Zhang et al. *New Zealand Journal of Forestry Science* (2026) 56:8
https://doi.org/10.33494/nzjfs562026x503x
E-ISSN: 1179-5395
published on-line: 06/07/2026

Bioeconomy Science Institute
Maiangi Taiao

RESEARCH ARTICLE Open Access

New Zealand Journal of Forestry Science

Comparative analysis of the chloroplast genomes of three *Populus* species: Insights into genetic relationships, evolution and classification

Liangxin Zhang[1], Jiaxuan Mi[1], Xueqin Wan[1], Shuaiying Zhang[1], Xin Huang[1], Xiandan Luo[1], Jianghong Qian[1], Lianghua Chen[1], Fan Zhang[2], Fang He[1], Xiong Huang[1], Yunjie Gu[3], Hongying Guo[3], Zhi Chen[3] and Hanbo Yang[1*]

[1] *Forestry Ecological Engineering in the Upper Reaches of the Yangtze River Key Laboratory of Sichuan Province, National Forestry and Grassland Administration Key Laboratory of Forest Resource Conservation and Ecological Safety on the Upper Reaches of the Yangtze River, College of Forestry, Sichuan Agricultural University, Chengdu, China*
[2] *College of Landscape Architecture, Sichuan Agricultural University, Chengdu, China*
[3] *Sichuan Academy of Forestry, Sichuan Key Laboratory of Ecological Restoration and Conservation for Forest and Wetland, Chengdu, China*

*Corresponding author: yanghanbo6@sicau.edu.cn

(Received for publication 17 October 2025; accepted in revised form 6 June 2026)

Editor: Clemens M. Altaner

Abstract

Background: Natural hybridisation is prevalent among *Populus* species in the Hengduan Mountains. Clarifying their interrelationships and hybrid formation mechanisms is critical for taxonomic classification and breeding of the genus.

Methods: This study sequenced the complete chloroplast genomes of three *Populus* species - *P. lasiocarpa* Oliv. (Sect. Leucoides Spach), *P. gonggaensis* N. Chao & J.R. He (Sect. Leucoides Spach), and *P. cathayana* Rehd. (Sect. Tacamahaca Spach) - to investigate their genetic relationships and evolutionary patterns.

Results: The chloroplast genome lengths were 156,554 bp (*P. lasiocarpa*), 156,494 bp (*P. gonggaensis*), and 156,812 bp (*P. cathayana*), with all sharing a 37% GC content and conserved quadripartite structure (large/small single-copy regions plus two inverted repeats). They contained 130 or 131 genes (85 or 86 protein-coding, 37 tRNA, 8 rRNA) and 129-142 simple sequence repeat (SSR) loci (primarily adenine/thymine-dominated mononucleotide repeats). Phylogenetic analysis revealed the three species form a monophyletic clade, with *P. gonggaensis* and *P. cathayana* showing close affinity.

Conclusions: Chloroplast evidence indicated that *P. gonggaensis* is maternally closely related to *P. cathayana*, supporting its maternal lineage from *P. cathayana*. Further nuclear genome evidence is needed to confirm the hybrid origin of this species.

Keywords: *Populus cathayana*; *Populus gonggaensis*; *Populus lasiocarpa*; phylogeny; genetic variation

Introduction

Populus belongs to the Salicaceae family and encompasses approximately 30 woody species that are widely distributed, primarily in the Northern Hemisphere (Racchi et al. 2011). In the Flora of China, *Populus* is classified into five sections, namely, Sect. *Aigeiros* Spach, Sect. *Leucoides* Spach, Sect. *Populus* Spach, Sect. *Tacamahaca* Spach, and Sect. *Turanga* Spach (Fang et al. 1999). Despite their economic value, rapid growth, and ecological significance, species of the genus *Populus* (genome size: 550 Mbp) face ongoing taxonomic challenges. These challenges arise from morphological plasticity, frequent natural hybridisation, especially within the sections *Aigeiros*

and *Tacamahaca*, and the limited genetic validation of traditional classifications (Eckenwalder 1977, 1996; Djerbi et al. 2005; Tuskan et al. 2006; Bylesjö et al. 2008; Ellis et al. 2010). Recent comprehensive genomic studies have revealed extensive gene flow among various *Populus* species. Furthermore, these studies have demonstrated that *P. ningshanica*, *P. wulianensis*, *P. tomentosa*, and *P. canescens* all originated from hybridisation (L. Zhang et al. 2018; M. C. Wang et al. 2020; Liu et al. 2022; H. Zhang et al. 2023). The Qinghai–Tibet Plateau has emerged as the central area of contemporary distributional diversity for the genus *Populus*, where this phenomenon is particularly pronounced. In this region, a total of 28 species have been identified, including three species in Sect. *Populus*, five species in Sect. *Leucoides*, and twenty species in Sect. *Tacamahaca* (Fang et al. 1999). Notably, approximately ten of these species are of natural hybrid origin (Wan & Zhang 2013). The complex species composition and frequent natural hybridisation events not only position the Qinghai–Tibet Plateau as a natural laboratory for investigating hybrid speciation mechanisms but also make it one of the most taxonomically challenging regions for *Populus* classification worldwide. In light of these characteristics, it is essential to systematically enhance germplasm identification and novel cultivar selection and conduct molecular breeding research to refine the *Populus* taxonomic framework (Shi et al. 2024). The representative species *Populus lasiocarpa* Oliv., *Populus gonggaensis* N. Chao & J.R. He, and *Populus cathayana* Rehd. are pivotal in addressing taxonomic controversies within the genus *Populus*. Notably, *P. gonggaensis*, which is confined to Kangding, Sichuan Province, China, displays phenotypic intermediacy in comparison to *P. lasiocarpa*. Initial molecular studies suggested that *P. gonggaensis* may have a hybrid origin, arising from the intersection of Sect. *Leucoides* Spach and Sect. *Tacamahaca* Spach (Wan & Zhang 2013; Zong, Zhou, et al. 2019; Y. C. Wang et al. 2022; Korbik & Kosinski 2023). Recent nuclear genome resequencing has confirmed the identification of *P. gonggaensis*, *P. lasiocarpa* (Sect. *Leucoides* Spach), and *P. cathayana* (Sect. *Tacamahaca* Spach) (Du et al. 2024). Molecular studies consistently position *P. cathayana* within Sect. *Tacamahaca*. In contrast, *P. gonggaensis* and *P. lasiocarpa* exhibit characteristics typical of Sect. *Leucoides*. Comparative analyses of their chloroplast genomes remain limited. Although the chloroplast genomes of the three species (*P. gonggaensis*, *P. lasiocarpa* and *P. cathayana*) have been reported, the phylogenetic relationships and hybrid origin of *P. gonggaensis* have not been investigated using maternal chloroplast genome data (Zong, Gan, et al. 2019; Shi et al. 2024).

Chloroplasts are the central organelles for photosynthesis in green plants and perform key functions, including carbon fixation and the regulation of plant growth (Palmer 1987). Compared with the nuclear genome, the chloroplast genome has a significantly slower evolutionary rate (Daniell et al. 2016). The wealth of valuable information contained within chloroplast genomes has established them widely in phylogenetic and taxonomic research. As a result, chloroplast genomes have been extensively sequenced and are frequently utilised for comparative and phylogenetic analyses (Djerbi et al. 2005; Lin et al. 2010; F.-H. Wu et al. 2010; Pan et al. 2012; Cho et al. 2015; Zhou et al. 2017; Shin et al. 2019; Androsiuk et al. 2020; Yue et al. 2023). Dong et al. (2013) sequenced and compared the plastomes of Saxifragales, including *Sedum*, *Paeonia*, and *Liquidambar*, and revealed conserved gene structures along with genus-specific nucleotide variations. Lin et al. (2020) conducted a comparative analysis of the plastomes of *Vasconcellea pubescens* A.DC. and *Carica papaya* L., revealing both conserved genes and genus-specific variations, as well as variable regions that are useful for phylogenetic studies. In summary, owing to their conserved structure, slow evolutionary rates, and abundant phylogenetic information, chloroplast genomes are indispensable tools for resolving taxonomic ambiguities and reconstructing evolutionary histories across various plant lineages.

In this study, we addressed the aforementioned gap by conducting *de novo* sequencing of chloroplast genomes from three individual trees per species of *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana*. Specifically, our objectives include reconstructing their phylogenetic network through comprehensive whole chloroplast genome comparisons, elucidating the maternal lineage of *P. gonggaensis* via detailed analyses of structural and sequence variations, and identifying hypervariable chloroplast genome regions suitable for molecular diagnostics. By integrating these analyses, this study aims to establish a robust genomic framework that clarifies their evolutionary relationships and supports molecular germplasm identification.

Methods

Plant materials and chloroplast DNA extraction

Fresh leaves were collected from three individual trees per species: *P. lasiocarpa* (102°13’41” E, 30°1’26” N, elevation 2,216 m), *P. gonggaensis* (103°38’43” E, 30°33’24” N, elevation 2,743 m), and *P. cathayana* (102°43’39” E, 30°48’2” N, elevation 2,423 m) in the Hengduan Mountains. All materials were morphologically identified based on standardised phenotypic characteristics, and the sampling locations and taxonomic identities of these three species were previously documented in Shi et al. (2024). The samples were deposited at the Chongzhou Modern Agricultural R&D Base (103°38’43” E, 30°33’24” N, elevation 512 m). We extracted the chloroplast DNA of *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana* using a Super Plant Genomic DNA DP360 Kit (Tiangen Biotech, Beijing, China). The quality and concentration of the chloroplast DNA were evaluated via 1.0% agarose gel electrophoresis and a Nanodrop 2000 (Thermo Fisher Scientific, USA).

Sequencing and assembly of the chloroplast genome
We then conducted chloroplast genome sequencing of the species *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana*. Library preparation was carried out using an Illumina TruSeq DNA Sample Preparation Kit (Illumina, USA) in accordance with the manufacturer's instructions. The prepared libraries were subsequently sequenced on the Illumina Nova-Seq platform with paired-end reads of 150 base pairs. The raw sequencing data obtained from the Illumina platform were subjected to quality control using Fastp v0.36 (https://github.com/OpenGene/fastp) (Chen et al. 2018). Low-quality reads and adapter sequences were removed using Trimmomatic v0.39 (Bolger et al. 2014). The sequences with cp-like reads were assembled using NOVOPlasty v4.3.1 (Dierckxsens et al. 2017) to obtain the complete chloroplast genomic sequence.

Genome annotation and analysis
The online tool GeSeq was used to annotate the chloroplast genomes of *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana*, which employs a blast-based approach to identify and annotate genes, tRNAs, and rRNAs (Tillich et al. 2017). The coding sequence (CDS) was predicted using Prodigal v2.6.3 (https://www.github.com/hyattpd/Prodigal), rRNA prediction was performed using HMMER v3.1b2 (http://www.HMMER.org/) (Oh et al. 2019), and tRNA prediction was performed using ARAGORN v1.2.38 (https://www.trna.se/ARAGORN/) (Lohse et al. 2007). The predicted genes and RNA structures were manually curated and verified using the BLAST tool against the NCBI (https://www.ncbi.nlm.nih.gov/) nucleotide and protein databases. A circular chloroplast genetic map was created using Chloroplot (https://irscope.shinyapps.io/Chloroplot/) (Zheng et al. 2020), which provided a convenient and quick interface for visualising and annotating organelle genomes.

Chloroplast genome comparison and IR boundary analysis
We conducted a comparative analysis of the chloroplast genomes of *P. lasiocarpa*, *P. gonggaensis* and *P. cathayana.* mVISTA (https://genome.lbl.gov/vista/index.shtml) (Mayor et al. 2000) was used in Shuffle-LAGAN mode to perform whole-genome comparisons of the three species. Moreover, we employed CGView online software (https://proksee.ca/) (Grant et al. 2023) to carry out GView analyses of the chloroplast genomes. Furthermore, to circumvent the limitation of studying only *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana*, we opted to analyse multiple *Populus* species to increase the representativeness and reliability of the research. The IRscope visualisation tool (https://irscope.shinyapps.io/irapp/) (Y. C. Wang et al. 2022) was utilised to examine the IR/LSC and IR/SSC boundary disparities across 24 *Populus* chloroplast genomes. The dataset comprised three newly sequenced genomes in this work and 21 publicly available plastomes. All species names, GenBank accession numbers, and corresponding references are compiled in Table S1. By comparing the chloroplast genomes of 24 *Populus* species, we obtained a more comprehensive understanding of the genetic diversity and evolutionary relationships of *Populus* species (Amiryousefi et al. 2018). Using DnaSP v6 software (http://www.ub.edu/dnasp) (Rozas et al. 2017), the nucleotide polymorphism (Pi) of the 24 chloroplast genomes was estimated with a sliding window 600 bp in length and a step size of 200 bp. The eight fragments exhibiting the highest Pi values were identified as variation hotspots among *Populus* species, and their precise locations were determined according to the annotation information.

Analysis of codon usage and microsatellites
CodonW software was used to conduct statistical analyses of codon usage (https://codonw.sourceforge.net/) (Peden 1999). The relative synonymous codon usage (RSCU) of chloroplast DNA was determined and visualised using the EMBOSS browser and bioinformatics cloud (http://cloud.genepioneer.com:9929/#/) (Rice et al. 2000). To compare the differences in the GC contents of protein-coding genes among *P. lasiocarpa*, *P. gonggaensis*, *P. cathayana* and other *Populus* species, we downloaded the complete chloroplast genomes (gb files) of 21 *Populus* species from NCBI (Table S1). All subsequent sequence processing and analyses (including GC content calculation) were performed independently by our research team. We subsequently utilised Microsoft Excel software to calculate and tabulate the GC content of all protein-coding genes separately. REPuter software in the BiBiServ2 browser (https://bibiserv.cebitec.uni-bielefeld.de/) (Kurtz et al. 2001) with parameters set at a minimum length of 8 bp was adopted to detect dispersed repetitive sequences, including direct, palindromic, inverted, and complementary repeats. The Hamming distance was set to 3. MISA v1.0 (http://pgrc.ipk-gatersleben.de/misa/misa.html) (Thiel et al. 2003; Beier et al. 2017) was used to detect simple sequence repeat sequences (SSRs) with the following parameters: at least 10 single-nucleotide repeat units, at least 6 double-nucleotide repeat units, and at least 5 triple, quadruple, quintuple, and hexanucleotide repeat units. The detection results were statistically analysed and plotted using Chiplot online software (https://www.chiplot.online/).

Phylogenetic analysis
To determine the phylogenetic positions of *P. lasiocarpa*, *P. gonggaensis*, *P. cathayana* in *Populus*, MAFFT software (https://mafft.cbrc.jp/alignment/software/) (Nakamura et al. 2018) was used to perform multiple sequence alignment of these chloroplast genome sequences compared with 21 other *Populus* species listed in Table S1. MEGA11 software (https://www.megasoftware.net/) (Tamura et al. 2021) with default parameter settings and 1,000 bootstrap replications was subsequently used to construct a maximum likelihood (ML) phylogenetic tree. iTOL v6.9.1 online software (https://itol.embl.de/) (Letunic & Bork 2024) was used to visualise the phylogenetic tree.

Results

Genome features of *Populus lasiocarpa*, *P. gonggaensis* and *P. cathayana*

The chloroplast sequence lengths of *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana* were 156,554 bp, 156,494 bp, and 156,812 bp, respectively. Their GC content was consistently 37%, and they all presented highly conserved quadripartite structures comprising of a large single-copy (LSC), a small single-copy (SSC), inverted repeat A (IRA) and inverted repeat B (IRB) regions (Figures 1 & 2). *Populus lasiocarpa*, *P. gonggaensis*, and *P. cathayana* contained 130, 131, and 131 genes in their whole chloroplast genomes, respectively. These genes consisted of 85-86 protein-coding genes (85 in *P. lasiocarpa* and 86 in the other two species), 37 tRNA genes, and 8 rRNA genes. Among these genes, 73 genes belong to the self-replicating class, and 37, 14, and 10 genes encode tRNAs, small ribosomal subunit proteins, and large ribosomal subunit proteins, respectively.

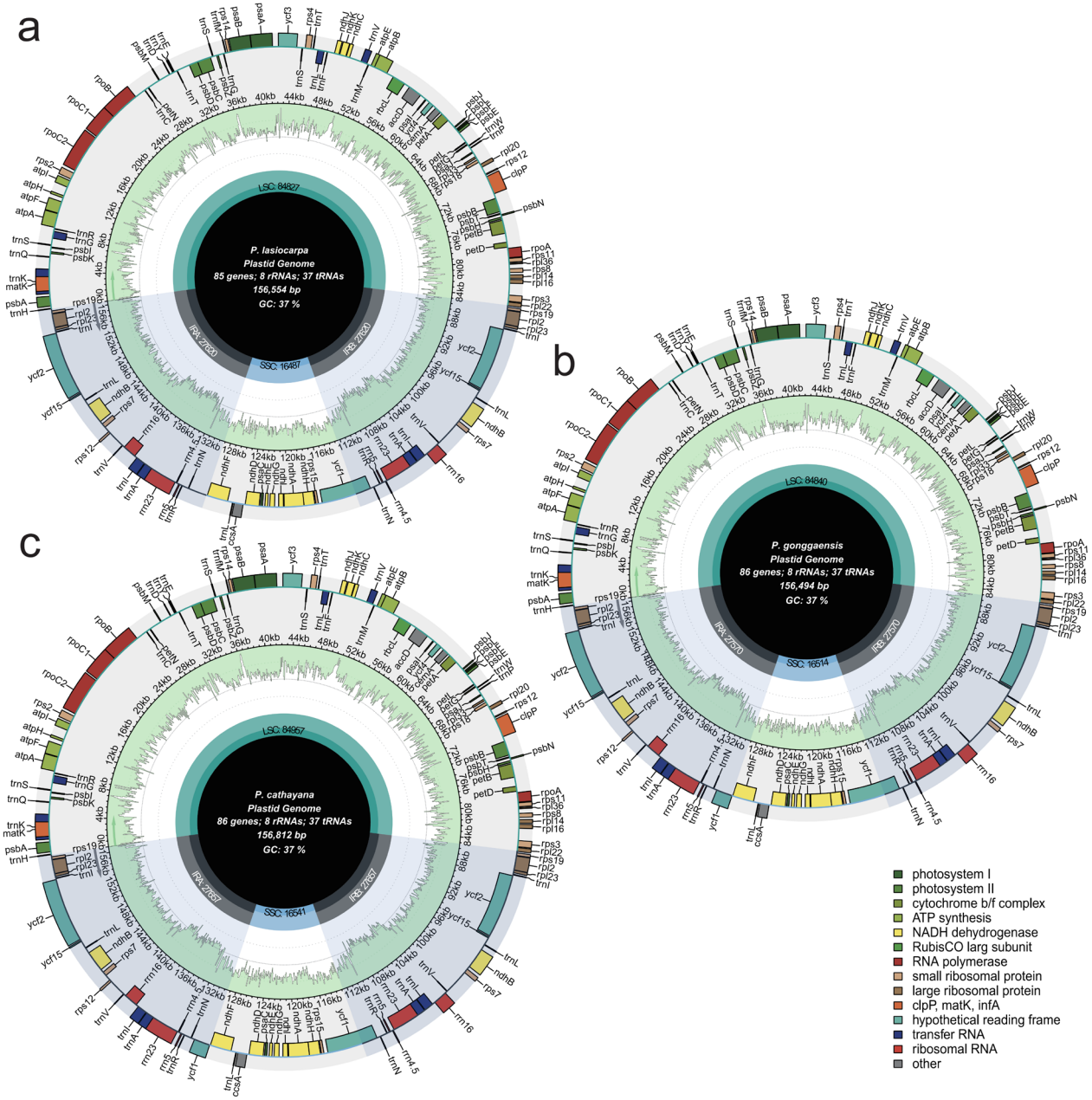

FIGURE 1: Chloroplast genome maps of *P. lasiocarpa* (a), *P. gonggaensis* (b), and *P. cathayana* (c), depicted in a circular format. From the innermost to the outermost regions, these maps consist of three concentric layers: GC content distribution, base sequencing depth, and gene element annotation. In the outermost gene element layer, the inner ring indicates forward-transcribed genes, whereas the outer ring corresponds to reverse-transcribed genes. The legend on the right identifies various functional gene categories, such as photosystem I/II and cytochrome b/f complex components, providing an intuitive visualisation of the structural organisation, gene distribution, and fundamental features (e.g., genome size, gene count, and GC content) of the chloroplast genomes of the *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana*.

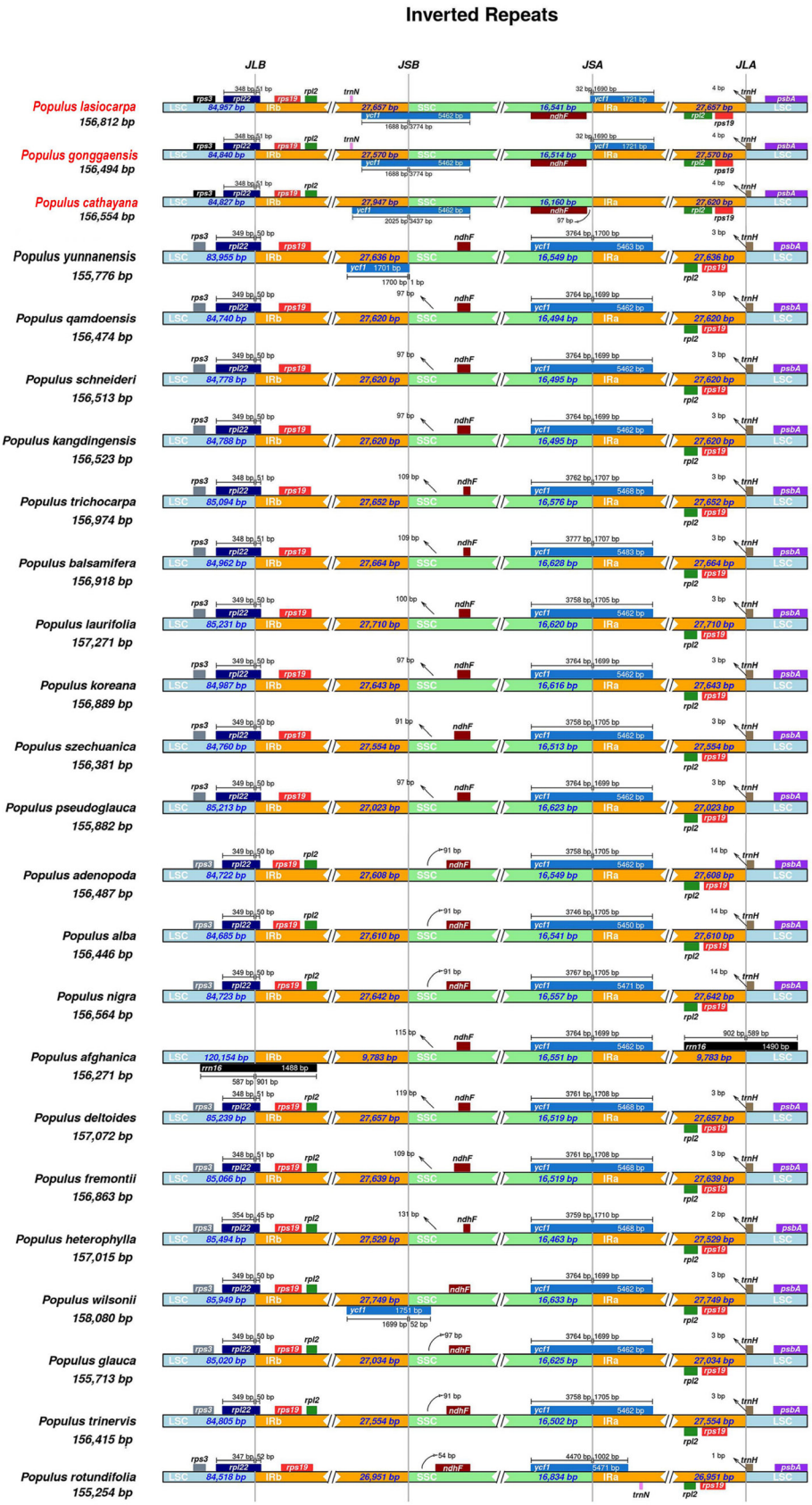

FIGURE 2: Comparison of the SSC, LSC, IRB and IRA regions in the chloroplast genomes of 24 *Populus* species. The figure illustrates the structural features of the chloroplast genomes from 24 *Populus* species. Moving from left to right, the regions are labeled as LSC (large single-copy region), IRb (inverted repeat b), SSC (small single-copy region), and IRa (inverted repeat a). The lengths of these regions are clearly indicated, with interregional boundaries highlighted in gray. Additionally, key genes and their boundary characteristics in selected areas are concurrently displayed.

Forty-eight genes were associated with photosynthesis, of which 6, 6, 12, 8, 15, and 1 encoded the ATP synthase subunit, NADH plastiquinone oxidoreductase subunit proteins, cytochrome b/f complex subunit proteins, photosystem I subunit enzymes, PSII subunits, and the Rubisco large subunit, respectively. There were also five groups belonging to the biosynthesis category, each containing one gene (Tables S2-S4).

Among the chloroplast genomes of 24 *Populus* species, the number of nucleotide polymorphisms (Pi) ranged from 0 to 0.1225 (Figure 3). The LSC regions presented the lowest average Pi values (< 0.004), reflecting minimal differentiation. Based on the Pi values, eight highly variable regions were identified (Pi > 0.11), including the *ndhA* gene; the intergenic regions between *ndhA* and *ndhI*, *ndhI* and *ndhG*, *ndhD* and *ccsA*, *trnL* and *ndhF*, *ndhF* and *trnN*; and the *rrn23* gene and the intergenic region between *trnA* and *trnI*. Notably, the *trnL-ndhF* region presented the highest level of polymorphism (Pi = 0.1225).

Frequency of codon usage and GC content

The total numbers of codons in protein-coding genes ranged from 52,162 to 52,270, and the codon usage patterns of the three target species in this study were nearly identical. Excluding termination codons, leucine (5,107-5,306 codons, 9.79-10.17%) was the most abundant amino acid, and tryptophan (817-874 codons) was the least abundant amino acid (Table S5). According to the RSCU values, synonymous codon preferences could be classified as nonpreferential (RSCU ≤ 1) or preferential (RSCU > 1). In the present study, 30 codons exhibited preferences, whereas 34 codons did not (Figure 4). Additionally, except for UUG, most of the preferred codons ended with A or U. The GC content is one of the most important features of the chloroplast genome. We determined the GC contents of all coding genes of *P. lasiocarpa*, *P. gonggaensis*, *P. cathayana*, and the 21 *Populus* species downloaded from NCBI (Table S1). There were identifiable differences in the GC contents of the coding genes among these

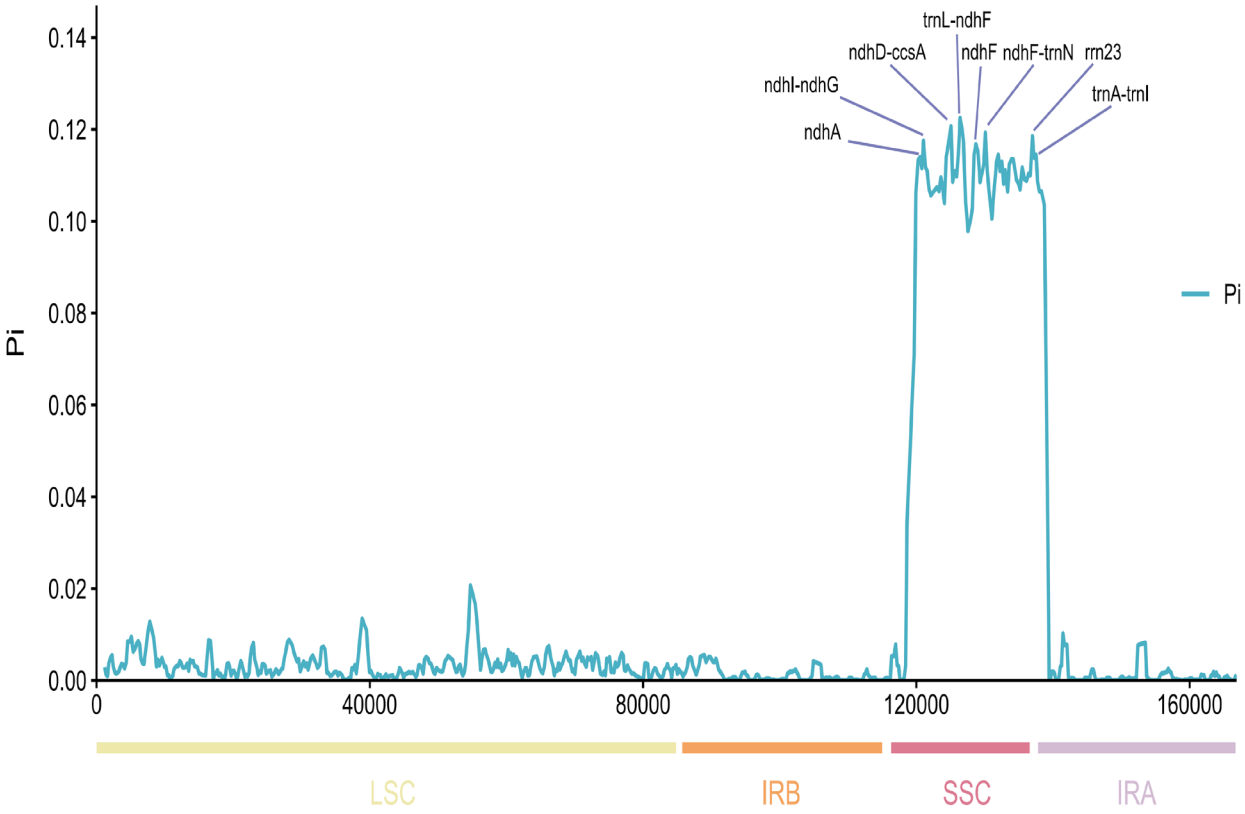

FIGURE 3: Analysis of nucleotide polymorphisms in the chloroplast genomes of 24 *Populus* species. The horizontal axis in the figure indicates the coordinates of the chloroplast genome, whereas the differently coloured regions (LSC, IRB, SSC, IRA) below the axis represent the structural divisions of the chloroplast genome. The vertical axis shows nucleotide diversity (Pi). A higher Pi value suggests a greater degree of variation at that site, indicating that it is a high-variation hotspot. The eight marked sites in the figure correspond to the regions with the highest Pi values: *ndhA*, the intergenic region between *ndhI* and *ndhG*, the intergenic region between *ndhD* and *ccsA*, the intergenic region between *trnL* and *ndhF*, *ndhF*, the intergenic region between *ndhF* and *trnN*, *rrn23*, and the intergenic region between *trnA* and *trnI*.

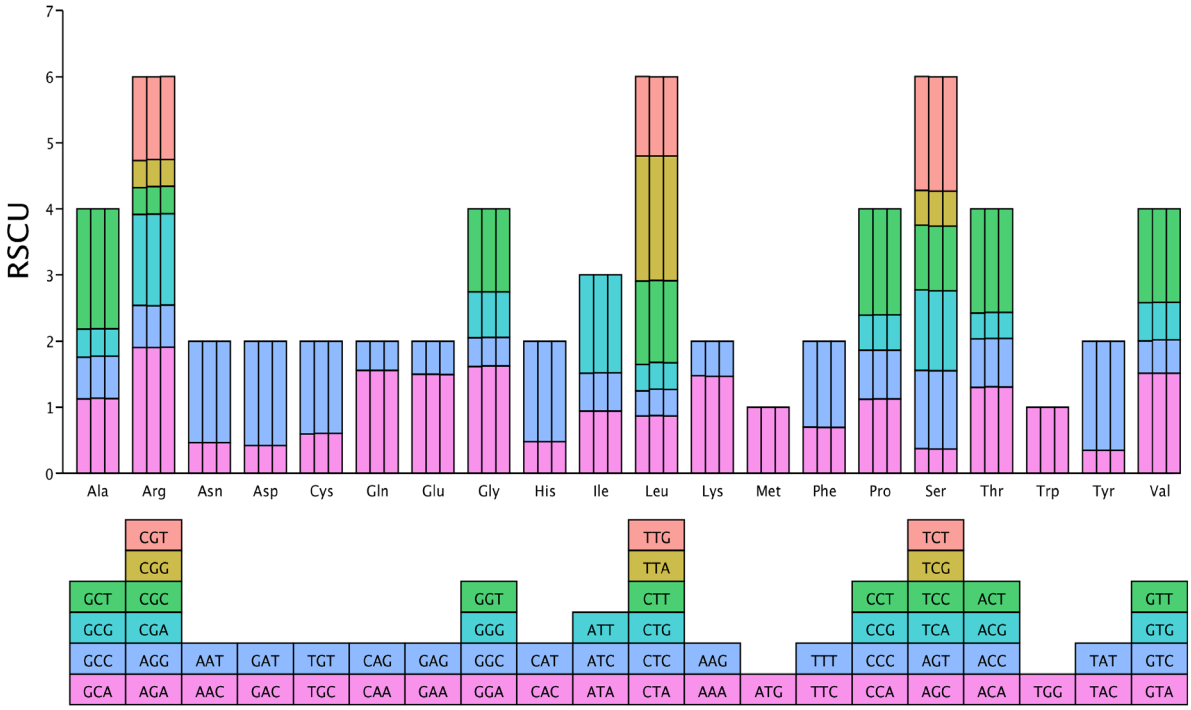

FIGURE 4: Frequency of codons in the chloroplast genomes of *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana* displayed in histograms. The y-axis represents the RSCU (relative synonymous codon usage), which reflects the degree of preference for specific codons. The x-axis lists the 20 amino acids and their corresponding codon types. The colour-coded blocks within each bar represent synonymous codons encoding the same amino acid, whereas the height of each colour block visually highlights the differences in codon usage patterns among the chloroplast genomes of *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana*.

Populus species, particularly significant variations in the *ndhE* and *psbZ* genes. Moreover, the GC contents of the coding genes among *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana* were minimally different, and the *clpP*, *ndhA*, *petB*, *psbZ*, *rpl16*, *rps12*, *ycf1*, and *ycf3* genes presented certain differences compared with those of other *Populus* species (Figure 5). Therefore, we could infer that there was kinship among *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana*.

Comparative analysis of the complete chloroplast genomes

We used *P. gonggaensis* as a reference and conducted BLAST analyses on the chloroplast genomes of *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana*. The entire chloroplast genome was highly conserved among species, especially in coding regions (Figure 6). Differences in the noncoding regions among the species were greater than those in the coding regions (Figure 7). CGView analysis revealed notable disparities in the genomic loci of the *ycf1* gene and one of the *rps7* genes between *P. lasiocarpa* and *P. gonggaensis*/*P. cathayana*. The chloroplast genome structure of *P. gonggaensis* maintained a high degree of consistency with that of *P. cathayana*. Under the prevalent matrilineal inheritance pattern in angiosperms, the aforementioned findings indicate that the chloroplast donor of *P. gonggaensis* is likely *P. cathayana* or its closely related taxa.

Repetitive sequence and chloroplast SSR analysis

A total of 129 to 142 SSR loci were identified in the chloroplast genomes of *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana* (Figure 8a). Mononucleotide repeats were the most prevalent, ranging from 99 to 113 (76.74-79.58%). *P. gonggaensis* and *P. cathayana* had four hexanucleotide repeats in their chloroplast genomes. Among dinucleotide repeats, AT/AT repeats were the most common (5, 3.52-3.88% of the total number of SSRs.), followed by AT/TA repeats (3-4, 2.11-3.10% of the total number of SSRs.) and AG/CT repeats (1, 0.70-0.84% of the total number of SSRs.) (Figure 8b-D). Among the three *Populus* species, single-nucleotide SSRs were the most common type of chloroplast SSR, and they were composed mainly of A and T, which might be associated with the high AT content. Among the four types of long repeats, forward repeats and palindromic repeats were the most abundant in the chloroplast genomes (Figure 9a). Additionally, sequences with lengths of 30-45 bp were the most common in these chloroplast genomes, followed by those with lengths of 54-68 bp and then those with lengths of 69-76 bp (Figure 10b). Sequences with lengths of 46-53 bp were only found in *P. cathayana*.

Phylogenetic analysis

To determine the phylogenetic placement of the three focal *Populus* species, a phylogenetic tree was

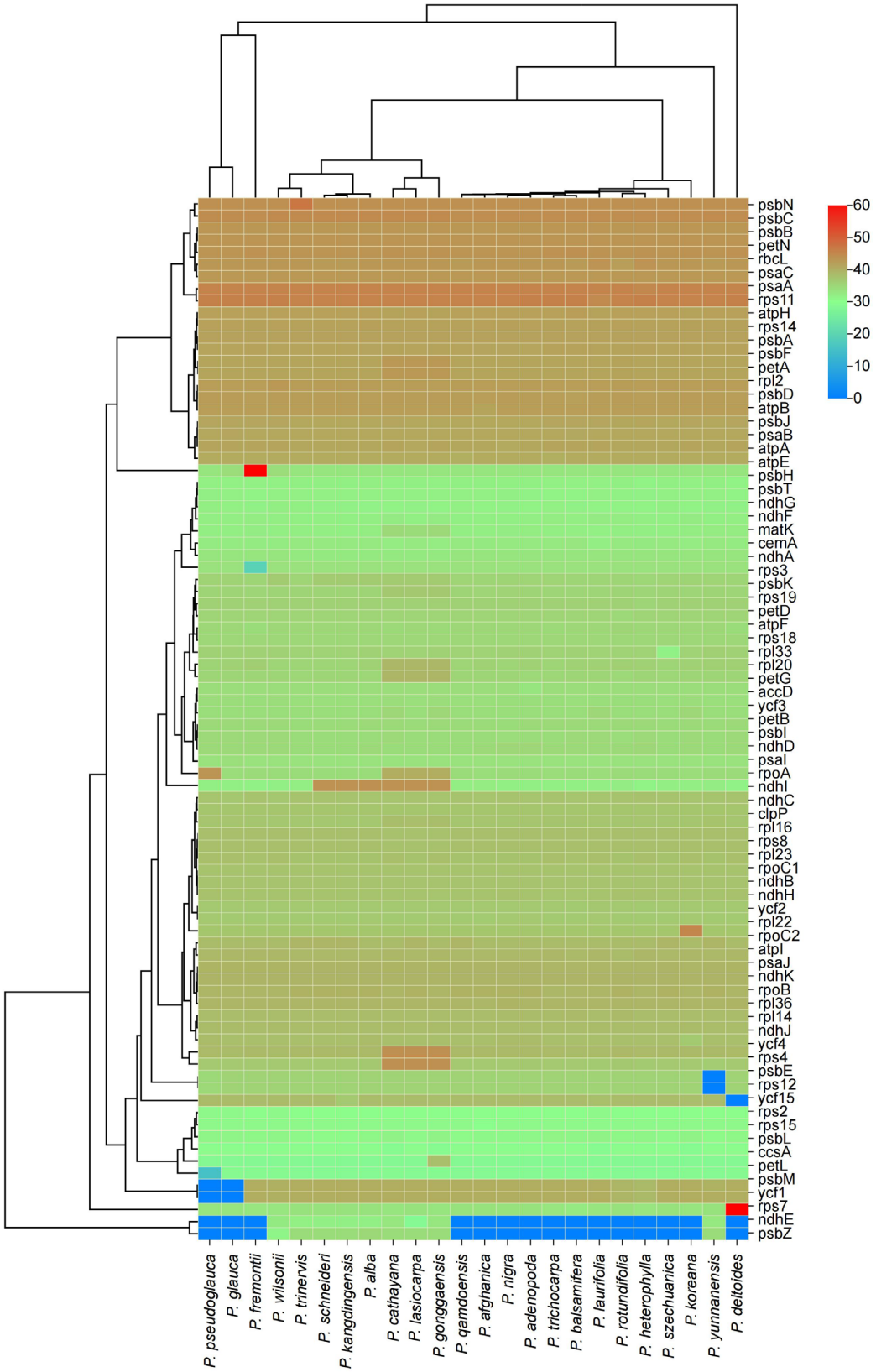

FIGURE 5: Heatmap of the GC contents of all protein-coding genes of 24 *Populus* species. This heatmap illustrates the GC contents of each protein-coding gene in 24 *Populus* species. Gene names are displayed on the right side, and species names are shown at the bottom. The colour gradient transitions from blue to green and then to red, representing an increasing GC content percentage. Combined with the hierarchical clustering tree, the map enabled phylogenetic relationship inference and species grouping based on the similarity of gene GC content profiles. Consistently similar colour intensity for the same gene across species indicates closer phylogenetic relatedness.

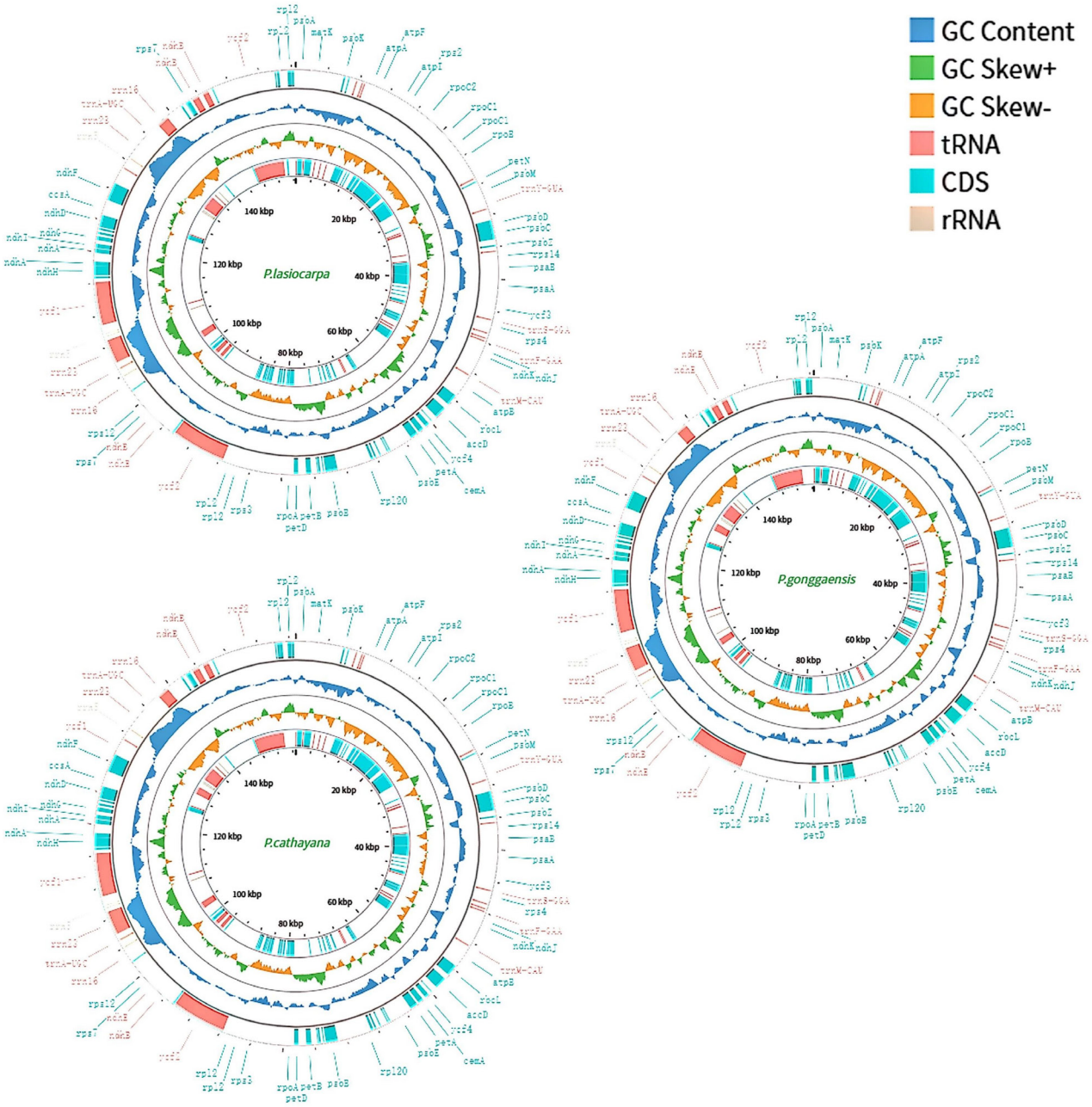

FIGURE 6: Genetic characteristics of *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana.* The inner circle marks the length of the interval zones in clockwise direction. In the middle layer, the green and orange peaks reflect the skew in the GC content on the positive and negative strands of the genome, respectively, and highlighting the asymmetry of the sequence base composition. The dark blue peaks denote the GC content of the corresponding regions. The coloured areas in the innermost and outermost circles represent various functional gene elements: red represents tRNAs (transfer RNA genes), blue–green represents CDSs (coding sequences), and gray–brown represents rRNAs (ribosomal RNA genes). The outermost circle labels gene names. This map systematically illustrates the base composition characteristics and distribution of functional elements in the chloroplast genomes of *P. lasiocarpa*, *P. gonggaensis* and *P. cathayana* through different colours, lines, and annotations.

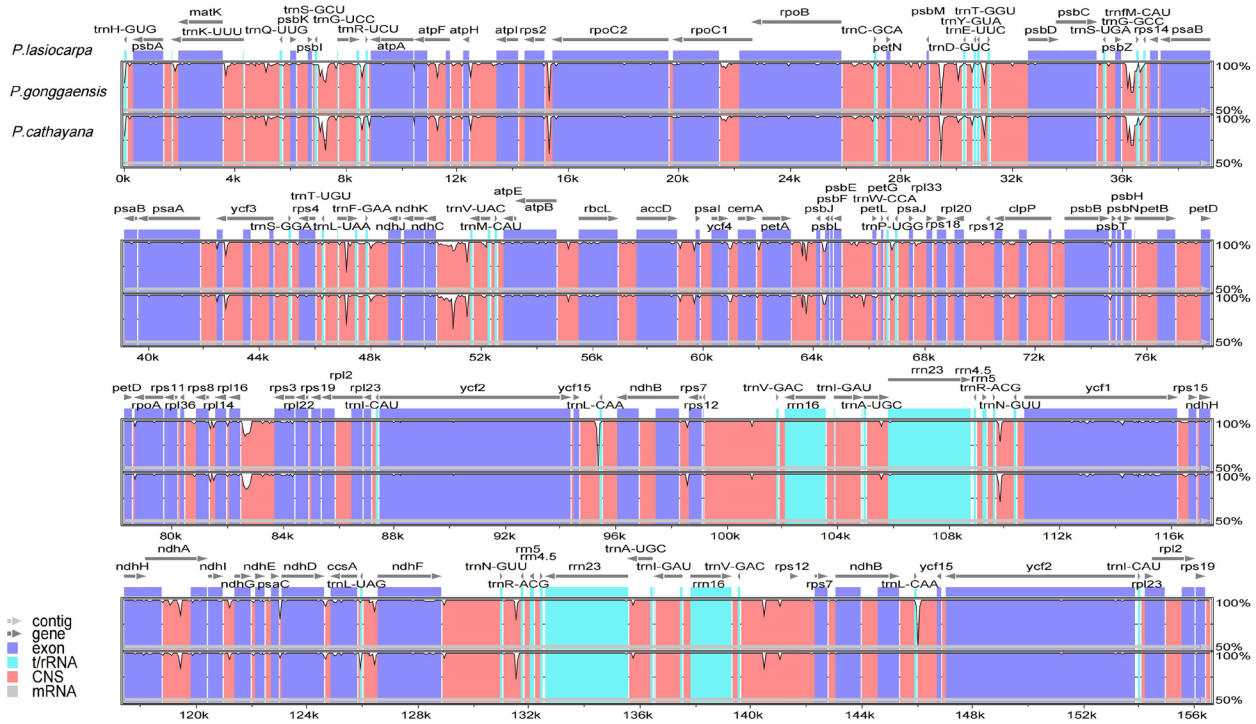

FIGURE 7: Comparative genomics of *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana*. The figure is divided into three rows, each representing a different *Populus* species, with *P. lasiocarpa* used as the reference genome. The linear genomic sequences are displayed horizontally, and the gene names are annotated (e.g., *matK*, *rpoC2*). Colour blocks, which are based on the legend (e.g., genes, exons), indicate the sequence conservation of elements such as genes, exons, conserved noncoding sequences (CNSs), and mRNAs. The percentages on the right represent the similarity thresholds; higher values indicate greater sequence similarity in the corresponding regions among the species.

constructed (Figure 10) using the chloroplast genomes of *P. lasiocarpa*, *P. gonggaensis*, *P. cathayana*, and 21 additional *Populus* species from various subgroups (Table S1). The topological structure presented phylogenetic relationships among *Populus* species from different groups. Among them, *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana* independently form a monophyletic clade. Additionally, *P. gonggaensis* and *P. cathayana* are assigned to one side of this clade, whereas *P. lasiocarpa* is assigned to the other side. Based on the genetic characteristics of the chloroplast genomes and our previous results of resequencing and phenotypic data, we speculated that *P. gonggaensis* may have originated from natural hybridisation between *P. lasiocarpa* and *P. cathayana*.

Discussion

Chloroplast genomes, which are maternally inherited and evolve more slowly than nuclear genomes, exhibit conserved structures and gene contents that make them valuable for phylogenetic inference (Palmer 1987; Carneiro Vieira et al. 2016). In this study, we sequenced and analysed the chloroplast genomes of three *Populus* species, namely, *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana*, revealing nucleotide polymorphism (Pi) values ranging from 0 to 0.1225 across 24 species of this genus. Although the LSC region exhibited minimal differentiation (average Pi <0.004) owing to the structural stability and functional constraints of the photosynthetic system, eight highly variable regions (Pi >0.11) were identified in the SSC region, including the *ndhA* gene; six intergenic spacers (*ndhA-ndhI*, *ndhI-ndhG*, *ndhD-ccsA*, *trnL-ndhF*, *ndhF-trnN*, *trnA-trnI*); and the *rrn23* gene. Notably, the *trnL-ndhF* region had the highest level of polymorphism (Pi = 0.1225) in our dataset, which was consistent with previous reports of high variability in this region (Djerbi et al. 2005; Pan et al. 2012). Comparative analyses with prior data revealed remarkable conservation. Zong, Gan, et al. (2019) reported chloroplast genome sizes of 156,789 bp (*P. cathayana*), 156,439 bp (*P. gonggaensis*), and 156,525 bp (*P. lasiocarpa*), all with identical GC contents (37%). Comparative genomic analyses with those of Zong, Gan, et al. (2019) demonstrated conserved quadripartite structures (156,812-156,554 bp; 37% GC) containing 130–131 genes (85–86 protein-coding genes, 37 tRNAs, and 8 rRNAs), corroborating the standard genomic framework (Table S6). Minor interstudy variations in genome size (<0.2%) likely reflect technical factors such as annotation discrepancies in repeat regions or potential low-frequency heteroplasmy within individuals (Z. Wu et al. 2015). Importantly, coding region alignments revealed 100% sequence identity between datasets, confirming that such technical variability does not compromise the phylogenetic

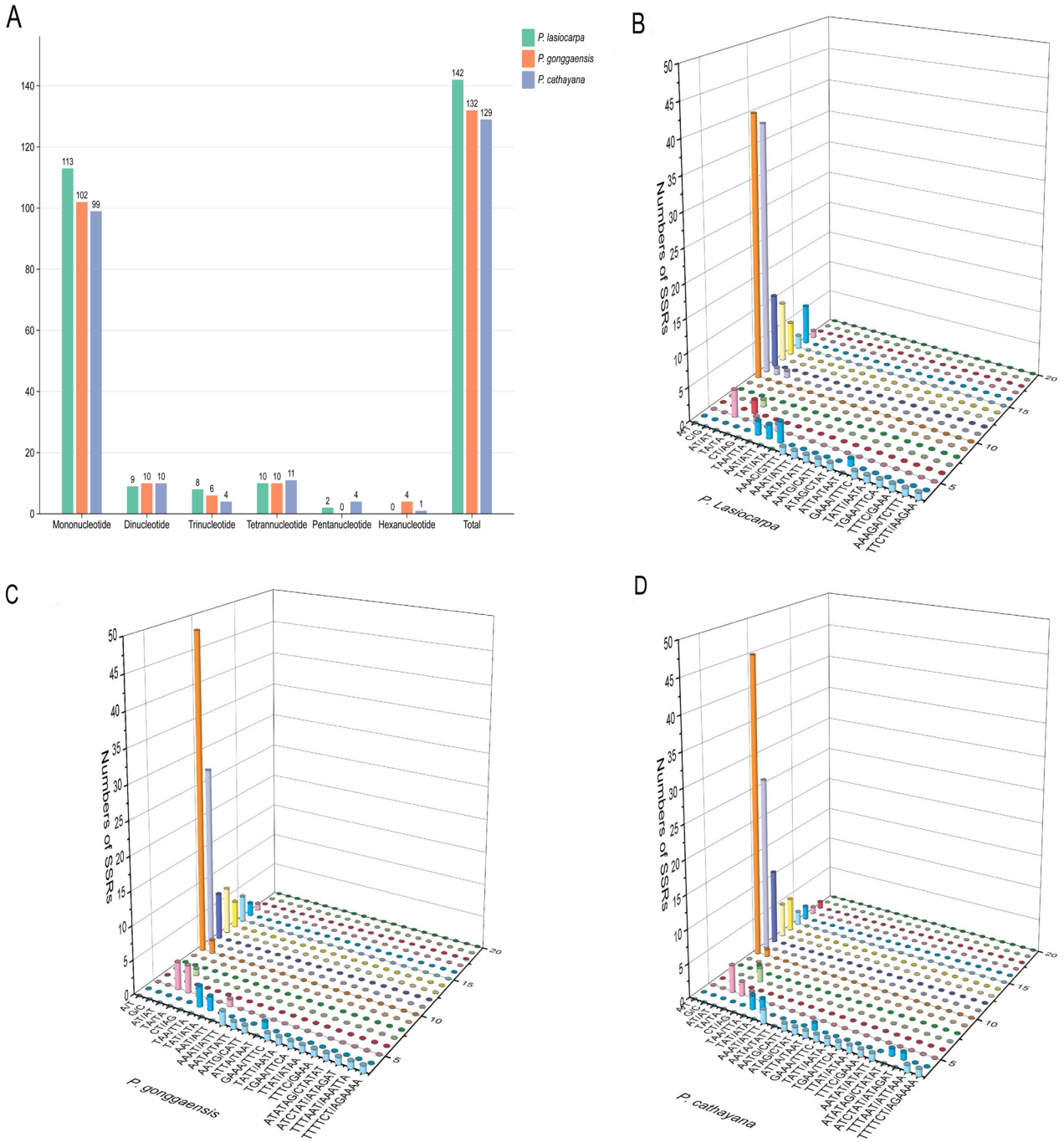

FIGURE 8: Chloroplast SSR analysis of *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana*. A: Total SSRs; B: SSR motif distribution in *P. lasiocarpa*; C: SSR motif distribution in *P. gonggaensis*; D: SSR motif distribution in *P. cathayana*. A summarises the numbers of mononucleotide to hexanucleotide repeat sequences and their total repeats (Total) across the three species. Different colours are used to distinguish the species, illustrating the variation in repeat sequence length distribution. B, C, and D depict the distribution of SSRs in *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana*. In these figures, the x-axis indicates the type of repeat sequence, the y-axis shows the number of repeats for each sequence type (e.g., a y-value of 8 for A/T bases signifies the presence of 8 A/T repeat SSRs), and the z-axis represents the total count of corresponding SSRs (e.g., a z value of 10 in the aforementioned case indicates that there are 10 SSRs with exactly 8 A/T repeats).

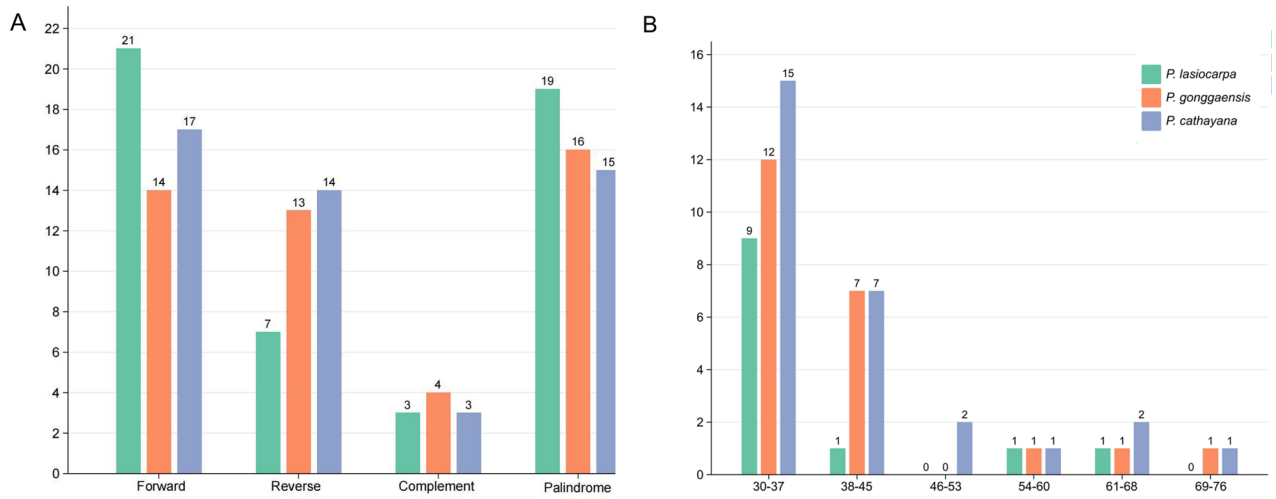

FIGURE 9: Comparison of long repeats in the chloroplast genomes of *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana*. A: Comparison of the distribution characteristics of four types of long repeat sequences (forward, reverse, complement, and palindrome) in the chloroplast genomes of *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana*. B: Analysis and visualisation of the distribution of long repeat sequences in the chloroplast genomes of *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana* based on repeat sequence length intervals (e.g., 30-37 bp, 38-45 bp). In the figure, different colours correspond to different species.

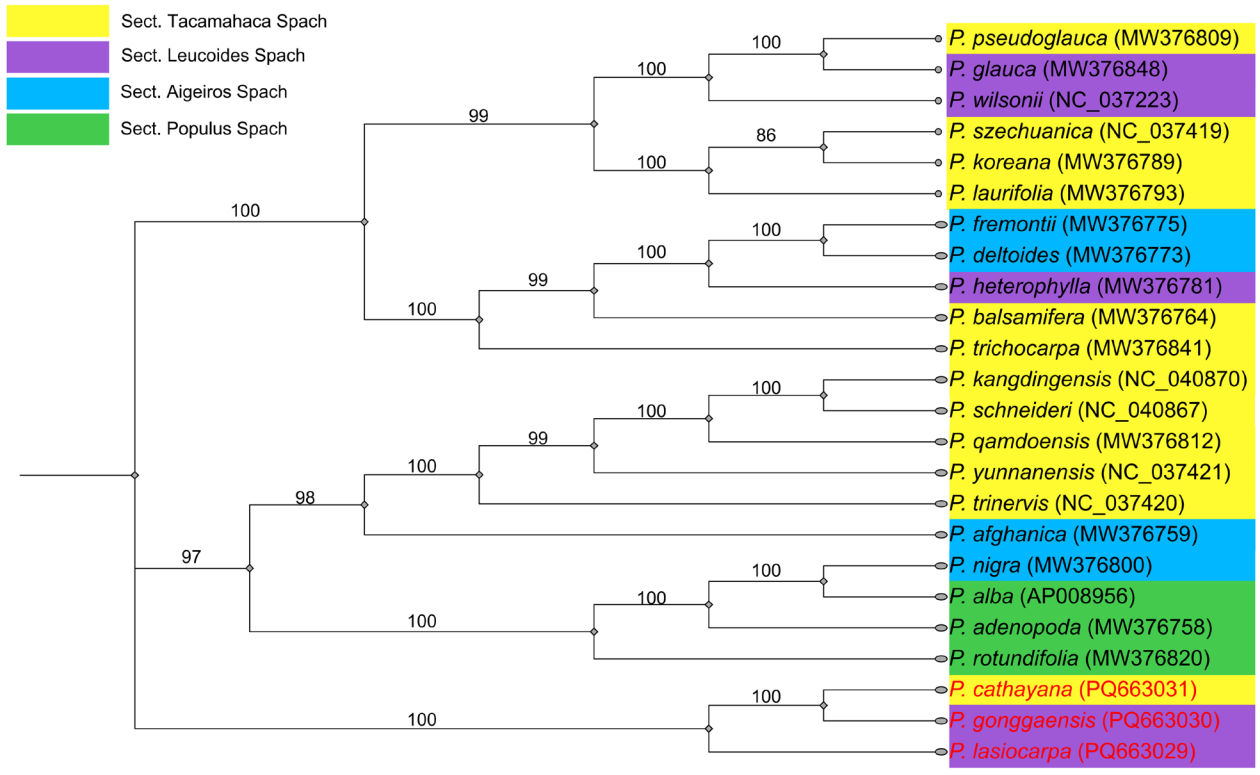

FIGURE 10: Phylogenetic analysis of the complete chloroplast genomes of 24 *Populus* species. Different colours correspond to four sections of the genus *Populus*. The values on the branches indicate bootstrap support rates; higher values suggest greater reliability of the results. This figure illustrates the phylogenetic relationships and taxonomic classifications among *Populus* species and elucidates the evolutionary relationships of the chloroplast genomes within this genus. All sequence information used in the phylogenetic analysis, including species names, GenBank accession numbers, and corresponding references, is provided in Table S1.

utility of these genomes. The observed conservation, particularly in coding regions, aligns with prior studies on Salicaceae chloroplast genomes, reflecting both the slow evolutionary rate of chloroplasts and functional constraints on the photosynthetic machinery (Djerbi et al. 2005; Pan et al. 2012).

This genomic stability underscored the critical role of chloroplast genomes in resolving plant phylogenies, as exemplified by their use in clarifying the phylogenetic placement of *Dendropanax morbifera* H.Lév (Araliaceae) and Saxifragales' evolutionary patterns (Dong et al. 2013; Kim et al. 2016). Our chloroplast genome data corroborated earlier nuclear SNP analyses and resequencing datasets, which proposed *P. gonggaensis* as a hybrid of *P. lasiocarpa* and *P. cathayana* (Du et al. 2024; Shi et al. 2024). Although Zong, Gan, et al. (2019) were the first to publish chloroplast sequences for these species, their study focused primarily on basic genomic characterisation without addressing phylogenetic implications or providing taxonomic validation, a limitation given the absence of voucher specimens and detailed morphological identification protocols. To ensure taxonomic precision, our samples were field-collected from the Hengduan Mountains using standardised phenotypic criteria (leaf morphology, bud structure, and trichome characteristics), with voucher samples deposited at Sichuan Agricultural University's Chongzhou R&D Base. This methodological rigor eliminated uncertainties regarding specimen identity that could affect cross-study comparisons. The convergence of maternal signals from chloroplast genomes strengthened the hypothesis of hybrid origin of *P. gonggaensis* and enhances our understanding of *Populus* speciation mechanisms.

Conclusions

Our study offered new insights into the chloroplast genomes of the three *Populus* species, and in combination with nuclear genomic data, supported the hybrid origin hypothesis of *P. gonggaensis* involving *P. lasiocarpa* and *P. cathayana*. These findings contributed to our understanding of *Populus* evolution and classification. However, to fully elucidate the evolutionary mechanisms driven by nuclear–cytoplasmic interactions, future research should integrate multiomics approaches, including mitochondrial genome analysis, to complement maternal inheritance patterns and reveal paternal signals alongside nuclear genomes and proteomes (Tam et al. 2010; Tang et al. 2023; Zhu et al. 2025). This integrated approach would further advance the exploration of *Populus* genetic diversity and may inform conservation and breeding strategies.

Competing interests

The authors declare that they have no competing interests.

Ethical approval

The authors declare that the collection of samples complies with the rules and regulations of relevant institutional, national guidelines and legislation, and they did not cause damage to the local environment.

Authors' contributions

HY and LZ designed the experiment and drafted the paper, JM, JQ, LC, FZ, XH, ZC, YG and FH collected samples and conducted experiments, LZ, SZ, XH, XL and JQ analysed the data. LZ and XW edited and revised the original draft. All authors participated in the revision of the paper and approved the final version.

Funding

This work was supported by Sichuan Province Forestry Breeding Research Project (2021YFYZ0032), University Student Innovation and Entrepreneurship Training Program (202410626053), and Research Interest Projects at SICAU (2025427X).

Data Availability

The data that support the findings of this study are openly available in GenBank of NCBI. The chloroplast genome of *Populus lasiocarpa*, *Populus gonggaensis* and *Populus cathayana* are available at GenBank under accession number PQ663029, PQ663030 and PQ663031. The associated BioProject, SRA, and BioSample numbers are PRJNA1192391, SRR31557152, SAMN45107138 (*P. lasiocarpa*), PRJNA1192394, SRR31557193, SAMN45107162 (*P. gonggaensis*), PRJNA1192399, SRR31557514 and SAMN45107214 (*P. cathayana*) respectively.

References

Amiryousefi A, Hyvonen J, Poczai P. 2018. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics, 34(17), 3030-3031. https://doi.org/10.1093/bioinformatics/bty220

Androsiuk P, Jastrzebski JP, Paukszto L, Makowczenko K, Okorski A, Pszczolkowska A, Chwedorzewska KJ, Gorecki R, Gielwanowska I. 2020. Evolutionary dynamics of the chloroplast genome sequences of six *Colobanthus* species. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-68563-5

Beier S, Thiel T, Münch T, Scholz U, Mascher M. 2017. MISA-web: a web server for microsatellite prediction. Bioinformatics, 33(16), 2583-2585. https://doi.org/10.1093/bioinformatics/btx198

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Bylesjö M, Segura V, Soolanayakanahally RY, Rae AM, Trygg J, Gustafsson P, Jansson S, Street NR. 2008. LAMINA: a tool for rapid quantification of leaf size and shape parameters. BMC Plant Biology, 8. https://doi.org/10.1186/1471-2229-8-82

Carneiro Vieira ML, Santini L, Diniz AL, Munhoz CdF. 2016. Microsatellite markers: what they mean and why they are so useful. Genetics and Molecular Biology, 39(3), 312-328. https://doi.org/10.1590/1678-4685-GMB-2016-0027

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics, 34(17), i884-i890. https://doi.org/10.1093/bioinformatics/bty560

Cho K-S, Yun B-K, Yoon Y-H, Hong S-Y, Mekapogu M, Kim K-H, Yang T-J. 2015. complete chloroplast genome sequence of tartary buckwheat (*Fagopyrum tataricum*) and comparative analysis with common buckwheat (*F. esculentum*). PLOS One, 10(5). https://doi.org/10.1371/journal.pone.0125332

Daniell H, Lin C-S, Yu M, Chang W-J. 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biology, 17(1), 134. https://doi.org/10.1186/s13059-016-1004-2

Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Research, 45(4).

Djerbi S, Lindskog M, Arvestad L, Sterky F, Teeri TT. 2005. The genome sequence of black cottonwood (*Populus trichocarpa*) reveals 18 conserved cellulose synthase (CesA) genes. Planta, 221(5), 739-746. https://doi.org/10.1007/s00425-005-1498-4

Dong WP, Xu C, Cheng T, Zhou SL. 2013. Complete chloroplast genome of *Sedum sarmentosum* and chloroplast genome evolution in *Saxifragales*. PLOS One, 8(10). https://doi.org/10.1371/journal.pone.0077965

Du W, Wang Y, Xie D, Li E, Bai Y, Shang C, Zhang Z. 2024. Phylogenomics reveal *Populus gonggaensis* as a hybrid between *P. lasiocarpa* and *P. cathayana* (*Salicaceae*). PhytoKeys (237), 161-177. https://doi.org/10.3897/phytokeys.237.103012

Eckenwalder JE. 1977. Systematics of *Populus* L. (Salicaceae) in Southwestern North America with special reference to Sect. Aigeiros Duby. (Dissertation/Thesis).

Eckenwalder JE. 1996. Systematics and evolution of *Populus*. In Settler RF, Bradshaw Jr HD, Heilman PE, & Hickley TM. (Eds.). Biology of *Populus* and its implications for management and conservation. National Research Council of Canada, Chapter 1. (pp. 7-32).

Ellis B, Jansson S, Strauss SH, Tuskan GA. 2010. Why and how *Populus* became a "Model Tree". In Jansson S, Bhalerao R Groover A (Eds.), Genetics and genomics of *Populus* (pp. 3-14). Springer New York, New York, NY. https://doi.org/10.1007/978-1-4419-1541-2_1

Fang ZF, Zhao SD, Skvortsov AK. 1999. "*Populus*" in flora of China. In Wu ZY & Raven PH (Eds.). Missouri Botanical Garden. Beijing: Science press, 162-274.

Grant JR, Enns E, Marinier E, Mandal A, Herman EK, Chen C-y, Graham M, Van Domselaar G, Stothard P. 2023. Proksee: in-depth characterization and visualization of bacterial genomes. Nucleic Acids Research, 51(W1), W484-W492. https://doi.org/10.1093/nar/gkad326

Han X-M, Wang Y-M, Liu Y-J. 2017. The complete chloroplast genome sequence of *Populus wilsonii* and its phylogenetic analysis. Mitochondrial DNA Part B-Resources, 2(2), 932-933. https://doi.org/10.1080/23802359.2017.1413291

Kim K, Lee SC, Yang TJ. 2016. The complete chloroplast genome sequence of *Dendropanax morbifera* (Leveille). Mitochondrial DNA Part A, 27(4), 2923-2924. https://doi.org/10.3109/19401736.2015.1060442

Korbik M, Kosinski P. 2023. Revisiting the taxonomy of *Populus lasiocarpa* x *P. wilsonii* hybrids. Dendrobiology, 90, 86-94. https://doi.org/10.12657/denbio.090.007

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Research, 29(22), 4633-4642. https://doi.org/10.1093/nar/29.22.4633

Letunic I, Bork P. 2024. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Research, 52(W1), W78-W82. https://doi.org/10.1093/nar/gkae268

Lin CP, Huang JP, Wu CS, Hsu CY, Chaw SM. 2010. Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. Genome Biology and Evolution, 2, 504-517. https://doi.org/10.1093/gbe/evq036

Liu SY, Zhang L, Sang YP, Lai Q, Zhang XX, Jia CF, Long ZQ, Wu JL, Ma T, Mao KS, Street NR, Ingvarsson PK, Liu JQ, Wang J. 2022. Demographic history and natural selection shape patterns of deleterious mutation load and barriers to introgression across *Populus* genome. Molecular Biology and Evolution, 39(2). https://doi.org/10.1093/molbev/msac008

Liu X, Sun C, Chen Z, Gong X. 2020. The complete chloroplast genome sequence of *Populus trinervis*, and a comparative analysis with other *Populus* species. Mitochondrial DNA Part B-Resources, 5(3), 2466-2468. https://doi.org/10.1080/23802359.2020.1775525

Lohse M, Drechsel O, Bock R. 2007. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Current Genetics, 52(5-6), 267-274. https://doi.org/10.1007/s00294-007-0161-y

Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, Pachter LS, Dubchak I. 2000. VISTA : visualizing global DNA sequence alignments of arbitrary length. Bioinformatics (Oxford, England), 16(11), 1046-1047. https://doi.org/10.1093/bioinformatics/16.11.1046

Nakamura T, Yamada KD, Tomii K, Katoh K. 2018. Parallelization of MAFFT for large-scale multiple sequence alignments. Bioinformatics, 34(14), 2490-2492. https://doi.org/10.1093/bioinformatics/bty121

Oh JH, Yun M, Park D, Ha IJ, Kim CK, Kim DW, Kim EO, Lee SG. 2019. *Papaver nudicaule* (Iceland poppy) alleviates lipopolysaccharide-induced inflammation through inactivating NF-B and STAT3. BMC Complementary and Alternative Medicine, 19. https://doi.org/10.1186/s12906-019-2497-5

Palmer JD. 1987. Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. The American Naturalist, 130, S6-S29. https://doi.org/10.1086/284689

Pan IC, Liao D-C, Wu F-H, Daniell H, Singh ND, Chang C, Shih M-C, Chan M-T, Lin C-S. 2012. Complete chloroplast genome sequence of an orchid model plant candidate: *Erycina pusilla* apply in tropical oncidium breeding. PLOS One, 7(4). https://doi.org/10.1371/journal.pone.0034738

Peden J. 1999. Analysis of codon usage: University of Nottingham. UK: Nottingham.

Racchi ML, Turchi A, Caparrini S, Camussi A. 2011. SSCP intron marker system is a convenient tool for clonal fingerprinting of poplar (*Populus*) cultivars of different species and interspecific hybrids. Genetic Resources and Crop Evolution, 58(4), 507-518. https://doi.org/10.1007/s10722-010-9594-0

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends in Genetics: TIG, 16(6), 276-277. https://doi.org/10.1016/S0168-9525(00)02024-2

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Molecular Biology and Evolution, 34(12), 3299-3302. https://doi.org/10.1093/molbev/msx248

Shi YJ, Huang JL, Mi JX, Li J, Meng FY, Zhong Y, He F, Tian FF, Zhang F, Chen LH, Yang HB, Hu HL, Wan XQ. 2024. A model of hybrid speciation process drawn from three new poplar species originating from distant hybridization between sections. Molecular Phylogenetics and Evolution, 190. https://doi.org/10.1016/j.ympev.2023.107966

Shin S, Kim S-C, Hong KN, Kang H, Lee J-W. 2019. The complete chloroplast genome of *Torreya nucifera* (*Taxaceae*) and phylogenetic analysis. Mitochondrial DNA Part B-Resources, 4(2), 2537-2538. https://doi.org/10.1080/23802359.2019.1640091

Tam PPC, Barrette-Ng IH, Simon DM, Tam MWC, Ang AL, Muench DG. 2010. The Puf family of RNA-binding proteins in plants: phylogeny, structural modeling, activity and subcellular localization. BMC Plant Biology, 10. https://doi.org/10.1186/1471-2229-10-44

Tamura K, Stecher G, Kumar S. 2021. MEGA11 Molecular evolutionary genetics analysis Version 11. Molecular Biology and Evolution, 38(7), 3022-3027. https://doi.org/10.1093/molbev/msab120

Tang DF, Huang SH, Quan CQ, Huang Y, Miao JH, Wei F. 2023. Mitochondrial genome characteristics and phylogenetic analysis of the medicinal and edible plant Mesona chinensis Benth. Frontiers in Genetics, 13. https://doi.org/10.3389/fgene.2022.1056389

Thiel T, Michalek W, Varshney R, Graner A. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theoretical and Applied Genetics, 106(3), 411-422. https://doi.org/10.1007/s00122-002-1031-0

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Research, 45(W1), W6-W11. https://doi.org/10.1093/nar/gkx391

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen G-L, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, dePamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé J-C, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai C-J, Uberbacher E, Unneberg P, Vahala J, Wall

K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science, 313(5793), 1596-1604. https://doi.org/10.1126/science.1128691

Wan X, Zhang F. 2013. An overview of *Populus* genetic resources in southwest China. Forestry Chronicle, 89(1), 79-87. https://doi.org/10.5558/tfc2013-013

Wang MC, Zhang L, Zhang ZY, Li MM, Wang DY, Zhang X, Xi ZX, Keefover-Ring K, Smart LB, DiFazio SP, Olson MS, Yin TM, Liu JQ, Ma T. 2020. Phylogenomics of the genus *Populus* reveals extensive interspecific gene flow and balancing selection. New Phytologist, 225(3), 1370-1382. https://doi.org/10.1111/nph.16215

Wang YC, Huang J, Li EZ, Xu SJ, Zhan ZF, Zhang XJ, Yang ZQ, Guo FY, Liu KJ, Liu D, Shen XL, Shang C, Zhang ZX. 2022. Phylogenomics and biogeography of *Populus* based on comprehensive sampling reveal deep-level relationships and multiple intercontinental dispersals. Frontiers in Plant Science, 13. https://doi.org/10.3389/fpls.2022.813177

Wu F-H, Chan M-T, Liao D-C, Hsu C-T, Lee Y-W, Daniell H, Duvall MR, Lin C-S. 2010. Complete chloroplast genome of *Oncidium* Gower Ramsey and evaluation of molecular markers for identification and breeding in *Oncidiinae*. BMC Plant Biology, 10. https://doi.org/10.1186/1471-2229-10-68

Wu Z, Tembrock LR, Ge S. 2015. Are differences in genomic data sets due to true biological variants or errors in genome assembly: an example from two chloroplast genomes. PLOS One, 10(2), e0118019. https://doi.org/10.1371/journal.pone.0118019

Yue M, Chen H, Xuan L, Yang Y, Chong X, Li M, Yu C, Lu X, Zhang F. 2023. Novel molecular markers for *Taxodium* breeding from the chloroplast genomes of four artificial *Taxodium* hybrids. Frontiers in Genetics, 14. https://doi.org/10.3389/fgene.2023.1193023

Zhang H, Zhang X, Wu GL, Dong CC, Liu JQ, Li MJ. 2023. Genomic divergence and introgression among three *Populus* species. Molecular Phylogenetics and Evolution, 180. https://doi.org/10.1016/j.ympev.2022.107686

Zhang L, Wang M, Ma T, Liu J. 2018. Taxonomic status of *Populus wulianensis* and *P. ningshanica* (*Salicaceae*). PhytoKeys (108), 117-129. https://doi.org/10.3897/phytokeys.108.25600

Zheng S, Poczai P, Hyvönen J, Tang J, Amiryousefi A. 2020. Chloroplot: An Online Program for the Versatile Plotting of Organelle Genomes. Frontiers in Genetics, 11. https://doi.org/10.3389/fgene.2020.576124

Zhou J, Chen X, Cui Y, Sun W, Li Y, Wang Y, Song J, Yao H. 2017. Molecular structure and phylogenetic analyses of complete chloroplast genomes of two *Aristolochia* medicinal species. International Journal of Molecular Sciences, 18(9). https://doi.org/10.3390/ijms18091839

Zhu W, Zhang D, Xu W, Gan Y, Huang J, Liu Y, Tan Y, Song Y, Xin P. 2025. Comparative genomics and phylogenetic analysis of mitochondrial genomes of *Neocinnamomum*. BMC Plant Biology, 25(1), 289. https://doi.org/10.1186/s12870-025-06238-x

Zong D, Gan P, Zhou A, Zhang Y, Zou X, Duan A, Song Y, He C. 2019. Plastome sequences help to resolve deep-level relationships of *Populus* in the family Salicaceae. Frontiers in Plant Science, 10. https://doi.org/10.3389/fpls.2019.00005

Zong D, Zhou A, Li D, He C. 2019. The complete chloroplast genome of *Populus xiangchengensis*, an endemic species in Southwest China. Mitochondrial DNA Part B, 4(1), 70-71. https://doi.org/10.1080/23802359.2018.1536463

SUPPLEMENTAL INFORMATION

TABLE S1: Information of *Populus* chloroplast genomes (section, GenBank accession numbers and references) used for phylogenetic analysis

Species	Section	Accession number	Reference
P. adenopoda	Sect. *Populus* Spach	MW376758	Wang et al. (2022)
P. afghanica	Sect. *Aigeiros* Spach	MW376759	Wang et al. (2022)
P. alba	Sect. *Populus* Spach	MW376764	Wang et al. (2022)
P. balsamifera	Sect. *Tacamahaca* Spach	AP008956	Wang et al. (2022)
P. deltoides	Sect. *Aigeiros* Spach	MW376773	Wang et al. (2022)
P. fremontii	Sect. *Aigeiros* Spach	MW376775	Wang et al. (2022)
P. glauca	Sect. *Leucoides* Spach	MW376848	Wang et al. (2022)
P. heterophylla	Sect. *Leucoides* Spach	MW376781	Wang et al. (2022)
P. kangdingensis	Sect. *Tacamahaca* Spach	NC_040870	Zong et al. (2019)
P. koreana	Sect. *Tacamahaca* Spach	MW376789	Wang et al. (2022)
P. laurifolia	Sect. *Tacamahaca* Spach	MW376793	Wang et al. (2022)
P. nigra	Sect. *Aigeiros* Spach	MW376800	Wang et al. (2022)
P. pseudoglauca	Sect. *Tacamahaca* Spach	MW376809	Wang et al. (2022)
P. qamdoensis	Sect. *Tacamahaca* Spach	MW376812	Wang et al. (2022)
P. rotundifolia	Sect. *Populus* Spach	MW376820	Wang et al. (2022)
P. schneideri	Sect. *Tacamahaca* Spach	NC_040867	Zong et al. (2019)
P. szechuanica	Sect. *Tacamahaca* Spach	NC_037419	Wang et al. (2022)
P. trichocarpa	Sect. *Tacamahaca* Spach	MW376841	Wang et al. (2022)
P. trinervis	Sect. *Tacamahaca* Spach	NC_037420	Liu et al. (2020)
P. wilsonii	Sect. *Leucoides* Spach	NC_037223	Han et al. (2017)
P. yunnanensis	Sect. *Tacamahaca* Spach	NC_037421	none

TABLE S2: List of genes in the chloroplast genome of *P. lasiocarpa*

Category	Gene group	Name of gene	Number
Self-replication	Proteins of the large ribosomal subunit	*rpl33,rpl20,rpl36,rpl4,rpl16,rpl22,rpl2*[ac],*rpl23*[c]	10
	Proteins of the small ribosomal subunit	*rps2,rps14,rps4,rps18,rps12*[bc],*rps11,rps8,rps3,rps19*[c],*rps7*[c],*rps15*	14
	Subunits of RNA polymerase	*rpoC2,ropC1*[a],*rpoB,rpoA*	4
	rRNAs	*rrn16*[c],*rrn23*[c],*rrn4.5*[c],*rrn5*[c]	8
	tRNAs	*trnH-GUG,trnK-UUU*[a],*trnQ-UUG,trnS-GCU,trnG-UCC*[a],*trnR-UCU,trnC-GCA,trnD-GUC,trnY-GUA,trnE-UUC,trnT-GGU,trnS-UGA,trnG-GCC,trnfM-CAU,trnS-GGA,trnT-UGU,trnL-UAA*[a],*trnF-GAA,trnV-UAC*[a],*trnM-CAU,trnW-CCA,trnP-UGG,trnI-CAU*[c],*trnL-CAA*[c],*trnV-GAC*[c],*trnI-GAU*[ac],*trnA-UGC*[ac],*trnR-ACG*[c],*trnN-GUU*[c],*trnL-UAG*	37
Photosynthesis	Subunits of photosynthesis I	*psaB,psaA,psaI,psaJ,psaC,ycf4,ycf15*[c]	8
	Subunits of photosynthesis II	*psbA*[a],*psbK,psbI,psbM,psbD,psbC,psbZ,psbJ,psbL,psbF,psbE,psbB,psbT,psbN,psbH*	15
	Subunits of NADH NADH-plastoquinone oxidoreductase	*petN,petA,petL,petG,petB*[a],*petD*	6
	Subunits of cytochrome b/f complex	*ndhJ,ndhK,ndhC,ndhB*[ac],*ndhH,ndhA*[a],*ndhI,ndhG,ndhE,ndhD,ndhF*	12
	Subunits of ATP synthase	*atpA,atpF*[a],*atpH,atpI,atpE,atpB*	6
	Large subunits of Rubisco	*rbcL*	1
Biosynthesis	Maturase	*matK*	1
	Protease	*clpP*[b]	1
	Envelope membrane protein	*cemA*	1
	Acetyl-CoA carboxylase	*accD*	1
	c-type cytochrome synthesis gene	*ccsA*	1
Unknown function	Conserved hypothetical chloroplast reading frames	*ycf3*[b],*ycf2*[c],*ycf1*	4

Note: [a]Genes containing two exons; [b]Genes containing three exons; [c]Two copies.

TABLE S3: List of genes in the chloroplast genome of *P. gonggaensis*

Category	Gene group	Name of gene	Number
Self-replication	Proteins of the large ribosomal subunit	*rpl33,rpl20,rpl36,rpl4,rpl16,rpl22,rpl2*[ac],*rpl23*[c]	10
	Proteins of the small ribosomal subunit	*rps2,rps14,rps4,rps18,rps12*[bc],*rps11,rps8,rps3,rps19*[c],*rps7*[c],*rps15*	14
	Subunits of RNA polymerase	*rpoC2,ropC1*[a],*rpoB,rpoA*	4
	rRNAs	*rrn16*[c],*rrn23*[c],*rrn4.5*[c],*rrn5*[c]	8
	tRNAs	*trnH-GUG,trnK-UUU*[a],*trnQ-UUG,trnS-GCU,trnG-UCC*[a],*trnR-UCU,trnC-GCA,trnD-GUC,trnY-GUA,trnE-UUC,trnT-GGU,trnS-UGA,trnG-GCC,trnfM-CAU,trnS-GGA,trnT-UGU,trnL-UAA*[a],*trnF-GAA,trnV-UAC*[a],*trnM-CAU,trnW-CCA,trnP-UGG,trnI-CAU*[c],*trnL-CAA*[c],*trnV-GAC*[c],*trnI-GAU*[ac],*trnA-UGC*[ac],*trnR-ACG*[c],*trnN-GUU*[c],*trnL-UAG*	37
Photosynthesis	Subunits of photosynthesis I	*psaB,psaA,psaI,psaJ,psaC,ycf4,ycf15*[c]	8
	Subunits of photosynthesis II	*psbA*[a],*psbK,psbI,psbM,psbD,psbC,psbZ,psbJ,psbL,psbF,psbE,psbB,psbT,psbN,psbH*	15
	Subunits of NADH NADH-plastoquinone oxidoreductase	*petN,petA,petL,petG,petB*[a],*petD*	6
	Subunits of cytochrome b/f complex	*ndhJ,ndhK,ndhC,ndhB*[bc],*ndhH,ndhA*[a],*ndhI,ndhG,ndhE,ndhD,ndhF*	12
	Subunits of ATP synthase	*atpA,atpF*[a],*atpH,atpI,atpE,atpB*	6
	Large subunits of Rubisco	*rbcL*	1
Biosynthesis	Maturase	*matK*	1
	Protease	*clpP*[b]	1
	Envelope membrane protein	*cemA*	1
	Acetyl-CoA carboxylase	*accD*	1
	c-type cytochrome synthesis gene	*ccsA*	1
Unknown function	Conserved hypothetical chloroplast reading frames	*ycf3*[b],*ycf2*[c],*ycf1*[c]	5

Note: [a]Genes containing two exons; [b]Genes containing three exons; [c]Two copies.

TABLE S4: List of genes in the chloroplast genome of *P. cathayana*

Category	Gene group	Name of gene	Number
Self-replication	Proteins of the large ribosomal subunit	*rpl33,rpl20,rpl36,rpl4,rpl16,rpl22,rpl2*[ac],*rpl23*[c]	10
	Proteins of the small ribosomal subunit	*rps2,rps14,rps4,rps18,rps12*[bc],*rps11,rps8,rps3,rps19*[c],*rps7*[c],*rps15*	14
	Subunits of RNA polymerase	*rpoC2,ropC1*[a],*rpoB,rpoA*	4
	rRNAs	*rrn16*[c],*rrn23*[c],*rrn4.5*[c],*rrn5*[c]	8
	tRNAs	*trnH-GUG,trnK-UUU*[a],*trnQ-UUG,trnS-GCU,trnG-UCC*[a],*trnR-UCU,trnC-GCA,trnD-GUC,trnY-GUA,trnE-UUC,trnT-GGU,trnS-UGA,trnG-GCC,trnfM-CAU,trnS-GGA,trnT-UGU,trnL-UAA*[a],*trnF-GAA,trnV-UAC*[a],*trnM-CAU,trnW-CCA,trnP-UGG,trnI-CAU*[c],*trnL-CAA*[c],*trnV-GAC*[c],*trnI-GAU*[ac],*trnA-UGC*[ac],*trnR-ACG*[c],*trnN-GUU*[c],*trnL-UAG*	37
Photosynthesis	Subunits of photosynthesis I	*psaB,psaA,psaI,psaJ,psaC,ycf4,ycf15*[c]	8
	Subunits of photosynthesis II	*psbA*[a],*psbK,psbI,psbM,psbD,psbC,psbZ,psbJ,psbL,psbF,psbE,psbB,psbT,psbN,psbH*	15
	Subunits of NADH NADH-plastoquinone oxidoreductase	*petN,petA,petL,petG,petB*[a],*petD*	6
	Subunits of cytochrome b/f complex	*ndhJ,ndhK,ndhC,ndhB*[bc],*ndhH,ndhA*[a],*ndhI,ndhG,ndhE,ndhD,ndhF*	12
	Subunits of ATP synthase	*atpA,atpF*[a],*atpH,atpI,atpE,atpB*	6
	Large subunits of Rubisco	*rbcL*	1
Biosynthesis	Maturase	*matK*	1
	Protease	*clpP*[b]	1
	Envelope membrane protein	*cemA*	1
	Acetyl-CoA carboxylase	*accD*	1
	c-type cytochrome synthesis gene	*ccsA*	1
Unknown function	Conserved hypothetical chloroplast reading frames	*ycf3*[b],*ycf2*[c],*ycf1*[c]	5

Note: [a]Genes containing two exons; [b]Genes containing three exons; [c]Two copies.

TABLE S5: The list of codon quantities for the chloroplast genomes of *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana*

Protein	Codon	*P. lasiocarpa*		*P. gonggaensis*		*P. cathayana*	
		Number	Total	Number	Total	Number	Total
Phe	UUU	2,342	3,725	2,364	3,752	2,386	3,785
	UUC	1,383		1,388		1,399	
Ser	UCU	1,208	4,803	1,256	4,823	1,229	4,936
	UCC	867		845		854	
	UCA	946		1,002		1,061	
	UCG	608		589		621	
	AGU	669		676		672	
	AGC	505		455		499	
Tyr	UAU	1,477	2,106	1,512	2,204	1,596	2,284
	UAC	629		692		688	
Cys	UGU	680	1,110	684	1,148	754	1,186
	UGC	430		464		432	
Leu	UUA	1,266	5,306	1,227	5,228	1,265	5,181
	UUG	1,054		1,052		967	
	CUU	1,100		1,090		1,070	
	CUC	646		605		616	
	CUA	766		775		810	
	CUG	474		479		453	
Trp	UGG	745	745	662	662	746	746
Pro	CCU	686	2,383	670	2,383	687	2,342
	CCC	563		583		575	
	CCA	748		758		701	
	CCG	386		372		379	
His	CAU	833	1,180	871	1,223	890	1,250
	CAC	347		352		360	
Arg	CGU	389	3,314	376	3,198	372	3,378
	CGC	250		244		248	
	CGA	587		571		631	
	CGG	399		386		399	
	AGA	1,119		1,050		1,078	
	AGG	570		571		650	
Gln	CAA	1,036	1,460	1,006	1,430	955	1,388
	CAG	424		424		433	
Ile	AUU	1,942	4,541	1,871	4,375	1,817	4,329
	AUC	1,061		1,003		1,012	
	AUA	1,538		1,501		1,500	
Thr	ACU	716	2,310	667	2,207	651	2,272
	ACC	534		512		565	
	ACA	711		691		719	
	ACG	349		337		337	
Asn	AAU	1,889	2,580	1,883	2,678	1,888	2,577
	AAC	691		795		689	

TABLE S5: *continued*

Protein	Codon	*P. lasiocarpa*		*P. gonggaensis*		*P. cathayana*	
		Number	**Total**	**Number**	**Total**	**Number**	**Total**
Met	AUG	871	871	859	859	817	817
Lys	AAA	2,265	3,279	2,221	3,219	2,285	3,256
	AAG	1,014		998		971	
Val	GUU	793	2,270	785	2,367	799	2,227
	GUC	434		447		404	
	GUA	649		715		635	
	GUG	394		420		389	
Ala	GCU	490	1,556	508	1,615	506	1,610
	GCC	350		379		361	
	GCA	457		462		497	
	GCG	259		266		246	
Asp	GAU	1,045	1,452	1,082	1,513	1,122	1,536
	GAC	407		431		414	
Gly	GGU	512	2,232	517	2,218	512	2,176
	GGC	349		336		351	
	GGA	805		818		773	
	GGG	566		547		540	
Glu	GAA	1,377	1,955	1,446	2,062	1,423	2,025
	GAG	578		616		602	
TER	UAA	1,223	3,006	1,205	3,000	1,209	2,969
	UGA	1,023		1,004		983	
	UAG	760		791		777	

Total: *P. lasiocarpa*: 52,184 codons; *P. gonggaensis*: 52,164 codons; *P. cathayana*: 52,270 codons.

TABLE S6: Comparison of chloroplast genome structural and compositional traits in *P. lasiocarpa*, *P. gonggaensis*, and *P. cathayana*: This study (Column a) vs. Zong et al. (2019) (Column b)

Characteristics	***P. lasiocarpa***		***P. gonggaensis***		***P. cathayana***	
	a	**b**	**a**	**b**	**a**	**b**
Length (bp)	156,554	156,525	156,494	156,439	156,812	156,789
GC content (%)	37	37	37	37	37	37
LSC (bp)	84,827	84,834	87,840	84,813	84,957	84,851
IR (bp)	27,620	27,620	27,570	27,570	27,657	27,672
SSC (bp)	16,487	16,451	16,514	16,513	16,541	16,594
Genes	130	130	131	130	131	130
CDS	85	85	86	85	86	85
tRNA	37	37	37	37	37	37
rRNA	8	8	8	8	8	8
GC content in LSC (%)	34.5	34.5	34.5	34.5	34.5	34.6
GC content in IR (%)	42.0	42.0	42.0	42.0	42.0	41.9
GC content in SSC (%)	30.5	30.6	30.7	30.7	30.7	30.6

Note: The reported GC content values were rounded to 37%, with the actual measured values being 36.7% for all three species.