized as an important cultivated variety owing to its distinctive pink flowers and robust growth performance. Although advances in high-throughput sequencing have included more than 7,000 plant chloroplast genomes in the NCBI Organelle Genome Database [13], the chloroplast genome features and phylogenetic placement of this cultivar remain unexplored.

Although the genetic code is highly conserved, the frequency of synonymous codon usage varies substantially across species, which is known as codon usage bias [14]. Population genetics studies indicate that although synonymous sites are under relatively weak selective pressure, codon usage patterns reflect the combined effects of mutation, natural selection, and genetic drift over long-term evolution [15]. Codon usage bias, defined as the non-random use of synonymous codons encoding the same amino acid, is shaped by complex evolutionary mechanisms. Investigating codon usage bias in plant chloroplast genomes can provide insights into phylogenetic relationships, horizontal gene transfer, and molecular evolution, while also contributing to the optimization of gene expression, enhancement of genetic transformation, and theoretical guidance for species conservation [16, 17].

In this study, we assembled, annotated, and performed comparative analyses of the chloroplast genome of *Paeonia suffruticosa* 'Cun Song Ying'. We systematically examined its structural features, distribution of repetitive sequences, codon usage bias, and IR boundary variation. In addition, we constructed a phylogenetic tree including closely related species to clarify the evolutionary characteristics of its plastid genome and its phylogenetic position within *Paeoniaceae*.

2 Materials and Methods

2.1 Sampling, DNA Extraction, and Chloroplast Genome Sequencing

Fresh leaf samples of *Paeonia suffruticosa* 'Cun Song Ying' were collected at Heze University. The samples were immediately frozen in liquid nitrogen and stored at − 80°C. Total genomic DNA was extracted using the HiPure SF Plant DNA Mini Kit (Magen). After DNA end repair and 3'-end adenylation,

sequencing adapters were ligated, and target fragments were recovered using magnetic bead purification. The fragments were amplified by PCR to construct sequencing libraries. Library quality was assessed prior to sequencing, and qualified libraries were sequenced on the Illumina NovaSeq X Plus platform.

2.2 Sequencing Data Quality Control, Assembly, and Chloroplast Gene Annotation

To ensure the accuracy of downstream analyses, raw reads were filtered using Cutadapt (v1.16) [18] to remove adapter containing reads, reads with more than 10% ambiguous bases (N), and reads in which low quality bases (Phred score < 5) accounted for more than 50% of the sequence length. After stringent filtering, clean reads were retained for subsequent analyses. Base composition distribution was examined to assess possible AT/GC bias. Quality profiles of sequencing reads across all cycles were evaluated using FastQC (v0.11.4) (http://www.bioinformatics.babraham.ac.

uk/projects/fastqc/), providing an overview of sequencing data quality. Genome assembly was performed using NOVOPlasty (v4.2) (https://github.com/ndierckx/ novoplasty) [19], Fast-Plast (v1.2.8) (https://github.com/mrmckain/Fast-Plast), and GetOrganelle (v1.7.0+) (https://github.com/Kinggerm/GetOrganelle) [20]. The assembly with the best overall performance was selected as the final result. Annotation of the assembled chloroplast genome was conducted using PGA [21] and GeSeq(https://chlorobox.mpimp-golm.mpg.de/geseq.html) [22], followed by manual correction of all annotation results.

Predicted protein coding sequences were compared against reference protein databases to obtain functional annotations. To ensure biological relevance, only the best alignment result for each gene was retained. Functional annotation databases included NR (http://www.ncbi.nlm.nih.gov/), Swiss-Prot (http://www.ebi.ac.uk/uniprot), EggNOG (http://eggnogdb.embl.de/), KEGG (http://www.genome.jp/kegg/), and GO (http://geneontology.org/). A circular gene map of the chloroplast genome was generated using the online tool OGDRAW [23] (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html).

2.3 Comparative Genomic Analysis

The chloroplast genome of *Paeonia suffruticosa* ‘Cun Song Ying’ was compared with 15 published *Paeoniaceae* chloroplast genomes (Table S1). Sequence similarity was assessed using mVISTA in LAGAN mode, which enables accurate multiple alignment regardless of potential inversions [24]. Genome sequences were aligned using MAFFT (7.525) [25], and nucleotide diversity (π) between ‘Cun Song Ying’ and other *Paeoniaceae* species was estimated with DnaSP v6.12 using a sliding window analysis (window length: 600 bp; step size: 200 bp) [26]. Results were visualized using ggplot2 package in R (R-4.2.3).

To detect potential genome rearrangements, complete chloroplast genome alignments were conducted with ProgressiveMauve v.1.1.3 implemented in Geneious Prime 2025.2.1 [27]. IRplus

(https://irscope.shinyapps.io/IRplus/) was used to examine the contraction or expansion of inverted repeat boundaries [28].

2.4 Analysis of Repetitive Sequences

Repetitive sequences, including direct, reverse, palindromic, and complementary repeats, were identified using REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer) [29] with the following parameters: minimum repeat size of 30 bp, maximum repeat length of 5,000 bp, Hamming distance of 3,000 bp, and sequence identity ≥ 90%. Simple sequence repeats (SSRs) were detected using MISA-web (http://pgrc.ipk-gatersleben.de/misa/), with the following minimum thresholds: 10 repeat units for mononucleotides, 5 for dinucleotides, 4 for trinucleotides, and 3 for tetranucleotides, pentanucleotides, and hexanucleotides. Relative synonymous codon usage (RSCU) values were calculated using CodonW (v1.4.4) [30].

2.5 Phylogenetic Analysis

Phylogenetic relationships were inferred using the shared protein coding genes (PCGs) from 16 *Paeoniaceae* chloroplast genomes. *Paeonia sterniana* and *Paeonia veitchii* were selected as outgroups (Table S1). Gene sequences were extracted, concatenated, and aligned with MAFFT (7.525) in ClustalW mode, followed by manual inspection to ensure reading frame accuracy. Maximum likelihood (ML) analyses were performed with RAxML v.8.2.4 using 1,000 bootstrap replicates to assess branch support [31]. Bayesian inference (BI) was conducted in MrBayes v.3.2.7 [32] with 1,000,000 Markov chain Monte Carlo (MCMC) generations. The first 25% of sampled trees were discarded as burn-in, and the remaining trees were used to generate a majority rule consensus tree.

For both ML and BI analyses, the best fitting nucleotide substitution model was determined using jModelTest v.2.1.9 [33] under the Akaike Information Criterion (AIC). Phylogenetic trees were visualized with FigTree v.1.4.5 (http://tree.bio.ed.ac.uk/software/figtree/).

3 Results

3.1 Assembly and annotation of the *Paeonia suffruticosa* ‘Cun Song Ying’ chloroplast genome

Illumina sequencing generated 8.5 Gb of raw data, with a Q30 value of 97.64%, indicating high-quality sequencing output. The assembled chloroplast genome of *Paeonia suffruticosa* ‘Cun Song Ying’ was 152,704 bp in length and exhibited the typical quadripartite structure. The LSC region (84,365 bp) and SSC region (17,047 bp) were separated by two IR regions (25,646 bp each) (Fig. 1, Table S2).

A total of 138 genes were annotated, including 4 pseudogenes and 134 functional genes. The functional genes include 87 protein coding genes (PCGs), 39 tRNA genes and 8 rRNA genes. Among these, 74 genes were associated with self-replication: 9 related to the large ribosomal subunit, 14 to the small ribosomal subunit, 4 encoding RNA polymerase subunits, 8 encoding rRNAs, and 39 encoding tRNAs. In

addition, 45 genes were involved in photosynthesis, including 5 encoding subunits of photosystem I, 15 for photosystem II, 12 for NADH dehydrogenase, 6 for the cytochrome b/f complex, 1 for the large subunit of Rubisco, and 6 for ATP synthase.

12 genes were annotated with other or putative functions (matK, clpP, cemA, accD, ccsA) or unknown functions (ycf1, ycf2, ycf3, ycf4, ycf15). 22 genes contained introns: 18 harbored a single intron (atpF, ndhA, ndhB ×2, petB, petD, rpl16, rpl2 ×2, rpoC1, rps16, trnA-UGC ×2, trnI-GAU ×2, trnK-UUU, trnL-UAA, trnV-UAC), while four genes (clpP, rps12 ×2, ycf3) contained 2 introns (Table S3).

3.2 Comparative analysis and nucleotide polymorphism

Sequence similarity across chloroplast genomes of *Paeoniaceae*, including 'Cun Song Ying', was assessed using mVISTA. Coding regions were more conserved than noncoding regions (Figure S1). No major structural rearrangements were detected across the compared genomes.

Nucleotide polymorphism (π) was calculated for 16 *Paeoniaceae* chloroplast genomes. The average π value was 0.0031, ranging from 0 to 0.01658 (Fig. 2). IR regions were the most conserved (average π = 0.0106), whereas higher variability was observed in the LSC (average π = 0.0391) and SSC regions (average π = 0.0458).

Six hypervariable regions (π > 0.01) were identified: psbA (π = 0.02097), ycf4 (π = 0.01122), psbJ–psbL–psbF–psbE (π = 0.01158), psaJ–rpl33–rps12 (π = 0.01186), rpoA–rps11 (π = 0.01174), and ndhF (π = 0.01181). These regions represent promising candidates for molecular markers in peony phylogenetics and population genetics (Fig. 2).

3.3 Tandem repeats and simple sequence repeats (SSRs)

In the *Paeonia suffruticosa* 'Cun Song Ying' genome, we identified a total of 43 repetitive sequences. Among them, there were 21 palindrome repeats and 22 forward repeats, which constituted the main repetitive sequences of this genome. In the 15 species of peonies used for comparison, a total of 616 repetitive sequences were identified. The types of these sequences were highly consistent: palindrome repeats (300) and forward repeats (309) dominated, and only 7 reverse repeats were found (Fig. 3A, Table S4). The lengths of the repetitive sequences in all species ranged from 30 to 79 bp, and repetitive sequences with lengths of 31, 32, 34, 35, 39, 52 and 54 bp were present in all 16 species, showing high conservation. While repetitive sequences with lengths of 38 bp (present in 14 species) and 41 bp (present in 13 species) were also relatively common (Fig. 3B).

The simple sequence repeat (SSR) analysis revealed that the chloroplast genome of the *Paeonia suffruticosa* 'Cun Song Ying' contained 73 SSRs, with mononucleotide repeats being the dominant type (49), followed by dinucleotide (12), trinucleotide (7), and tetranucleotide repeats (5) (Fig. 3C, Table S5). Among the compared species of the Paeonia, the total number of SSRs varied, ranging from 55 (*Paeonia sterniana*) to 75 (*Paeonia jishanensis*). Among these, the SSR numbers of species such as *Paeonia*

sterniana and *Paeonia veitchii* were relatively low (55–57). The pentanucleotide SSRs was a rare type, only detected in *Paeonia delavayi* and *Paeonia jishanensis* (Fig. 3D, TableS5). Additionally, the analysis of the IR region boundaries indicated that *Paeonia suffruticosa* 'Cun Song Ying' had a typical tetrad structure: the LSC/IRb and LSC/IRa boundaries were located at the rps19 and trnH genes, while the IRa/SSC and IRb/SSC boundaries were located at the ycf1 and ndhF genes. Comparative analysis showed that the length variation range of the IR region in all Paeonia species was narrow (24,863–25,651 bp), and the gene composition of this region was completely conserved (Fig. 4).

3.4 Relative synonymous codon usage (RSCU)

RSCU values were calculated from the complete coding sequence of the *Paeonia suffruticosa* 'Cun Song Ying' chloroplast genome, based on 69 PCGs and 24,297 codons. Leucine (Leu, 10.40%) was the most abundant amino acid, whereas cysteine (Cys, 1.13%) was the least frequent. The most frequently used codon was AUU (1,011 counts; encoding isoleucine), while UGC (71 counts; encoding cysteine) was the least common (Table S6).

Eighteen codons displayed codon usage bias with RSCU > 1.0. ATG and TGG showed no codon usage bias (RSCU = 1.0). Codon usage bias was particularly evident in leucine: UUA exhibited the highest RSCU value (1.97), while CUG had the lowest (0.37). Overall, codons ending in A/U were preferred over G/C-ending codons, consistent with trends reported for other *Paeonia* species (Fig. 5).

3.5 Phylogenetic analysis

Phylogenetic relationships were reconstructed using a concatenated nucleotide dataset of 69 PCGs (64,654 bp) from 16 *Paeoniaceae* chloroplast genomes (Fig. 6). The topologies of the ML and BI trees were highly congruent and strongly supported.

Herbaceous peonies (*Paeonia veitchii* and *Paeonia sterniana*) formed distinct outgroups. All woody peonies clustered into a strongly supported monophyletic clade. Within this clade, *Paeonia ludlowii* and *Paeonia delavayi* diverged first, forming an independent lineage. The core woody peonies were divided into two major sister lineages.

The first lineage included wild ancestors such as *Paeonia baokangensis*, *Paeonia qiui, Paeonia decomposita*, and *Paeonia jishanensis*, together with the cultivated varieties *Paeonia suffruticosa* 'Luo Yang Hong' and its closely related Japanese cultivar *Paeonia suffruticosa* 'Cun Song Ying'. The second lineage comprised both wild and cultivated taxa, including *Paeonia ostii* and *Paeonia rockii*, along with well-known cultivars such as *Paeonia suffruticosa* 'Yao Huang', 'Dou Lv', 'Cao Zhou Hong' and interspecific hybrids such as *Paeonia Itoh*.

This phylogenetic framework highlights the multiple hybrid origins of cultivated peonies and strongly supports the extremely close genetic relationship between *Paeonia suffruticosa* 'Cun Song Ying' and *Paeonia suffruticosa* 'Luo Yang Hong'.

4 Discussion

This study presents the first complete chloroplast genome of *Paeonia suffruticosa* ‘Cun Song Ying’, including its sequencing, assembly, and annotation, followed by a systematic analysis of genome architecture, repeat content, codon usage bias, and phylogeny. Our findings enrich chloroplast genomic resources for *Paeoniaceae* and provide new insights for cultivar identification, phylogenetic reconstruction, and plastid evolutionary mechanisms.

4.1 Structure and conservation of the ‘Cun Song Ying’ chloroplast genome

The chloroplast genome of ‘Cun Song Ying’ is 152,704 bp and displays the canonical circular quadripartite structure comprising the LSC, SSC, and two IR regions. We annotated 134 functional genes, including 87 protein-coding genes, 39 tRNA genes, and 8 rRNA genes. Gene content, order and copy number closely match those reported for other *Paeonia* species [34], indicating strong overall conservation of chloroplast genomes within *Paeoniaceae*. This pattern is consistent with the structural stability of plastid genomes across core angiosperms.

IR regions are widely considered critical for maintaining chloroplast genome stability. In most plants they are highly conserved. For example, Xiao et al. (2025) examined 21 Camellia species and reported conserved SC/IR boundaries across the genus [35]. Similarly stable IR regions have been documented in *Capsicum* [36], *Sapindaceae* [37], and wild Prunus species [38]. In our analysis, boundary genes at the IR junctions (such as rps19, trnH, ycf1, and ndhF) and their positions showed no notable variation among *Paeonia* taxa, supporting strong IR conservation among close relatives. By contrast, the LSC and SSC regions exhibited greater sequence variability, particularly in noncoding intervals, highlighting promising loci for developing high resolution molecular markers.

4.2 Sequence variation and identification of hypervariable regions

Nucleotide polymorphism (π) analyses showed higher conservation in coding regions than in noncoding regions, in line with trends observed in most plant chloroplast genomes. π values in the IR were markedly lower than those in the LSC and SSC, reinforcing the stabilizing role of the IR.

We identified six hypervariable regions ($\pi > 0.01$), spanning coding and intergenic segments associated with psbA, ycf4, psbJ–psbL–psbF–psbE, psaJ–rpl33–rps12, rpoA–rps11, and ndhF. These regions also display elevated variability in related taxa and thus represent candidate DNA barcodes for cultivar identification and population genetic studies in peonies [39, 40]. Notably, ycf1 and ndhF carry strong phylogenetic signals across diverse plant groups [41, 42], making them suitable for resolving interspecific relationships.

4.3 Repeat elements and SSR distribution

This study identified 43 repetitive sequences (21 palindromic repeats and 22 forward repeats) in *Paeonia suffruticosa* 'Cun Song Ying'. Its length is concentrated in two specific intervals of 30–35 bp and 39–54 bp, and this pattern is highly conserved within the *Paeonia*. This discovery suggests that these repetitive sequences may not have been randomly generated, but rather constrained by some evolutionary mechanism. We speculate that these length conserved repetitive sequences constitute potential "hotspots" for homologous recombination, and they are not only the source of genomic sequence microevolution, but also an important intrinsic driving force for macroscopic structural variations such as IR region boundary expansion/contraction. Therefore, indepth study of repetitive sequences provides a new perspective for understanding the mechanisms of maintaining and breaking through the stability of chloroplast genome structure in *Paeonia*.

SSR analysis revealed 73 SSR sites, dominated by mononucleotide repeats (49), followed by dinucleotide (12), trinucleotide (7), and tetranucleotide (5) repeats. The number of SSRs varies among *Paeonia* species (approximately 55–75), a polymorphism useful for cultivar discrimination and genetic diversity assessment. Pentanucleotide SSRs were rare and detected only in a few species, suggesting possible lineage specificity.

4.4 Codon usage bias and evolutionary implications

Codon usage analysis showed a pronounced preference for A/U ending codons, a common pattern in land plants likely shaped by combined effects of mutational bias and natural selection. Leucine (Leu) was the most frequently encoded amino acid, whereas cysteine (Cys) was least frequent. The RSCU value for TTA (Leu) reached 1.97, indicating strong preferential use.

These biases may reflect optimization of translational efficiency and accuracy over evolution. A/U ending codons may better match the chloroplast tRNA pool and translation machinery, potentially enhancing expression efficiency. Understanding codon usage patterns helps elucidate molecular evolutionary dynamics and can inform strategies to optimize heterologous gene expression in chloroplasts.

4.5 Phylogenetic relationships and implications for cultivar origins

Phylogenetic analysis based on chloroplast genome strongly supports the formation of sisters group relationship (Bootstrap = 100) between *Paeonia suffruticosa* 'Cun Song Ying' and *Paeonia suffruticosa* 'Luo Yang Hong'. The substantial progress brought about by this study lies in the fact that this result, from the perspective of the plastid genome inherited through matrilineal inheritance, firmly confirms that the direct maternal source of the rare variety 'Cun Song Ying' cultivated in Japan should belong to the *Paeonia cathayana* variety group.

Woody peonies formed a well supported monophyletic group that segregated into two major clades. Clade I included wild species such as *Paeonia baokangensis*, *Paeonia qiui, Paeonia decomposita*, and *Paeonia jishanensis*, together with cultivars including 'Luo Yang Hong' and 'Cun Song Ying'. Clade II

comprised *Paeonia rockii*, *Paeonia ostii*, and their derivative cultivars such as ‘Yao Huang’, ‘Dou Lv’, and *Paeonia Itoh*, a lineage that appears to have evolved independently in both genetic and morphological traits relative to Clade I.

In summary, our chloroplast genome analyses elucidate the precise phylogenetic relationship between *Paeonia suffruticosa* ‘Cun Song Ying’ and ‘Luo Yang Hong’, providing a theoretical basis for cultivar identification and determining the genetic origin of tree peonies. Future studies that integrate nuclear genomic data will further illuminate the history of their hybridization and domestication.

5 Conclusions

This study is the first to assemble and annotate the rare Japanese variety *Paeonia suffruticosa* ‘Cun Song Ying’. Through analysis of 16 chloroplast genomes from the *Paeoniaceae* family, it was found that all genomes exhibited typical tetrad structures with highly similar sizes (152153–152958 bp). Comparative genomics analysis shows that the expansion/contraction of the IR region and repetitive sequence variations are the main causes of genome size differences. We identified a large number of SSR loci and 6 high frequency variation regions (Pi > 0.01), providing valuable resources for subsequent population genetics research. Phylogenetic analysis strongly supports a sister group relationship (Bootstrap = 100) between ‘Cun Song Ying’ and the *Paeonia cathayana* cultivar ‘Luo Yang Hong’ in China. Furthermore, maternal inheritance analysis confirms that the female parent of 'Cun Song Ying' is derived from the *Paeonia cathayana* cultivar group. This study lays an important genomic foundation for the identification, conservation, and breeding within the genus *Paeonia*.

Abbreviations

LSC, large single-copy region; SSC, small single-copy region; SC, single copy region; IR, inverted repeat region; PCGs, protein coding genes; SSRs, simple sequence repeats; RSCU, Relative synonymous codon usage.

Declarations

Conflicts of Interest

The authors declare that there is no conflict of interest.

Funding

Doctoral Fund Project of Heze University (XY23BS28).

Author Contribution

Wenzhen Cheng and Mingyu Li: designed the study. Wenzhen Cheng: Samples collection, manuscript writing and funding acquisition. Mingyu Li and Conghao Hong: data analysis. Changyong Gao: project ministration, writing-revision and editing.

Acknowledgments

We are grateful to Professor Hongbo Gao from Beijing Forestry University for his suggestions on this work.

Data Availability

The sequence data generated in this study are available in GenBank of the National Center for Biotechnology Information (NCBI) under the access number PX252292.

References

1. Bock R. Witnessing Genome Evolution: Experimental Reconstruction of Endosymbiotic and Horizontal Gene Transfer. *Annu Rev Genet*. 2017;51:1–22. doi:10.1146/annurev-genet-120215-035329.
2. Zimorski V, Ku C, Martin WF, Gould SB. Endosymbiotic theory for organelle origins. *Curr Opin Microbiol*. 2014;22:38–48. doi:10.1016/j.mib.2014.09.008.
3. Wu XX, Mu WH, Li F, et al. Cryo-EM structures of the plant plastid-encoded RNA polymerase. *Cell*. 2024;187(5):1127–1144.e21. doi:10.1016/j.cell.2024.01.026
4. Zhang Y, Tian L, Lu C. Chloroplast gene expression: Recent advances and perspectives. *Plant Commun*. 2023;4(5):100611. doi:10.1016/j.xplc.2023.100611.
5. Qin Q, Dong Y, Chen J, et al. Comparative analysis of chloroplast genomes reveals molecular evolution and phylogenetic relationships within the Papilionoideae of Fabaceae. *BMC Plant Biol*. 2025;25(1):157. Published 2025 Feb 6. doi:10.1186/s12870-025-06138-0.
6. Dong W, Xu C, Li W, et al. Phylogenetic Resolution in *Juglans* Based on Complete Chloroplast Genomes and Nuclear DNA Sequences. *Front Plant Sci*. 2017;8:1148. Published 2017 Jun 30. doi:10.3389/fpls.2017.01148.
7. Dong W, Xu C, Wu P, et al. Resolving the systematic positions of enigmatic taxa: Manipulating the chloroplast genome data of Saxifragales. *Mol Phylogenet Evol*. 2018;126:321–330. doi:10.1016/j.ympev.2018.04.033.
8. Dong W, Xu C, Wen J, Zhou S. Evolutionary directions of single nucleotide substitutions and structural mutations in the chloroplast genomes of the family Calycanthaceae. *BMC Evol Biol*. 2020;20(1):96. Published 2020 Jul 31. doi:10.1186/s12862-020-01661-0.

9. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol*. 2016;17(1):134. Published 2016 Jun 23. doi:10.1186/s13059-016-1004-2.
10. Barrett CF, Wicke S, Sass C. Dense infraspecific sampling reveals rapid and independent trajectories of plastome degradation in a heterotrophic orchid complex. *New Phytol*. 2018;218(3):1192–1204. doi:10.1111/nph.15072.
11. Barrett CF, Sinn BT, Kennedy AH. Unprecedented Parallel Photosynthetic Losses in a Heterotrophic Orchid Genus. *Mol Biol Evol*. 2019;36(9):1884–1901. doi:10.1093/molbev/msz111.
12. Jiang D, Cai X, Gong M, et al. Complete chloroplast genomes provide insights into evolution and phylogeny of Zingiber (Zingiberaceae). *BMC Genomics*. 2023;24(1):30. Published 2023 Jan 18. doi:10.1186/s12864-023-09115-9.
13. Hua Z, Tian D, Jiang C, et al. Towards comprehensive integration and curation of chloroplast genomes. *Plant Biotechnol J*. 2022;20(12):2239–2241. doi:10.1111/pbi.13923.
14. Shabalina SA, Spiridonov NA, Kashina A. Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity. *Nucleic Acids Res*. 2013;41(4):2073–2094. doi:10.1093/nar/gks1205.
15. Hershberg R, Petrov DA. Selection on codon bias. *Annu Rev Genet*. 2008;42:287–299. doi:10.1146/annurev.genet.42.110807.091442.
16. Parvathy ST, Udayasuriyan V, Bhadana V. Codon usage bias. *Mol Biol Rep*. 2022;49(1):539–565. doi:10.1007/s11033-021-06749-4.
17. Chen Y, Zhao Y, Yan Q, et al. Characterization and Phylogenetic Analysis of the First Complete Chloroplast Genome of *Shizhenia pinguicula* (Orchidaceae: Orchideae). *Genes (Basel)*. 2024;15(11):1488. Published 2024 Nov 20. doi:10.3390/genes15111488.
18. Kechin A, Boyarskikh U, Kel A, Filipenko M. cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing. *J Comput Biol*. 2017;24(11):1138–1143. doi:10.1089/cmb.2017.0096.
19. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res*. 2017;45(4):e18. doi:10.1093/nar/gkw955.
20. Jin JJ, Yu WB, Yang JB, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. 2020;21(1):241. Published 2020 Sep 10. doi:10.1186/s13059-020-02154-5.
21. Qu XJ, Moore MJ, Li DZ, Yi TS. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods*. 2019;15:50. Published 2019 May 21. doi:10.1186/s13007-019-0435-7.
22. Tillich M, Lehwark P, Pellizzer T, et al. GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 2017;45(W1):W6-W11. doi:10.1093/nar/gkx391.
23. 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-W64.

doi:10.1093/nar/gkz238.

24. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res*. 2004;32(Web Server issue):W273-W279. doi:10.1093/nar/gkh458.
25. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–780. doi:10.1093/molbev/mst010.
26. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, et al. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol Biol Evol*. 2017;34(12):3299–3302. doi:10.1093/molbev/msx248.
27. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*. 2004;14(7):1394–1403. doi:10.1101/gr.2289704.
28. Díez Menéndez C, Poczai P, Williams B, Myllys L, Amiryousefi A. IRplus: An Augmented Tool to Detect Inverted Repeats in Plastid Genomes. *Genome Biol Evol*. 2023;15(10):evad177. doi:10.1093/gbe/evad177.
29. 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):4633–4642. doi:10.1093/nar/29.22.4633.
30. Sharp PM, Li WH. Codon usage in regulatory genes in Escherichia coli does not reflect selection for 'rare' codons. *Nucleic Acids Res*. 1986;14(19):7737–7749. doi:10.1093/nar/14.19.7737.
31. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312–1313. doi:10.1093/bioinformatics/btu033.
32. Ronquist F, Teslenko M, van der Mark P, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61(3):539–542. doi:10.1093/sysbio/sys029.
33. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods*. 2012;9(8):772. Published 2012 Jul 30. doi:10.1038/nmeth.2109.
34. Guo S, Guo L, Zhao W, et al. Complete Chloroplast Genome Sequence and Phylogenetic Analysis of *Paeonia ostii*. *Molecules*. 2018;23(2):246. Published 2018 Jan 26. doi:10.3390/molecules23020246.
35. Xiao X, Chen J, Ran Z, Huang L, Li Z. Comparative Analysis of Complete Chloroplast Genomes and Phylogenetic Relationships of 21 Sect. *Camellia* (*Camellia* L.) Plants. *Genes (Basel)*. 2025;16(1):49. Published 2025 Jan 3. doi:10.3390/genes16010049.
36. Sebastin R, Kim J, Jo IH, et al. Comparative chloroplast genome analyses of cultivated and wild Capsicum species shed light on evolution and phylogeny. *BMC Plant Biol*. 2024;24(1):797. Published 2024 Aug 24. doi:10.1186/s12870-024-05513-7.
37. Li, J., Wang, H., Wang, L., Wang, X., Jia, L., & Chen, Z (2025). Comprehensive analysis of the complete chloroplast genome of the cultivated soapberry and phylogenetic relationships of Sapindaceae. *Industrial Crops and Products*. 2025;228(120952):0926–6690. https://doi.org/10.1016/j.indcrop.2025.120952.

38. Cui M, Liu C, Yang X, et al. Comparative and Phylogenetic Analysis of the Chloroplast Genomes of Four Wild Species of the Genus *Prunus*. *Genes (Basel)*. 2025;16(3):239. Published 2025 Feb 20. doi:10.3390/genes16030239.
39. Frigerio J, Agostinetto G, Mezzasalma V, De Mattia F, Labra M, Bruno A. DNA-Based Herbal Teas' Authentication: An ITS2 and *psbA-trnH* Multi-Marker DNA Metabarcoding Approach. *Plants (Basel)*. 2021;10(10):2120. Published 2021 Oct 6. doi:10.3390/plants10102120.
40. Zhang, J. H., Liu, S. H., Zhang, Z. F., Shi, Y., Man, J. H., Yin, G. Y., Wang, X., Liu, F. B., Wang, X. H., & Wei, S. L. Identification and quality evaluation of germplasm resources of commercial Scutellaria baicalensis based on DNA barcode and HPLC. *Zhongguo Zhong Yao Za Zhi*. 2022;47(7):1814–1823. https://doi.org/10.19540/j.cnki.cjcmm.20220117.101.
41. Yang X, Yu X, Zhang X, et al. Development of Mini-Barcode Based on Chloroplast Genome and Its Application in Metabarcoding Molecular Identification of Chinese Medicinal Material Radix *Paeoniae Rubra* (Chishao). *Front Plant Sci*. 2022;13:819822. Published 2022 Mar 31. doi:10.3389/fpls.2022.819822.
42. Amar MH. ycf1-ndhF genes, the most promising plastid genomic barcode, sheds light on phylogeny at low taxonomic levels in Prunus persica. *J Genet Eng Biotechnol*. 2020;18(1):42. Published 2020 Aug 14. doi:10.1186/s43141-020-00057-3.

Figures

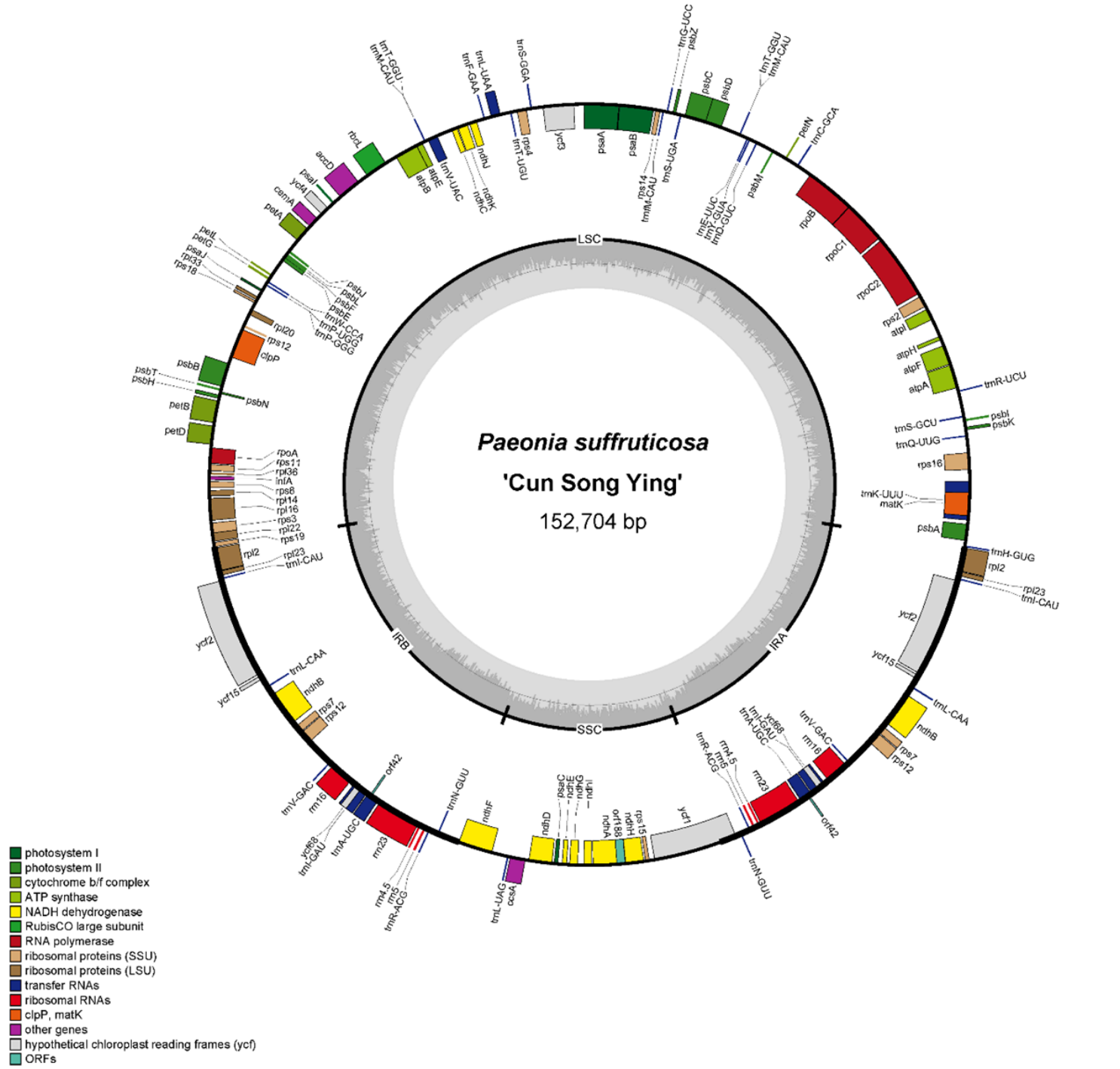

Figure 1

Chloroplast genome map of *Paeonia suffruticosa* 'Cun Song Ying'. Genes inside the circle are transcribed clockwise, while those outside are transcribed counter clockwise. Functional groups are indicated by color. The inner dark gray circle represents GC content and the light gray circle represents AT content.

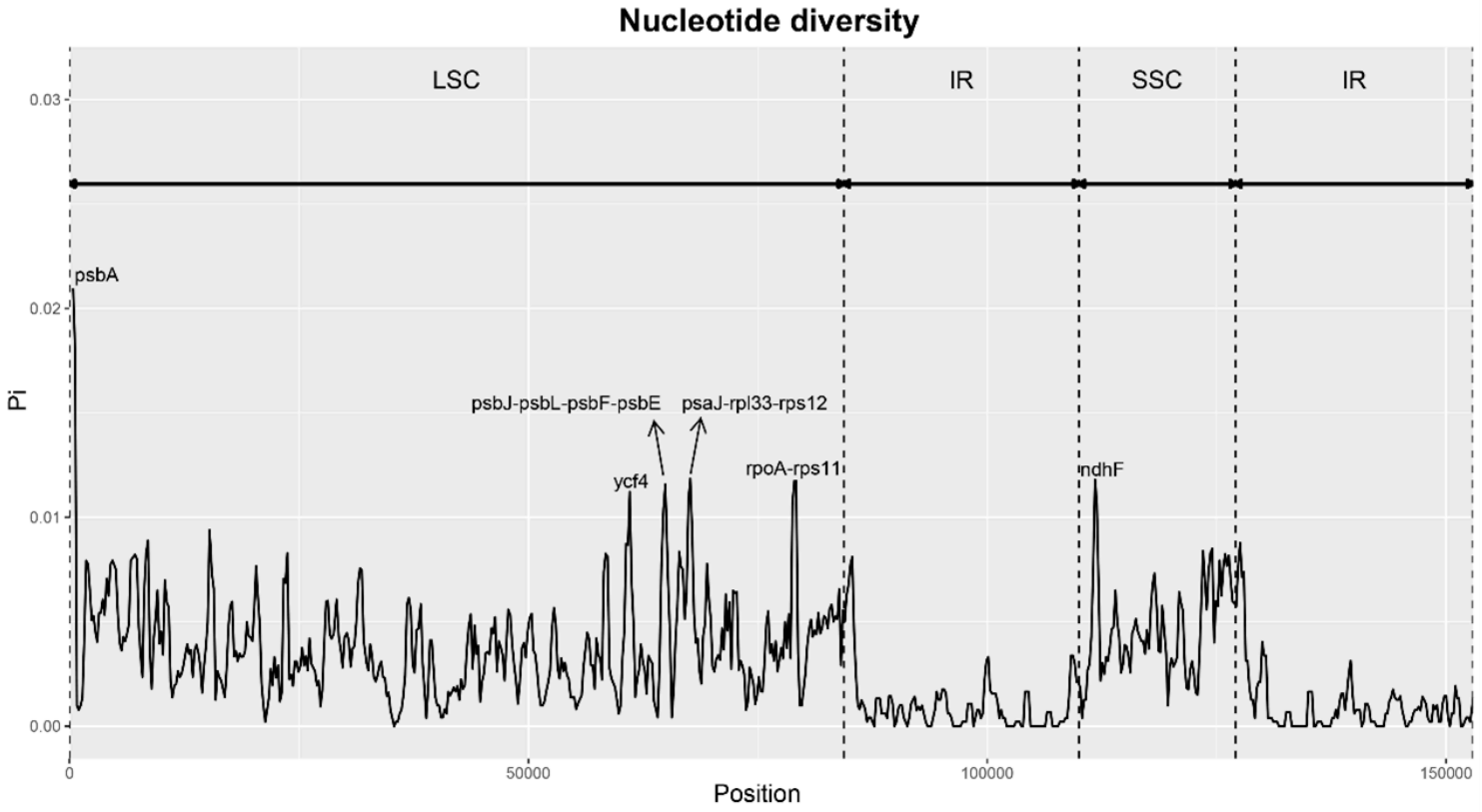

Figure 2

Nucleotide diversity (π) across 16 peonies chloroplast genomes.

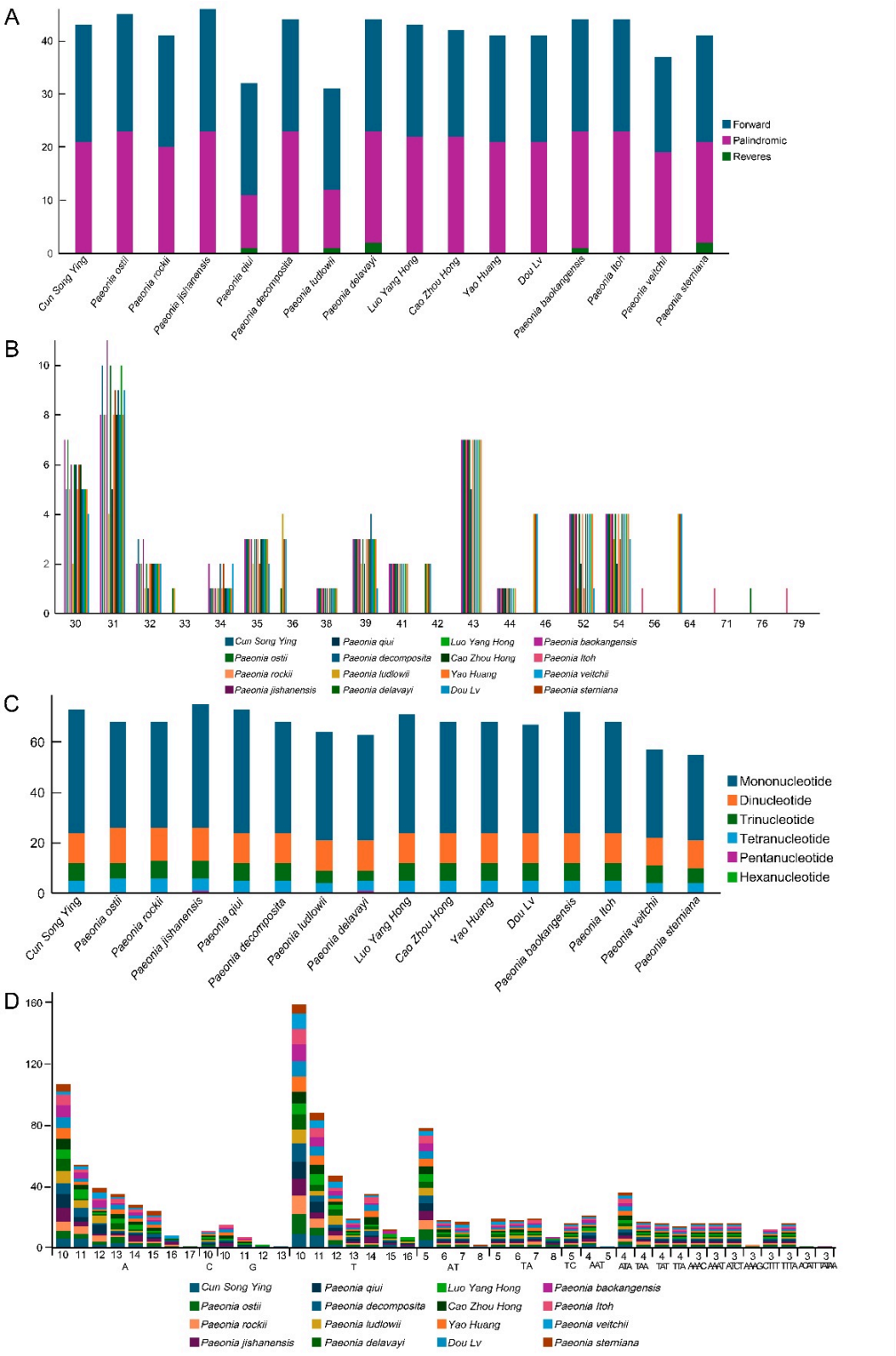

Figure 3

Repeat sequence analyses across 16 *Paeoniaceae* chloroplast genomes.

(A) Distribution of tandem repeat types. (B) Frequency of tandem repeats of different lengths. (C) Distribution of SSR motifs. (D) Frequency of SSR motif types.

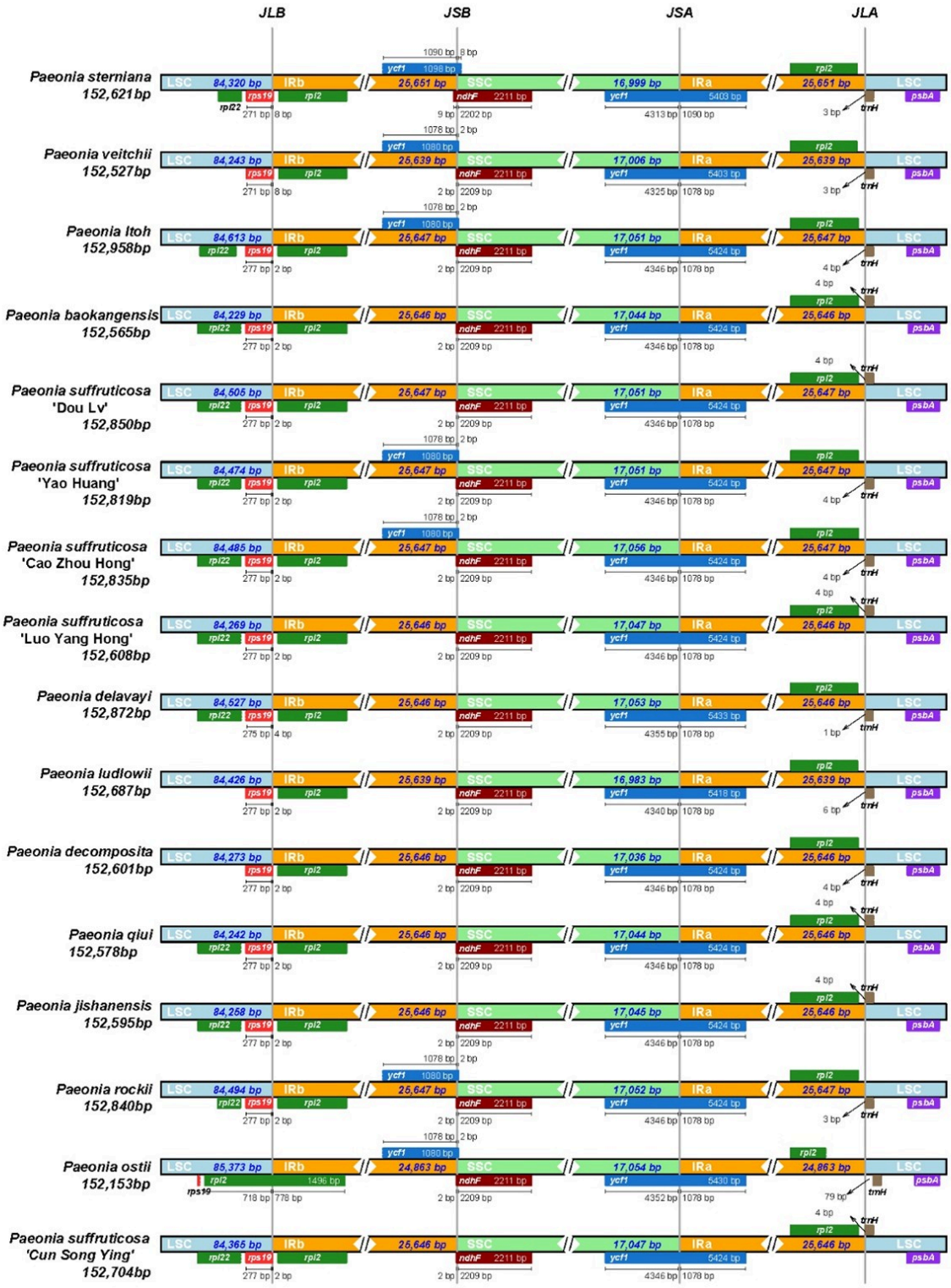

Figure 4

Comparison of LSC, IR, and SSC boundaries among *Paeoniaceae* species.

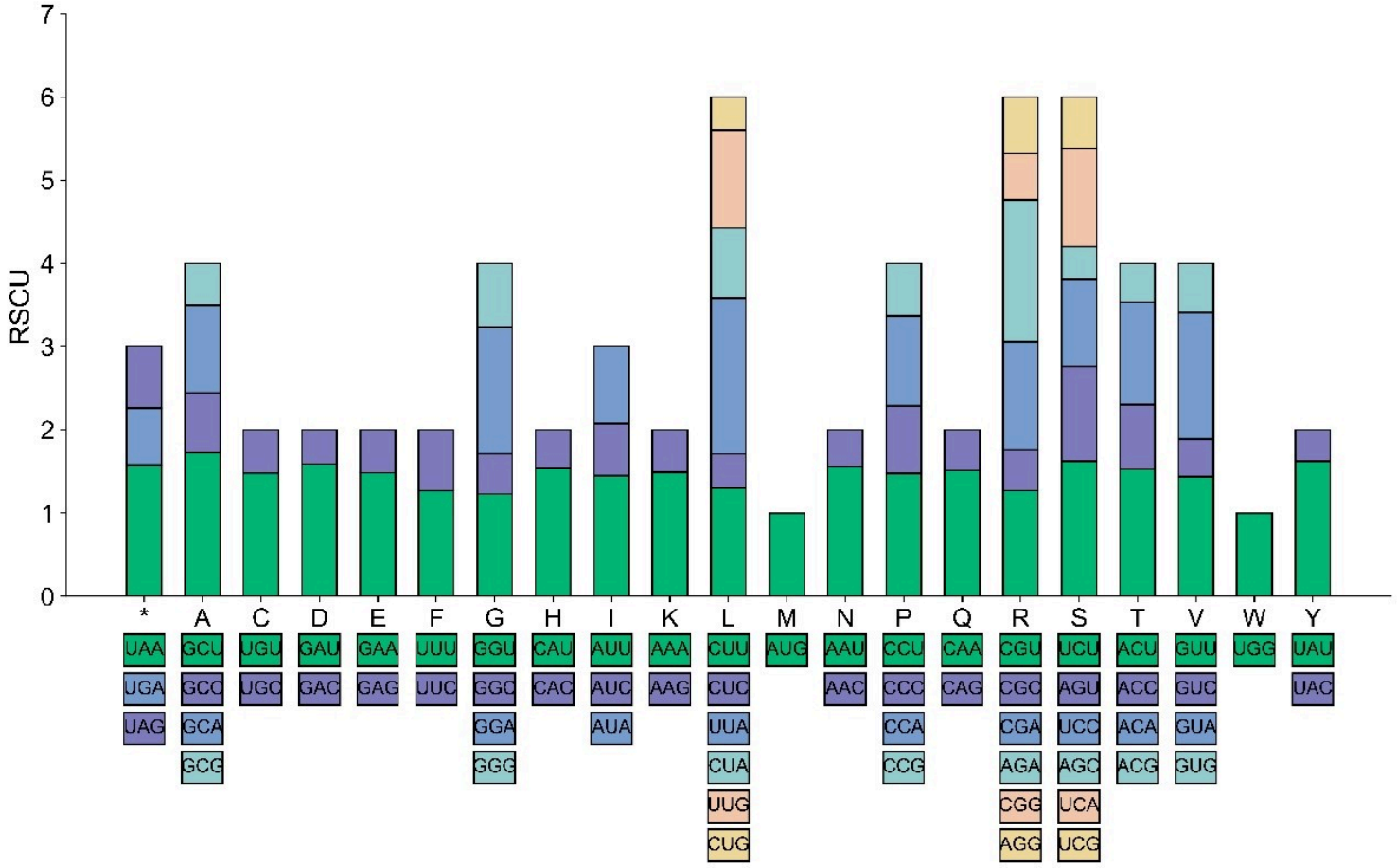

Figure 5

Relative synonymous codon usage (RSCU) values for 20 amino acids in chloroplast protein coding genes of *Paeonia*. The x-axis represents amino acids, the y-axis represents RSCU values, and colors indicate different synonymous codons. Note: “*” stands for the termination codon.

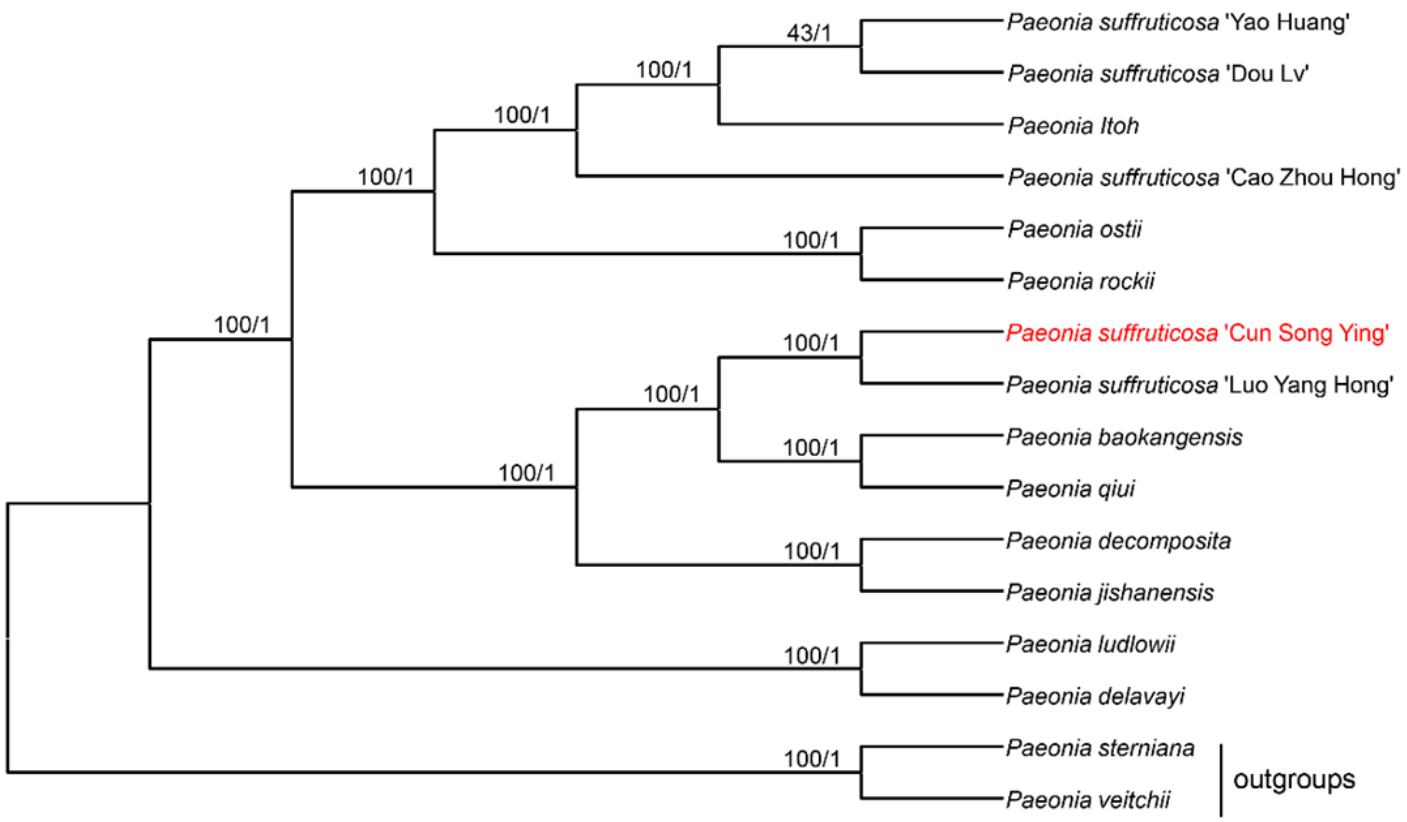

Figure 6

Phylogenetic tree of *Paeoniaceae* inferred from concatenated chloroplast CDS sequences of 69 PCGs. The tree was constructed using ML and BI methods (Bootstrap = 100, Posterior Probability = 1.00). Numbers above branches represent bootstrap values and posterior probabilities.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementaltable.xlsx