BMC Plant Biology

RESEARCH

Open Access

Comparative chloroplast genomes and phylogenetic analyses shed new insights on the phyloevolution of different ploidy in *Camellia reticulata*

Zhifeng Fan[1], Xiaodan Xu[2], Qi Wang[2] and Wei Zheng[2*]

Abstract

Background *Camellia reticulata* Lindl. (*C. reticulata*) is the tallest ornamental camellia globally, with wild populations comprising a polyploid complex of diploids (2×), tetraploids (4×), and hexaploids (6×). The type specimen of *C. reticulata* is a heteroploid hexaploid derived from 2× ancestors, including *C. pitardii*, *C. saluenensis*, and 2× *C. reticulata*. Currently, limited information exists regarding the evolutionary characteristics of the chloroplast genomes of *C. reticulata* at different ploidy levels, and the phylogenetic position of 2× and 4× *C. reticulata* remains unclear.

Results This study sequenced, assembled, and annotated the chloroplast genomes of 2× and 4× *C. reticulata*, comparing them with those of 6× *C. reticulata* and other closely related species. The results indicated that the chloroplast genome sizes of *C. reticulata* ranged from 156,519 to 156,927 bp, with gene counts, distributions, GC content, and codon usage being similar across different ploidy levels. The *ycf*1 gene exhibited significant differentiation among species, and was identified as a candidate for adaptive evolution in *C. reticulata*. Additionally, 11 highly differentiated intergenic regions were identified, with six hotspots of variation that can serve as molecular markers for genetic studies in *C. reticulata* populations. Analysis of selection pressure indicated that four genes were under positive selection. Phylogenetic analysis revealed that the polyploid complex of *C. reticulata*, along with *C. pitardii*, *C. saluenensis*, and *C. mairei*, formed a well-supported clade. The genetic distances between 6× *C. reticulata* and its three 2× ancestors were relatively small.

Conclusion *Camellia pitardii*, *C. saluenensis*, and *C. mairei* may have participated in the allopolyploidization of *C. reticulata*, with both 2× and 4× *C. reticulata* have the potential for independent classification. These findings provide valuable insights into chloroplast genome alterations following allopolyploidization, establishing a crucial foundation for understanding the systematic evolutionary history of various ploidy levels in *C. reticulata*.

Keywords *Camellia reticulata*, Heteroploid hexaploid, Chloroplast genome, Phylogenetic relationship, Systematic evolution

*Correspondence:
Wei Zheng
57280234@qq.com
Full list of author information is available at the end of the article

Introduction

Polyploid genotypes can induce variations in the morphological, physiological, and molecular characteristics of plants [1]. *Camellia reticulata* Lindl. (*C. reticulata*), an evergreen tree of the Theaceae family, is naturally distributed in southwestern China [2] and has been cultivated for more than 1,300 years [3]. Recognized as the tallest ornamental camellia globally, *C. reticulata* is characterized by large, vibrant flowers with a long flowering period and diverse varieties [4]. In addition to its ornamental value, *C. reticulata* is also a notable oil plant [5]. The formal nomenclature of *C. reticulata* was established in 1827, with the type specimen being a semidouble cultivar introduced from Tengchong, Yunnan, by John Damper Parks in 1824 and later confirmed as a hexaploid (6×) [5]. Wild 6×*C. reticulata* was discovered in the 1940s [6]. Tetraploid (4×) and diploid (2×) *C. reticulata* were subsequently identified in the Jinshajiang Basin in 1994 and 1997, respectively [7, 8]. Genome in situ hybridization (GISH) studies revealed that 2×*C. reticulata, C. pitardii* (2×), and *C. saluenensis* (2×) are common ancestors of 6×*C. reticulata*. Initially, hybridization between 2×*C. reticulata* and *C. pitardii,* followed by polyploidization, produced a 4×*C. reticulata*. This tetraploid subsequently underwent an additional hybridization and polyploidization event with *C. saluenensis*, leading to the formation of 6×*C. reticulata* [9]. Consequently, the 2×, 4×, and 6×forms of *C. reticulata* constitute a polyploid complex (Fig. 1) [9, 10].

Chloroplasts are organelles that are essential for photosynthesis in green plants and play crucial roles in energy conversion in higher plants [11, 12]. In angiosperms, most chloroplast genomes are maternally inherited, displaying stable structures and low genetic recombination rates [13]. Owing to their slow evolutionary characteristics and conserved sequences, chloroplast genomes are extensively utilized for species identification and phylogenetic analysis [14, 15]. During allopolyploidization, not only do two distinct nuclear genomes merge, but chloroplast genomes from different origins also interact within the same cell [16]. Consequently, chloroplast genome variations can serve as valuable genetic markers

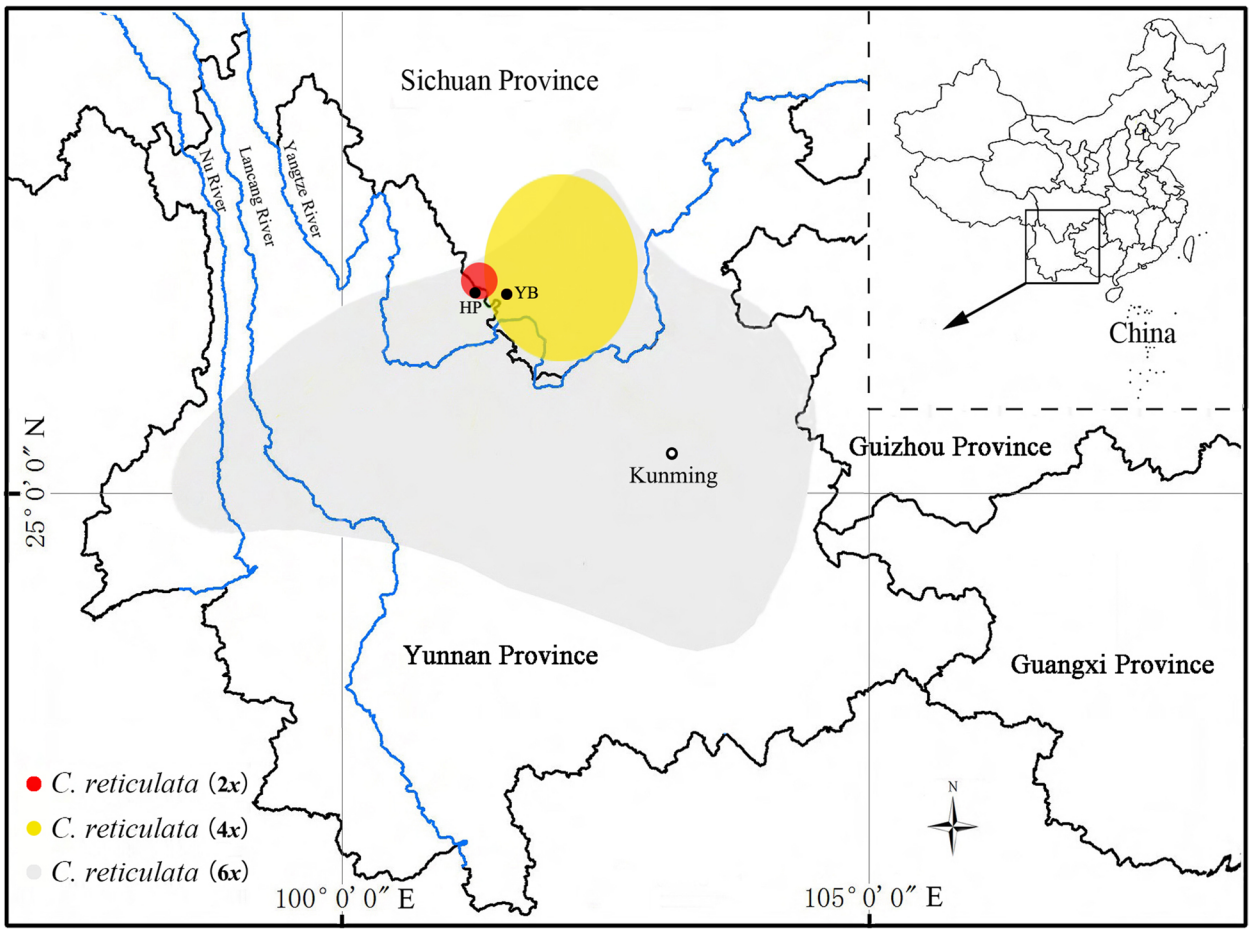

Fig. 1 Distribution of the *C. reticulata* polyploid complex. HP, Huaping; YB, Yanbian

for polyploid analyses [17], making them ideal models for studying reticulate evolution and comparative genomics [18]. While the chloroplast genome of 6×*C. reticulata* has been reported within the polyploid complex [19], information on the chloroplast genomes of 2×and 4×*C. reticulata* is lacking. As a result, the evolutionary characteristics of chloroplast genomes across different ploidy levels remain unclear.

In the context of allohexaploid species, their diploid progenitors are typically classified as distinct species [20, 21]. This raises the question of whether 2×*C. reticulata* should be similarly categorized as an independent species, analogous to its diploid counterparts, *C. pitardii* and *C. saluenensis*. Morphologically, 2×*C. reticulata* is a typical tree that is readily distinguishable from closely related diploid species. Notably, the fruit peel thickness of 2×*C. reticulata* ranges from 1.3 to 1.7 cm, significantly exceeding that of *C. saluenensis* and *C. pitardii* (0.1 to 0.3 cm) [7]. Chromosomal karyotype analyses revealed that 2×*C. reticulata* possesses a karyotype of 24 m + 4sm + 2st [7], distinct from *C. pitardii* (22 m + 6sm + 2st) and *C. saluenensis* (20 m + 8sm + 2st) [8]. Furthermore, in comparison with 4×and 6×*C. reticulata*, 2×*C. reticulata* exhibits the thickest fruit peel and hairy flower filaments [7]. These morphological and chromosomal data suggest that 2×*C. reticulata* may warrant consideration for independent classification.

The geographic distributions of 4×*C. reticulata* show both continuity and overlap with those of the diploid ancestors of 6×*C. reticulata* [8]. The meiotic chromosome pairing configuration in 4×*C. reticulata* is predominantly bivalent (30II), and the chromosome sets are heterozygous [8]. The heterologous nature of 4×*C. reticulata* contributes significantly to the evolution of polyploidy within the *C. reticulata* complex. Morphologically, 4×*C. reticulata* found at elevations between 1100 and 1800 m exhibited notable difference with 2×and 6×*C. reticulata* [8]. Furthermore, 4×*C. reticulata* demonstrates greater cold hardiness compared to both 2×and 6×*C. reticulata* [22]. Given that allopolyploid species are often recognized as distinct species, 4×*C. reticulata* may possess the potential for independent classification.

The evolutionary characteristics of the chloroplast genome across various ploidy levels of *C. reticulata* and the phylogenetic position of 2×and 4×*C. reticulata* remain unresolved. Therefore, this study aims to sequence and analyze the chloroplast genomes of 2×and 4×*C. reticulata*, comparing them with those of closely related species. This research endeavors to identify genetic variations influenced by allopolyploidization and elucidate their systematic position within the section (sect.) *Camellia*. This study aims to provide a theoretical foundation for understanding the origin and evolution of *C. reticulata* and to inform its conservation and utilization strategies.

Materials and methods

Experimental materials, DNA extraction, and sequencing

The diploid material of *C. reticulata* was collected from the Huaping population in Yunnan Province (collection number Zheng2409), whereas the tetraploid material was obtained from the Yanbian population in Sichuan Province (collection number Zheng2429), with reference to the sampling locations of the different ploidy populations as described by Xia et al. (1994) [7] and Gu (1997) [8]. Voucher specimens identified by Professor Wei Zheng were maintained in the Landscape Architecture, Ecology, and Plant Application Laboratory at Kunming University of Science and Technology (Figs. 1 and 2). Total genomic DNA was extracted from young leaves of both species via the TianGen DNA Extraction Kit (TianGen Biotechnology, Beijing, China). The DNA concentration was assessed via a Qubit 4.0 fluorometer (Life Invitrogen, USA), and agarose gel electrophoresis was performed to

Fig. 2 Diploid and tetraploid samples of *C. reticulata*. **A** diploid; **B**: tetraploid

evaluate DNA integrity and purity. The quality-verified DNA samples were submitted to Boyun Huakang Gene Technology Co., Ltd. (Beijing, China) for low-depth genome sequencing via the Illumina NovaSeq 6000 platform. Furthermore, chloroplast genome sequencing data for 6×*C. reticulata* (KY406793), *C. pitardii* (KF156837), and *C. saluenensis* (ON525353) were acquired from the NCBI GenBank database, and the annotation information of the raw data was rectified for subsequent analyses.

Chloroplast genome assembly and annotation

The chloroplast genomes were reconstructed via GetOrganelle [23]. The chloroplast genome of 6×*C. reticulata* (KY406793) served as a reference sequence, and gene annotation was conducted via Geneious R11 (https://www.geneious.com). Chloroplast genome maps were generated via OGDRAW v 1.3.158 [24]. The annotated sequences were subsequently deposited in GenBank (accession numbers PQ152957 for 2×*C. reticulata* and PQ152958 for 4×*C. reticulata*).

Identification of repetitive sequences

Long repetitive sequences in the chloroplast genomes of the five species (2×*C. reticulata*, 4×*C. reticulata*, 6×*C. reticulata*, *C. pitardii* and *C. saluenensis*) were identified via REPuter, which included forward, reverse, palindromic, and complementary repeats [25]. The detection parameters were as follows: repfind -c -f -p -r -l 30 -h 3—best 1000. Simple sequence repeats (SSRs) were detected via MISA [26], with minimum thresholds set at 10, 5, 4, 3, 3, and 3 for mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats, respectively.

Sequence variation analysis

To evaluate nucleotide substitutions and genetic distances among the five species, 1,000 bootstrap replicates were conducted via MEGA X [27]. mVISTA [28] was employed to visualize comparisons of the chloroplast genomes of three different ploidy levels of *C. reticulata* and its closely related species, utilizing 2×*C. reticulata* as the reference sequence within the shuffle-LAGAN model. The boundaries of four distinct regions in the chloroplast genomes of the five species were visualized via IRscope [29]. To examine nucleotide diversity (Pi), coding and noncoding regions were extracted and analyzed via the biological information cloud platform of Genepioneer Co., Ltd. (Nanjing, China, http://cloud.genepioneer.com), with a window length of 600 bp and a step length of 200 bp. Interspecific nucleotide diversity (Pi) for 2×, 4×and 6×*C. reticulata* were calculated using DnaSP6 [30]. Furthermore, single-nucleotide polymorphisms (SNPs) were identified from pairwise comparisons of species. Synteny analysis was performed on the chloroplast genomes of the five species via MUMmer software with 2×*C. reticulata* as the reference sequence, and the SNP sites were marked [31].

Selection pressure analysis

To assess evolutionary selection pressure on the 81 unique protein-coding genes across the five species, the ratio of the nonsynonymous substitution rate (Ka) to the synonymous substitution rate (Ks) was calculated via KaKs_Calculator 2.0 [32]. The MLWL model was employed for selection pressure analysis [33]. Multiple sequence alignment was performed using MAFFT, with the codon alignment mode selected as the parameter [34]. Genes exhibiting a Ka/Ks ratio < 1 (particularly < 0.5) were interpreted as undergoing purifying selection, whereas a Ka/Ks ratio > 1 indicated positive selection. A ratio of approximately 1 suggests neutral evolution. In instances where Ks = 0, the Ka/Ks value was denoted as NA, signifying minimal nonsynonymous site substitutions, and was consequently excluded from the analysis. The sequence alignment was converted to the fas-clustal format via ALTER for compatibility in downstream analysis. A Ka/Ks ratio heatmap was then generated using HemI software to identify positively selected genes among species pairs [35].

Codon usage

To accurately determine the effective number of codons [36], all coding sequences (CDSs) shorter than 300 bp were excluded to minimize sample bias. Consequently, 50 out of the 81 CDSs were utilized for codon usage analysis. CodonW v.1.4.2 was employed to calculate relative synonymous codon usage (RSCU) rates [37], with the results visualized via R software. The RSCU represents the ratio of the observed frequency of a specific codon to its expected frequency. An RSCU value exceeding 1 indicates higher-than-expected codon usage, whereas a value below 1 signifies lower-than-expected usage.

Phylogenetic analysis

To investigate the phylogenetic relationships within the polyploid complex of *C. reticulata* in sect. *Camellia*, chloroplast genome data for 14 subg. *Camellia* species were obtained from GenBank (Table S1). This dataset encompassed all 12 species in sect. *Camellia*, one species from sect. *Stereocarpus*, and one from sect. *Paracamellia*. *Polyspora axillaris* and *P. hainanensis* were selected as outgroups. In total, the phylogenetic analysis incorporated 18 species, including the newly sequenced diploid (2×) and tetraploid (4×) *C. reticulata*.

Phylogenetic trees were constructed via the ML and BI methods on the PhyloSuite platform [38]. The optimal nucleotide substitution model, GTR + I + G, was selected

via the Akaike information criterion (AIC) method implemented in Modelfinder [39]. For the ML tree, 1,000 bootstrap replicates were performed, whereas the BI tree utilized 1 million generations with sampling every 100 generations. The initial 25% of the trees were discarded, and a 50% majority rule consensus tree was constructed to provide posterior probabilities. The resulting tree was visualized and refined via Figtree (http://tree.bio.ed.ac.uk/), with subsequent integration and enhancement via Adobe Illustrator CS6 software.

Results

Structure and characteristics of the chloroplast genome of *Camellia reticulata*

Illumina sequencing results demonstrated that the 2×*C. reticulata* exhibited 24,887,576 reads (2.43 GB, average read length of 149.5 bp), while the 4×*C. reticulata* yielded 16,680,876 reads (2.18 GB, average read length of 149.1 bp). The two ploidy levels of *C. reticulata* achieved organelle base coverages of 293×and 540×, respectively. The chloroplast genome lengths of the five species ranged from 156,519 bp for 4×*C. reticulata* to 156,927 bp for 2×*C. reticulata* (Fig. 3). This finding suggests that genome length is not directly correlated with ploidy or parental lineage. All the species presented chloroplast genomes with a characteristic quadripartite structure comprising a large single-copy region (LSC, 86,212–86,670 bp), a small single-copy region (17,850–18,277 bp), and a pair of inverted repeat regions (IR, 26,023–26,048 bp). The small single-copy region of 4×*C. reticulata* was the shortest, while the other regions presented negligible length differences among the five species. The GC content of the chloroplast genomes was consistently 37.3% across all the species. The 2×*C. reticulata* genome contained 132 genes, including 87 protein-coding genes, 37 tRNA genes, and 8 rRNA genes (Table S2). The *clp*P gene was absent in 4×*C. reticulata*, whereas other gene types and counts were consistent with those in 2×*C. reticulata*. In the IR region, 18 genes were duplicated, comprising 7 tRNA genes, 4 rRNA genes, and 7 protein-coding genes. Thirteen genes contained one intron, including 7 protein-coding genes (*atp*F, *ndh*A, *ndh*B, *rpl*2, *rps*12, *rps*16, and *rpo*C1) and 6 tRNA genes (*trn*A-UGC, *trn*G, *trn*I-GAU, *trn*K-UUU, *trn*L-UAA, and *trn*V-UAC), whereas two protein-coding genes (*clp*P and *ycf*3) possessed two introns (Table S3).

Expansion and contraction of inverted repeat regions

Figure 4 depicts the expansion and contraction of the IR regions among *C. reticulata* and its closely related species. Throughout the evolutionary process of chloroplast genomes, variations among the five species have been minimal. The length of the IR regions (26,023–26,048 bp) and their connection to the single-copy regions were highly conserved. The *rps*19 gene maintained a consistent position across all species, spanning 233 bp in the LSC region and 46 bp in the IRb region, whereas the *rpl*12 gene was situated entirely within the IRb region. In both 2×and 4×*C. reticulata*, the *ndhF* gene spans the IRb/SSC boundary, occupying 23 bp within the IRb region in the diploid and 24 bp in the tetraploid. Conversely, the *ndh*F genes of the remaining

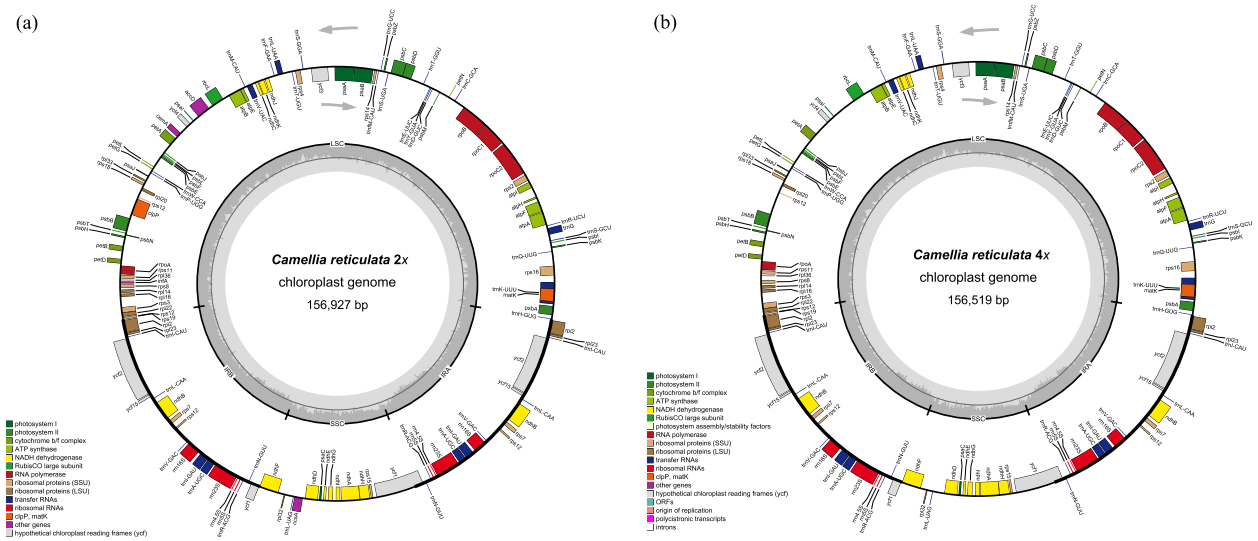

Fig. 3 Chloroplast genome map of 2×and 4×*C. reticulata*. Genes located on the outer edge of the circle are transcribed in a counterclockwise direction, whereas those situated within the circle are transcribed clockwise. Genes associated with distinct functional groups are represented in different colors. The dark gray shading in the inner circle indicates the guanine–cytosine content

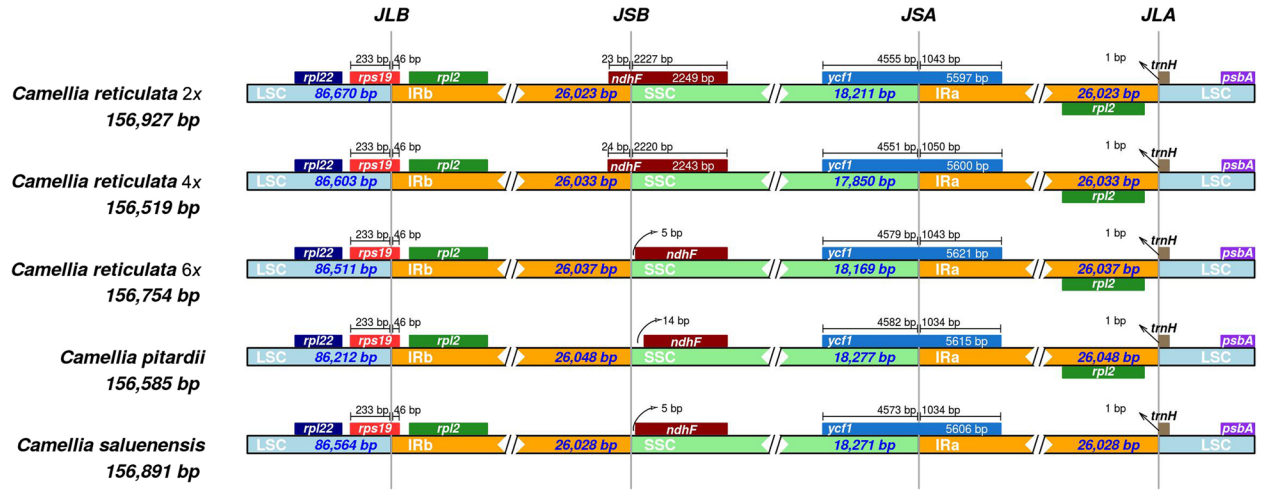

Fig. 4 Comparative analysis of the border regions in the chloroplast genomes of *C. reticulata* and its related species. JLB (IRb/LSC), JSB (IRb/SSC), JSA (SSC/IRa), and JLA (IRa/LSC) indicate the junction sites between each corresponding region in the genome

three species were located within the SSC region. The *ndh*F genes of 6×*C. reticulata* and *C. saluenensis* were located 5 bp from the IRb/SSC junction, whereas that of *C. pitardii* was 14 bp away. The *ycf*1 gene in all species extended across the boundaries of SSC and IRa, exhibiting variations in gene length among the different species. The *trn*H gene was consistently located in the LSC region across all species, just 1 bp from the IRa/LSC boundary. Overall, the chloroplast genomes of *C. reticulata* presented relatively high structural stability at the IR and SC boundaries, indicating that minor alterations in boundary genes may reflect evolutionary adaptations of these species' chloroplast genomes.

Analysis of long repetitive sequences

A comprehensive analysis using REPuter identified 208 long repetitive sequences in the chloroplast genomes of the five species studied (Table S4). Forward and palindromic repeats were observed in all species, whereas reverse repeats were present only in 2×*C. reticulata* and *C. saluenensis*. Complementary repeat sequences were unique to *C. saluenensis* (Fig. 5). The number of forward repeats varied from 15 to 19, with minimal variation in

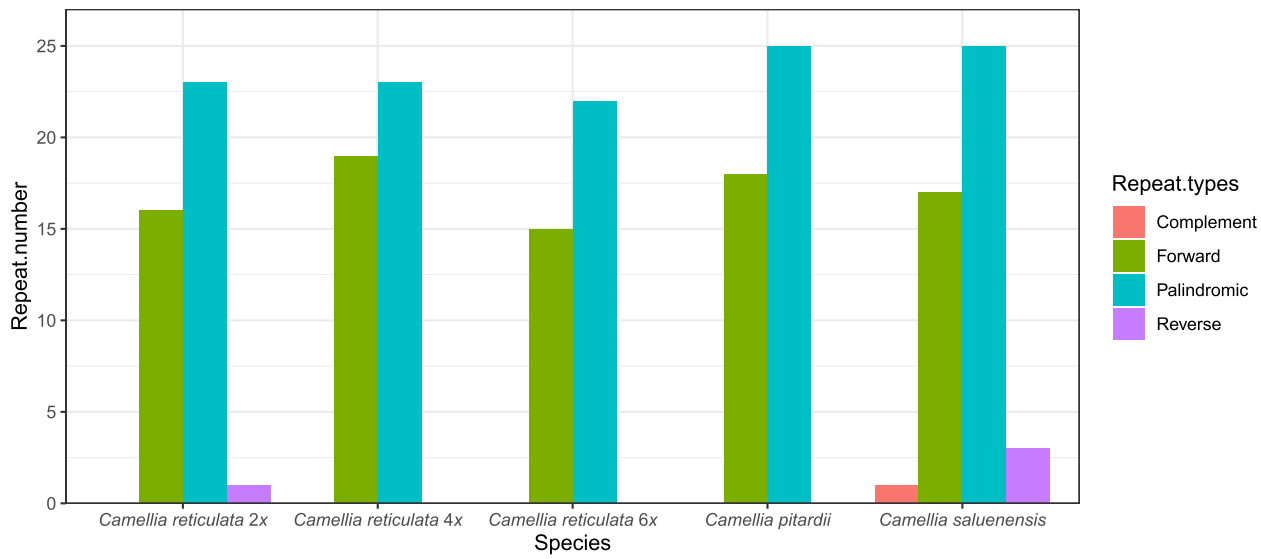

Fig. 5 Quantity of extended repetitive sequences in the chloroplast genomic DNA of *C. reticulata* and its associated species

palindromic repeats across species (Fig. 5, Table S4). Among the five species, 6×*C. reticulata* presented the lowest number of repeats, comprising 15 forward and 22 palindromic repeats (Fig. 5, Table S4). The lengths of repetitive sequences, excluding the IR regions, ranged from 30 to 72 bp, with 30, 42, and 64 bp being the most prevalent lengths (Fig. 6, Table S5). The majority of long repetitive sequences were located in the IR region (Table S5), with the *ycf*2 gene containing the longest repetitive sequence at 64 bp.

SSR analysis

The number of SSRs in the five species ranged from 70 to 75 (Table S6). The SSR count remained consistent across all three ploidy levels, with each level containing 70 SSRs. Five types of SSRs were identified: mononucleotide, dinucleotide, trinucleotide, tetranucleotide, and hexanucleotide. Among *C. reticulata* and its closely related species, mononucleotide repeats occurred most frequently (51–56 times), whereas all five species presented a single occurrence of trinucleotide repeats, with *C. pitardii* and *C. saluenensis* each containing two occurrences of pentanucleotide repeats (Fig. 7). A/T mononucleotide repeats significantly outnumbered the total number of occurrences of the other three repeat types. In each species, the AAG/CTT trinucleotide repeat and the ACAG/CTGT tetranucleotide repeat were each found only once; AGAT/ATCT and AGGG/CCCT tetranucleotide repeats occurred with the same frequency across all species, repeated three times and twice, respectively; and aligned with the patterns observed in most angiosperms [40]. Additionally, a unique six-base SSR (AAATTC/AATTTG) hexanucleotide repeat was present only once in *C. pitardii* (Fig. 8, Table S7). SSRs were predominantly distributed in intergenic regions, with a few occurring in the coding genes *atp*A, *rpo*C2, *rpo*B, *atp*B, *rpo*A, *ycf*2, *ccs*A, *ndh*D, *ndh*G, *ycf*1, and *ycf*2. Specifically, the genes *rpo*C2, *ndh*D, and *ycf*2 each contained two SSRs, whereas the *ycf*1 gene possessed five SSRs (Table S8).

Sequence variation analysis

To assess the variations in the chloroplast genomes of *C. reticulata* and its closely related species, nucleotide substitutions, and genetic distances were compared across the five species, utilizing the annotated 2×*C. reticulata* chloroplast genome as the reference for mVISTA analysis. The nucleotide difference values among the five species ranged from 657 to 1,446, with genetic distances ranging from 0.000582 to 0.001308 (Table 1). The chloroplast genome sequences across the five species presented relative similarity, with sequence variation primarily concentrated in noncoding regions. Conversely, the exons and coding regions demonstrated minimal differences. The *ycf*1 gene and 11 intergenic regions (*trn*H-*psb*A, *rpo*C1-*rpo*B, *trn*C-*pet*N, *trn*F-*trn*L, *ycf*4-*cem*A, *pet*L-*psa*J, *pet*D-*rpo*A, *trn*R-*ndh*F, *rpl*32-*ccs*A, *ndh*G-*ndh*I, and *trn*L-*ycf*15) presented a high degree of differentiation (Fig. 9).

In comparisons between 4×*C. reticulata* and its 2×progenitor species, 91 SNPs and 74 insertions/deletions (Indels) were detected. Comparisons between 4×*C. reticulata* and *C. pitardii* revealed 202 SNPs and 78 Indels (Table 2). When comparing 6×*C. reticulata*

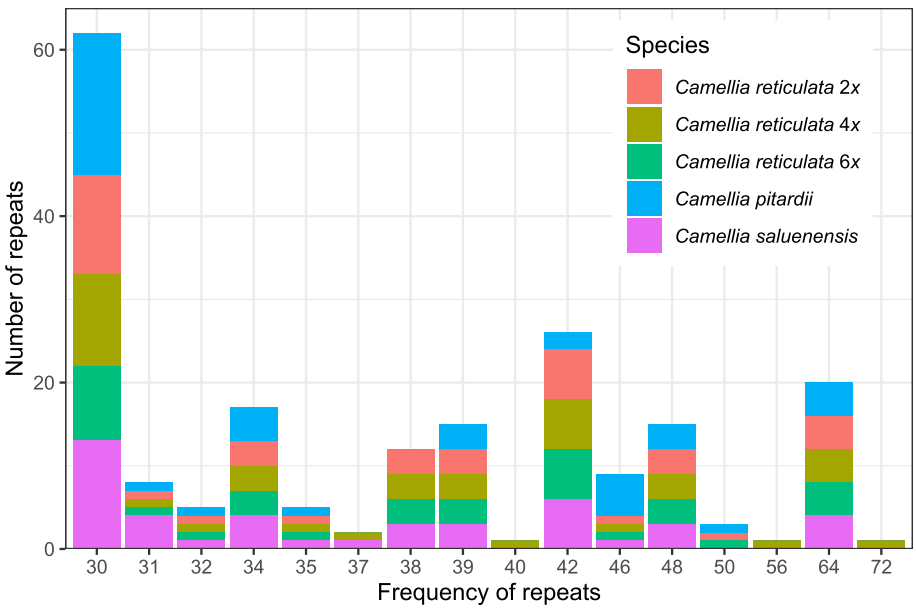

Fig. 6 Frequency distribution of repeats exceeding 30 bp in length

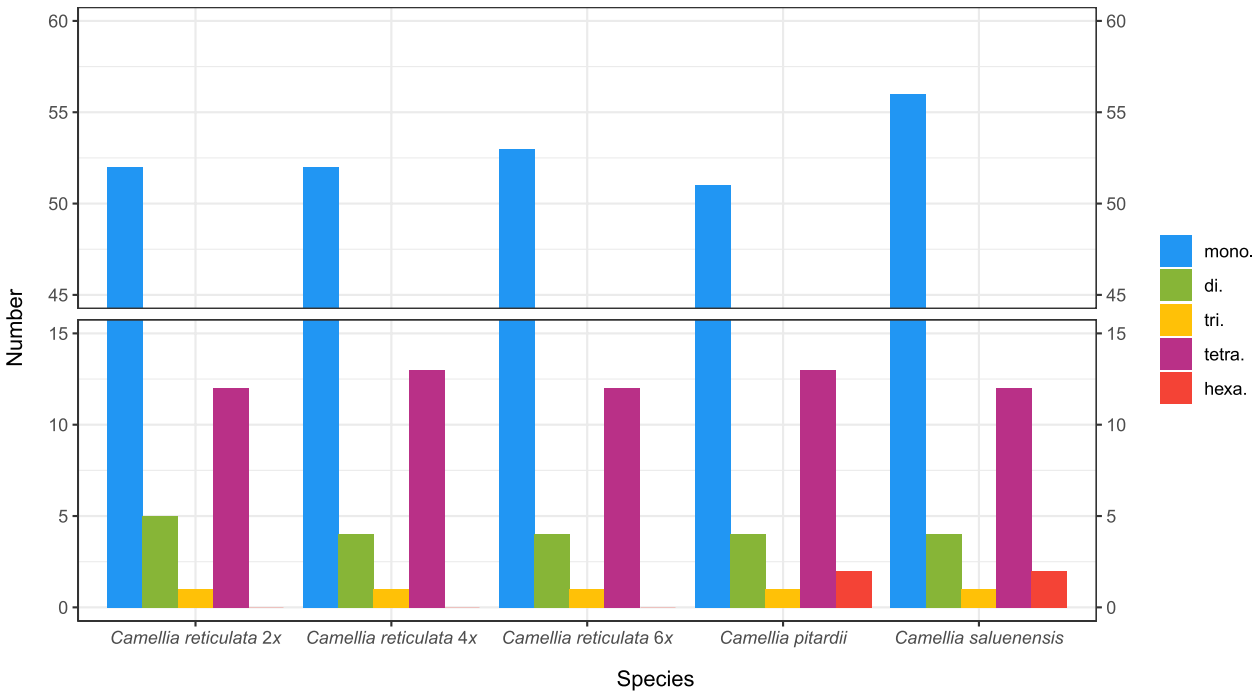

Fig. 7 Comparison of SSR distributions in the chloroplast genomes of *C. retculata* and related species

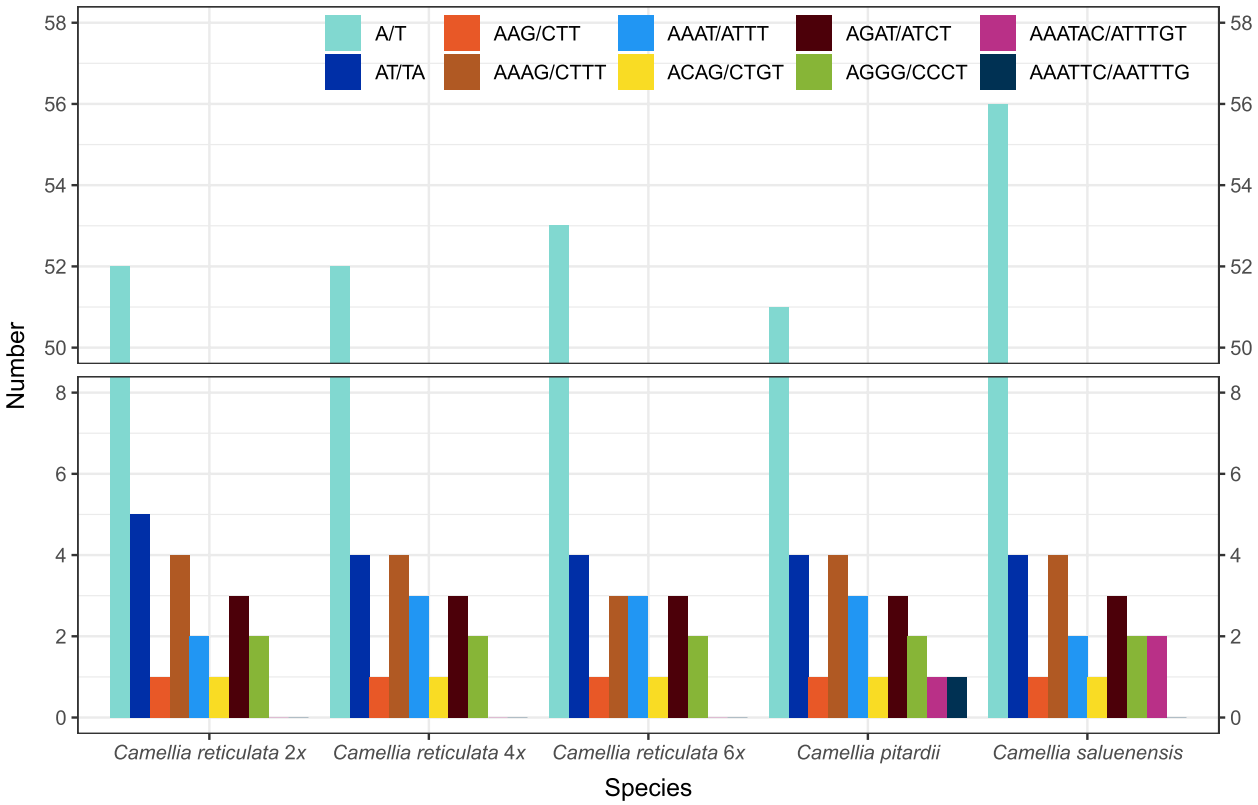

Fig. 8 Distribution of various SSR types identified in the chloroplast genome of *C. reticulata* and related species

Table 1 Numbers of nucleotide substitutions and genetic distance in five complete chloroplast genomes

Species	*C. reticulata* 2×	*C. reticulata* 4×	*C. reticulata* 6×	*C. pitardii*	*C. saluenensis*
C. reticulata 2×		0.000582	0.000728	0.001293	0.000977
C. reticulata 4×	875		0.000755	0.001308	0.001005
C. reticulata 6×	657	853		0.001223	0.000939
C. pitardii	1174	1446	1132		0.001056
C. saluenensis	717	1057	726	1119	

Note: The lower left triangle displays the quantity of nucleotide substitutions, whereas the upper right triangular region indicates the genetic distance

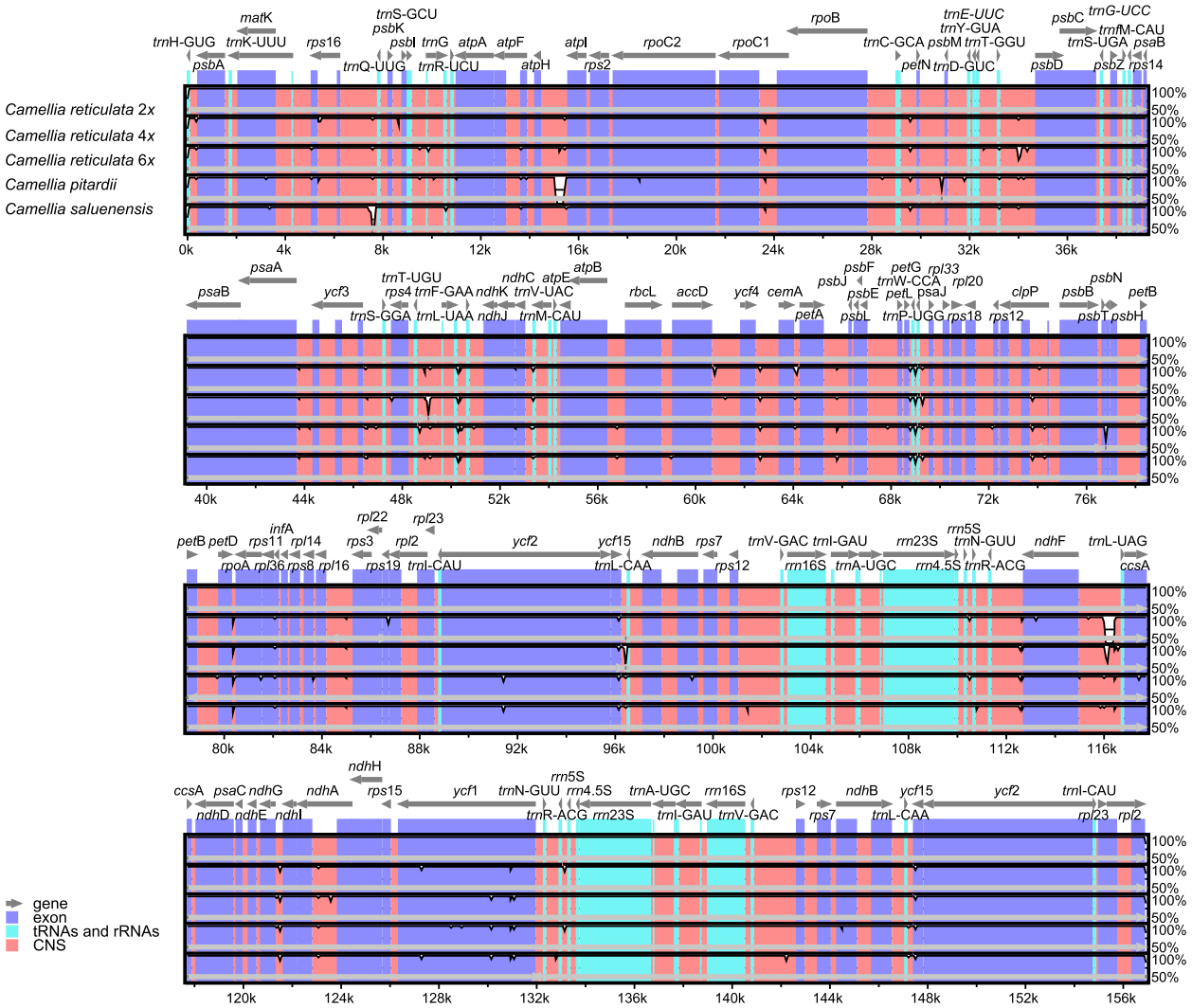

Fig. 9 Alignment of chloroplast genomes in mVISTA

with its progenitor species, 117 SNPs and 61 Indels were observed between 6×*C. reticulata* and 4×*C. reticulata*, whereas 144 SNPs and 70 Indels were identified between 6×*C. reticulata* and *C. saluenensis* (Table 2). Among the five species, a total of 350 SNPs and 156 Indels were identified. The locations of the SNP sites were presented in Tables S9–S18 and were illustrated in the synteny analysis diagram (Fig. S1).

Table 2 SNP and Indel analysis of the chloroplast genomes of the *C. reticulata* polyploid complex, *C. pitardii*, and *C. saluenensis*

Interspecies	SNP	Indel
C. reticulata 4× vs. *C. reticulata* 2×	91	74
C. reticulata 4× vs. *C. pitardii*	202	78
C. reticulata 6× vs. *C. reticulata* 4×	117	61
C. reticulata 6× vs. *C. saluenensis*	144	70
Among five species	350	156

Polymorphic hotspot analysis

Sliding window analysis of the chloroplast genome sequences of *C. reticulata* identified 350 mutation sites, with Pi values ranging from 0.00067 to 0.00767 and an average value of 0.00098. Six regions (*mat*K, *trn*S-*trn*G, *trn*G-*trn*R, *pet*N-*psb*M, *trn*P, and *trn*N-*ndh*F) demonstrated high variability (Fig. 10). Four of these regions were situated within the LSC region, whereas the remaining two were in the SSC region. Only one highly variable gene (*mat*K) was found in coding regions, whereas the other highly variable segments occurred in noncoding regions. Compared with the IR regions, the LSC and SSC regions presented greater levels of differentiation, suggesting that protein-coding genes are more conserved than noncoding genes. The *trn*N-*ndh*F region had the highest differentiation value at 0.00767 (Fig. 10). The interspecific nucleotide diversity of different ploidy *C. reticulata* demonstrated that the Pi value between 4× and 6× *C. reticulata* was 0.00075, which was higher than that between 2× and 4× *C. reticulata* (Pi = 0.00058).

Selection pressure analysis

The ratios of nonsynonymous (Ka) to synonymous (Ks) substitutions were calculated for 81 protein-coding genes. Across the five chloroplast genomes, the average Ka/Ks ratio for these genes was 0.3186 (Table S19). Significant evolutionary rate variations were observed among the five species, with 26 of the 81 protein genes exhibiting positive selection sites. Four genes (*mat*K, *ndh*B, *rpo*C1, and *ycf*1) presented Ka/Ks ratios exceeding 1, suggesting that they may be under positive selection (Fig. 11, Table S20). Notably, the Ka/Ks ratio for the *ycf*1 gene surpassed 1.5 in three pairwise comparisons (2× *C. reticulata*/4× *C. reticulata*, 4× *C. reticulata*/6× *C. reticulata*, 6× *C. reticulata*/*C. saluenensis*), indicating that *ycf*1 is a potential candidate gene for adaptive evolution. The majority of the remaining genes presented Ka/Ks ratios less than 0.5, suggesting that most genes were subject to strong purifying selection.

Further analysis of the Ka/Ks ratios in the chloroplast genome of 2× *C. reticulata* in relation to its closely related species revealed notable findings (Fig. 12, Table S21). The results indicated that 19 out of the 81 protein-coding genes presented positive selection sites, with three genes (*mat*K, *ndh*B, and *ycf*1) displaying Ka/Ks ratios exceeding 1. Notably, the Ka/Ks ratio for the *ycf*1 gene between 2× and 4× *C. reticulata* reached 1.85.

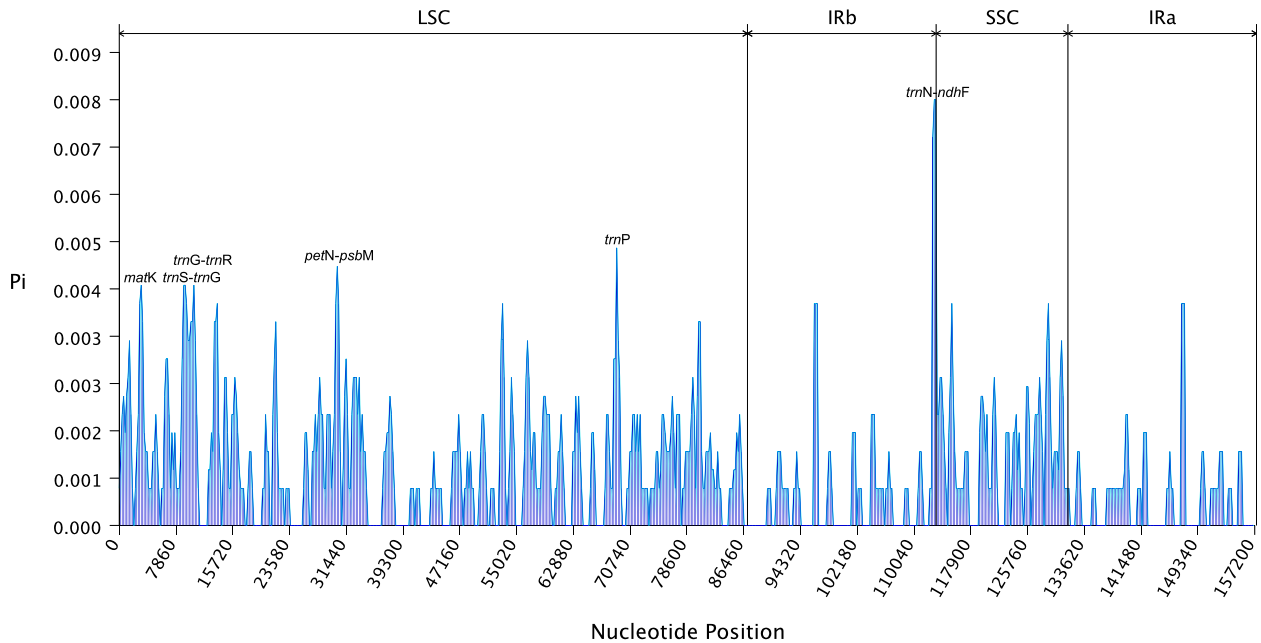

Fig. 10 Nucleotide diversity values among the five *Camellia* species. X-axis, position of the midpoint of a window; Y-axis, nucleotide diversity of each window. (window length: 600 bp, step size: 200 bp)

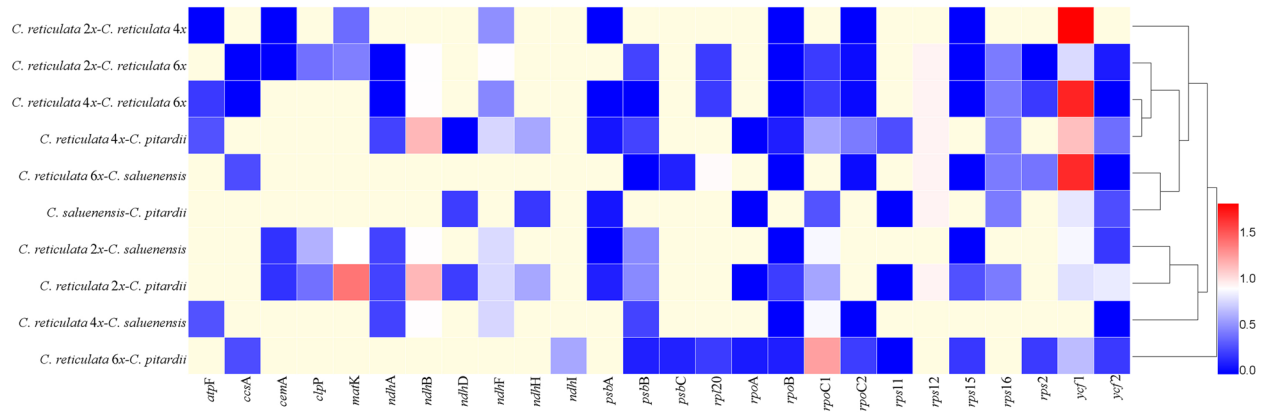

Fig. 11 Pairwise Ka/Ks ratios of chloroplast protein-coding genes of *C. reticulata* and related species

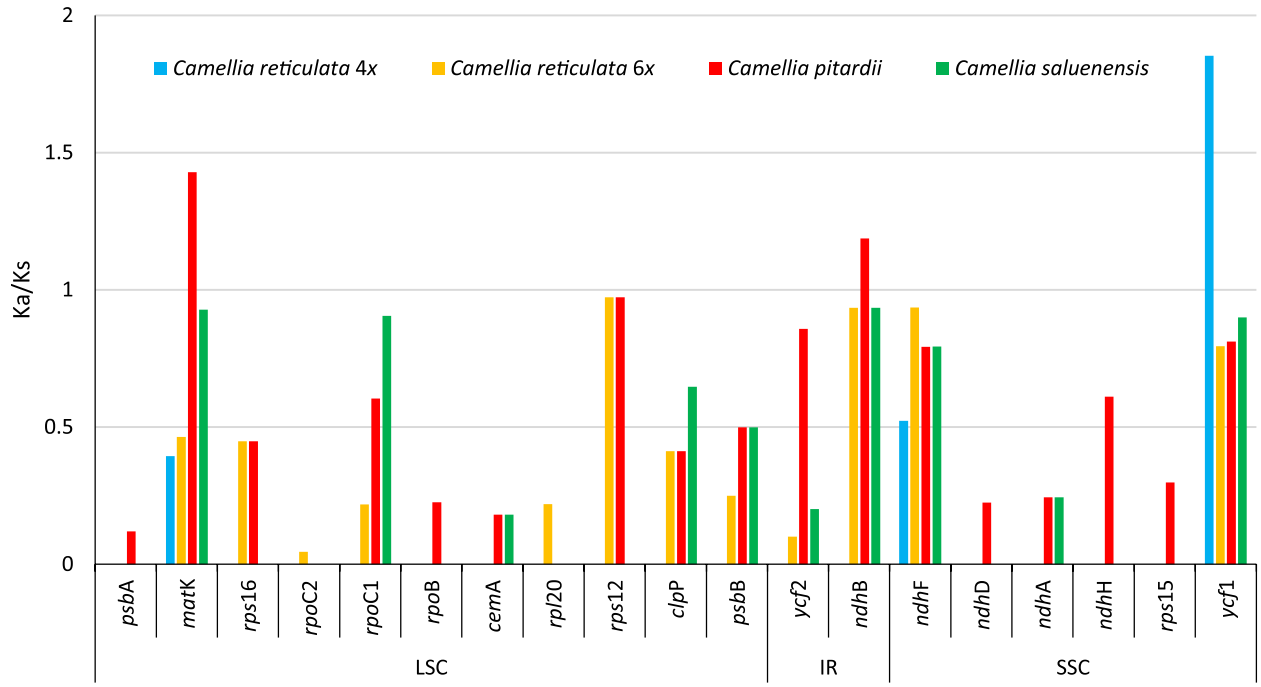

Fig. 12 The Ka/Ks ratios of 19 protein-coding genes in the chloroplast genome of diploid *C. reticulata* compared with those of related species

Moreover, in ten species comparisons, the *rpo*C1, *rps*12, *clp*P, *ycf*2, *ndh*F, and *ndh*H genes presented Ka/Ks ratios exceeding 0.5. The analysis of Ka/Ks ratios across different regions (LSC, IR, and SSC) did not reveal any specific trends.

Codon usage

This study analyzed codon usage in 50 protein-coding sequences across five species. The number of codons ranged from 20,427 (4×*C. reticulata*) to 21,104 (*C. pitardii*), with RSCU values ranging from 0.32 (CGC and AGC) to 1.98 (TTA) (Table S22, Fig. S2). Among the chloroplast protein-coding genes of *C. reticulata* and its closely related species, 64 codons encode 20 amino acids. A detailed examination of codon usage bias in 2×*C. reticulata* revealed that leucine was the most abundant amino acid, comprising approximately 10.21% of the total with 2,118 codons, whereas cysteine was the least common amino acid, accounting for approximately 1.08% with 225 codons. Among the 64 codons, ATT (encoding isoleucine) occurred most frequently (873 times), whereas TGC (encoding cysteine) appeared least often

(55 times) (Table S22). The codon AGC (serine) presented the lowest RSCU value, whereas TTA (leucine) presented the highest. Methionine (ATG) and tryptophan (TGG), which are encoded by single codons, showed no codon bias. Thirty codons had RSCU values exceeding 1, indicating preferential usage (Fig. 13, excluding stop codons). Among these preferred codons, 16 ended with T, 13 with A, and only 1 with G. This pattern suggests a strong A/T bias at the third codon position in the chloroplast genome of *C. reticulata*, which aligns with the codon usage biases observed in most angiosperms.

Phylogenetic analysis

This study utilized chloroplast genomes from 16 species within the subgenus *Camellia* for phylogenetic analysis, including the newly sequenced 2×and 4×*C. reticulata.* ML and BI methods were employed to infer phylogenetic relationships. Both the ML and BI trees partitioned the sect. *Camellia* species into two major clades, with most nodes receiving strong support (Fig. 14). Clade A comprised species from sect. *Paracamellia,* including *C. japonica*, *C. chekiangoleosa*, *C. edithae*, *C. subintegra*, *C. polyodonta*, *C. hongkongensis*, *C. azalea*, *C. semiserrata*, *and C. oleifera.* Clade B consisted of the polyploid complex of *C. reticulata*, *C. pitardii*, *C. saluenensis*, *C. mairei,* and sect. *Stereocarpus* member *C. pubipetala,* with support values and posterior probabilities of all internal nodes in this clade exceeding 90% and 0.90. Within the *C. reticulata* clade, *C. pubipetala* occupied the basal position, followed by divergences leading to *C. pitardii* and *C. saluenensis*. The *C. reticulata* complex and *C. mairei* formed robust sister groups in both phylogenetic trees, with support and posterior probability values reaching 100% and 1.00, respectively. The 2×*C. reticulata* and 4×*C. reticulata*, along with the 6×*C. reticulata* and *C. mairei*, formed two distinct sister lineages.

Discussion

The effect of allopolyploidization on the genetic diversity of chloroplast genomes

Cellular ploidy enhances genetic diversity among species and potentially provides advantages for adaptive

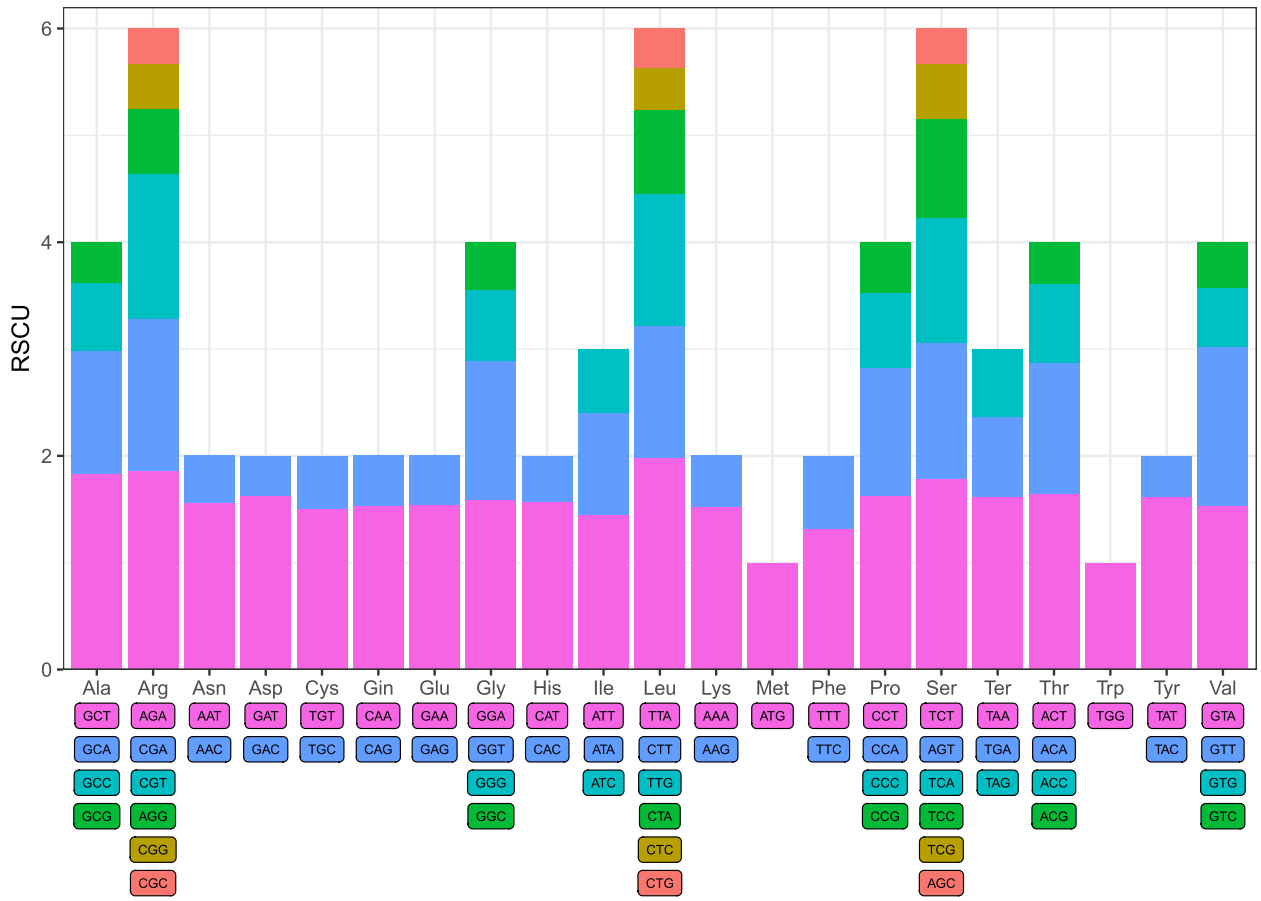

Fig. 13 Codon contents of 20 amino acids and stop codons in all protein-coding genes of the chloroplast genome of 2×*C. reticulata*. The colors correspond to the codons listed underneath the columns

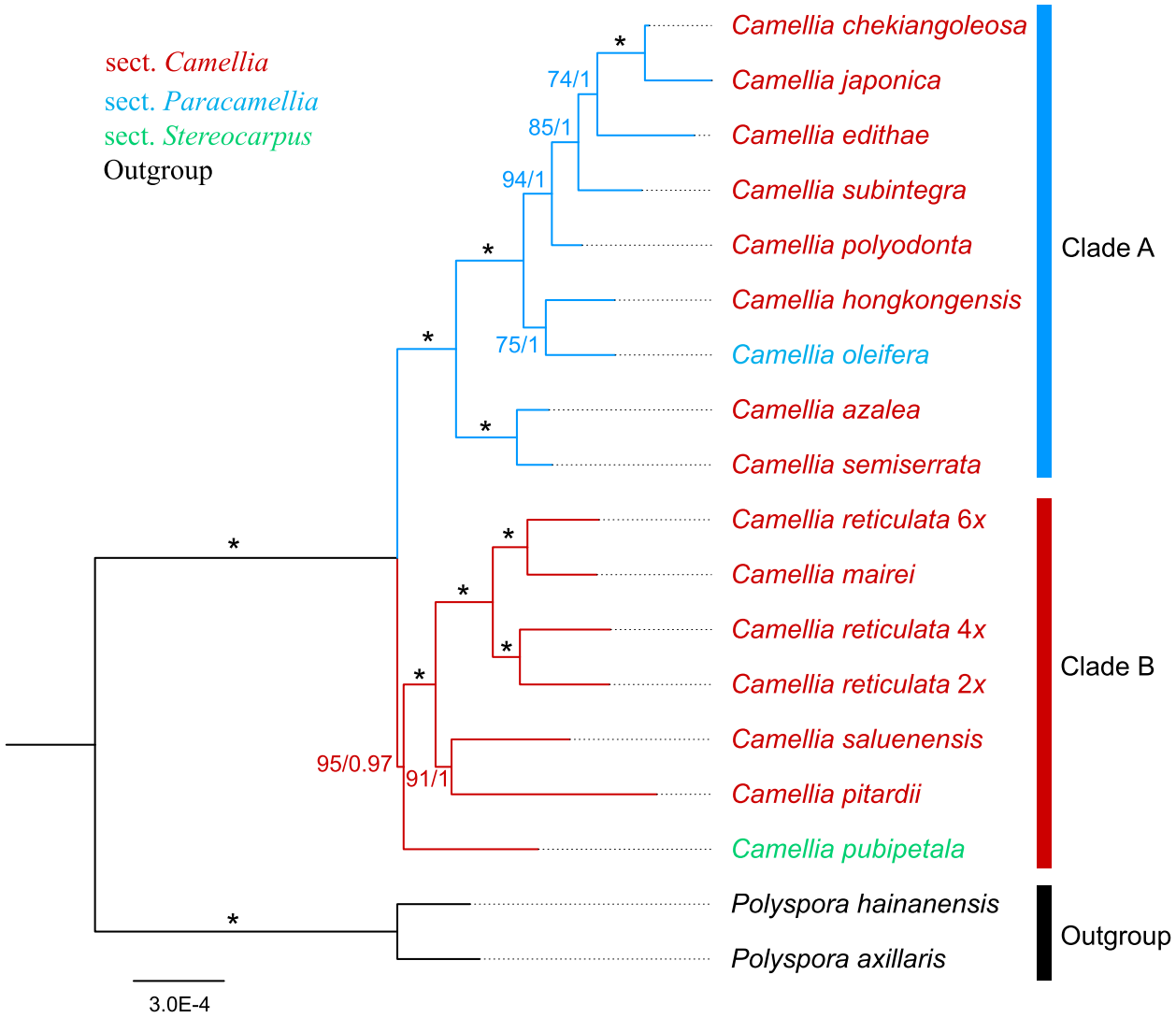

Fig. 14 ML/BI phylogenetic tree of sect. *Camellia* based on chloroplast genome sequences. Asterisks represent nodes with maximal support, ML support rates of 100% and BI posterior probability values of 1.00. The section colors in the upper left corner correspond to the species colors in the phylogenetic tree

evolution by influencing genome structure. Most seed plants have undergone polyploidy during their evolutionary history [41], and the early evolution of polyploid genomes epitomize the overall evolution of plant genomes [16]. Subsequent to heterologous polyploidization, nuclear genes may take various evolutionary pathways, including gene loss, pseudogenization, de novo functionalization, subfunctionalization, and the redistribution of expression among duplicate copies across different tissues or developmental stages [42]. Nonetheless, the chloroplast genome has been regarded as genetically and evolutionarily conserved. Our findings reveal that the chloroplast genomes of *C. reticulata* at various ploidy levels exhibit high structural conservation, all displaying a typical quadripartite structure. Additionally, the stability at the IR and SC boundaries is relatively high, which is consistent with the findings of chloroplast genome studies in other angiosperms [17]. Moreover, the chloroplast genome length, boundary gene length, gene number, types and quantities of long repetitive sequences, spacer regions, single nucleotide polymorphisms (SNPs), and nucleotide insertions and deletions (Indels) varied during the allopolyploidization of *C. reticulata* (Table 2, Table S23). The chloroplast genome length of 2×*C. reticulata* exceeds that of 4×and 6×*C. reticulata*. While previous research on allopolyploid cucumber (*Cucumis*×*hytivus*) revealed that the chloroplast genome size increases with ploidy [16], our results do not support

this trend. Despite the high conservation of chloroplast genomes, structural variations, gene losses, and transfers can occur in some species due to evolutionary processes [43]. In this study, we observed intron loss in the *clp*P gene of 4×*C. reticulata*, a phenomenon previously reported in other plant families, including *Jasminum* (Oleaceae) [18] and *Medicago* (Fabaceae) [44]. The findings regarding interspecific nucleotide polymorphisms across different ploidy levels of *C. reticulata* indicate that the molecular diversity of its chloroplast genome increases following allopolyploidization.

The variability of SSRs in *C. reticulata* may significantly contribute to the relatively minor changes in its cpDNA size [41]. The distribution of long repetitive sequences and SSRs in the chloroplast genome of *C. reticulata* indicates that the quantity and types of repetitive sequences do not significantly correlate with species ploidy. However, these repetitive sequences potentially play crucial roles in the structural evolution of chloroplast genomes. Notably, the long repetitive sequences in the *ycf*2 gene suggest its potential key role in genomic recombination and structural maintenance. In *C. reticulata* and other species, SSRs predominantly comprise A/T mononucleotides, which are consistent with patterns observed in most angiosperms [40]. Additionally, a unique six-base SSR (AAATTC/AATTTG) was identified in *C. pitardii*. These repetitive sequences not only reflect genomic stability but also offer potential markers for species identification and phylogenetic analysis.

Through mVISTA and the biological information cloud platform, we identified several highly variable regions (Figs. 9 and 10), including the *mat*K and *trn*P genes, as well as various intergenic regions (*trn*H-*psb*A, *rpo*C1-*rpo*B, *trn*C-*pet*N, *trn*F-*trn*L, *ycf*4-*cem*A, *pet*L-*psa*J, *pet*D-*rpo*A, *trn*R-*ndh*F, *rpl*32-*ccs*A, *ndh*G-*ndh*I, *trn*L-*ycf*15, *trn*S-*trn*G, *trn*G-*trn*R, *pet*N-*psb*M, *trn*N-*ndh*F) that present high mutation frequencies across different species. These variations in genes or intergenic regions may be associated with the adaptive evolution of *C. reticulata* and its closely related species. The majority of variations occur in noncoding regions, which align with previous findings and likely represent a common characteristic of chloroplast genomes in angiosperms [45]. These hotspot regions play crucial roles in the identification and characterization of species within sect. *Camellia*, with *mat*K and *trn*H-*psb*A already established as standard plant barcodes. Further investigation is necessary to determine which additional genes or intergenic regions may serve as reliable DNA barcodes.

Parental inference of the allopolyploid *Camellia reticulata*

The maternal inheritance of chloroplast genomes serves as a valuable tool for analyzing the parental composition of allopolyploids, as hybrid offspring retain a greater proportion of genetic information from the maternal species [16]. Previous phylogenetic investigations utilizing the chloroplast *rpl*16 intron suggested that the 6×*C. reticulata* originated through hybridization between 4×*C. reticulata* (maternal progenitor) and *C. saluenensis* (paternal progenitor), and the 4×*C. reticulata* was proposed to derive from *C. pitardii* (maternal ancestor) and 2×*C. reticulata* (paternal parent) [3]. The chloroplast genome SNPs and Indels analysis between paired species in this study showed that the parental relationship of 6×*C. reticulata* was consistent with the inference from the *rpl*16 intron. However, the parentage of 4×*C. reticulata* was resolved as 2×*C. reticulata* (maternal progenitor) and *C. pitardii* (paternal progenitor), contradicting the earlier *rpl*16-based inferences.

Adaptive evolution of the chloroplast genome

The analysis of adaptive evolution in chloroplast genomes is essential for understanding gene function and structural changes [46]. During the evolutionary process of chloroplast genomes, most genes have undergone purifying selection, while some participate in environmental adaptation and positive selection, and others may have experienced neutral evolution [47]. Selection pressure analysis revealed that four genes (*ndh*B, *rpo*C1, *mat*K, and *ycf*1) exhibited strong positive selection in *C. reticulata* species (Fig. 11). These genes belong to four distinct functional groups: photosynthesis (*ndh*B), self-replication (*rpo*C1), maturation enzymes (*mat*K), and genes of unknown function (*ycf*1) (Table S3). The *ndh* gene encodes chloroplast NADH dehydrogenase [48], which is involved in photosynthetic electron transport [49] and contributes to tolerance against photoxidative stress [50]. The *rpo*C1 gene encodes the β' subunit of the chloroplast RNA polymerase [48]. In the species pairs of 6×*C. reticulata* and *C. pitardii*, the Ka/Ks ratio for *rpo*C1 exceeded 1, potentially related to its functional role in chloroplast RNA polymerase. The *mat*K gene, one of the fastest-evolving genes, encodes the only maturation enzyme associated with the splicing of type II introns in land plant chloroplasts [51]. Owing to its rapid evolution, the *mat*K gene has become a valuable molecular marker in plant phylogenetic research [52]. The *ycf*1 gene is one of the largest open reading frames in the chloroplast genome [48]. Although its specific function remains unclear, its low sequence similarity among different species suggests that it is one of the most promising chloroplast DNA barcodes in terrestrial plants [53]. With Ka/Ks ratios exceeding 1.5 in three pairwise species comparisons and with gene length increasing with the ploidy level of *C. reticulata*, we hypothesize that *ycf*1 plays a crucial role in the polyploidization and adaptive evolution of *C.*

reticulata. The identification of these positively selected genes provides new insights into the role of the chloroplast genome of *C. reticulata* in plant adaptive evolution.

Phylogenetic relationships in section *Camellia*

Phylogenetic trees constructed from a limited number of gene sequences may display inconsistent or conflicting topologies due to variations in evolutionary rates and horizontal gene transfer among genes, complicating the accurate representation of species' evolutionary relationships [45, 54]. The combination of chloroplast genome conservation and variations in specific hotspot regions effectively elucidates phylogenetic relationships among different species. Through phylogenetic analysis of the chloroplast genomes of *C. reticulata* and other species within sect. *Camellia*, we further clarified the evolutionary relationships among these species. The phylogenetic trees constructed using ML and BI methods indicate that sect. *Camellia* was divided into two distinct clades. Clade A included nine species, such as *C. japonica*, *C. chekiangoleosa*, and *C. edithae*, etc. Clade B, which was strongly supported, encompasses the *C. reticulata* complex along with *C. pitardii*, *C. saluenensis*, *C. mairei*, and *C. pubipetala* (Fig. 14). This finding is consistent with the conclusions of several previous studies [17, 43, 55, 56], yet the present study examines a broader range of species within sect. *Camellia* and offers a more comprehensive depiction of the branching relationships. Morphologically and geographically, species in Clade A, characterized by a bract calyx that is more or less persistent and an ovary that is either glabrous or hairy, are predominantly found in southern China, eastern China, and Japan. In contrast, species in Clade B, which are distinguished by a deciduous bract calyx and a densely hairy ovary, are mainly located in southwestern China. The distribution areas of the former clade are largely heterogeneous or show only minor overlaps, whereas the latter clade is frequently sympatric or overlapping [6]. Sect. *Camellia* possesses the same chromosome base number (x = 15) and exhibits significant variability in ploidy levels (2×, 4×, 6×, and 8×), with polyploidy primarily observed in southern China [6]. Cytogeographical studies have identified the Nanling Mountains and adjacent regions in southern China as a center for distribution and differentiation of sect. *Camellia* [6]. Within the Jinshajiang Basin in southwestern China, taxa with 2×, 4×, and 6× ploidy levels coexist, designating this area as the secondary divergence center for sect. *Camellia* [6]. In clade B, 2× and 4× *C. reticulata* exhibit sister relationships, as both species are codistributed in the Jinshajiang Valley. However, they display significant differences in fruit morphology and notable differences in chloroplast genome structure and sequence polymorphisms, suggesting that they may not be treated as the same species. The 6× *C. reticulata* has formed a sister relationship with *C. mairei*, indicating the involvement of other species in the hybridization and polyploidization processes of *C. reticulata*, thus confirming that *C. reticulata* is an allopolyploid species. Furthermore, this implies that 4× and 6× *C. reticulata* may represent different species. The genetic distances among 2×, 4×, and 6× *C. reticulata* and their respective distances from *C. pitardii* and *C. saluenensis* are relatively similar (Table 1). Therefore, it may be more appropriate to treat 2× and 4× *C. reticulata* as distinct species and assign them new names. On the basis of the phylogenetic trees and genetic variation results, the other two progenitors of the polyploid complex, *C. pitardii* and *C. saluenensis*, are also suggested to be distinct species. Our findings support the relationships among the *C. reticulata* species established through genomic in situ hybridization [9], reaffirming the earlier confirmation that *C. pitardii* and *C. saluenensis* are the ancestors of the heteroploid hexaploid *C. reticulata* [57]. Additionally, we postulate that the polyploid complex of *C. reticulata* and *C. mairei* evolved from *C. pitardii*. These findings further support the notion of dynamic genomic changes in Theaceae plants during chloroplast transfer and polyploidization processes [17]. However, owing to the effects of hybridization and polyploidization, chloroplast genomes and nuclear genomes evolve independently, making it insufficient to rely solely on chloroplast genomic phylogenetics for classification decisions [58, 59]. Future studies could combine morphological, cellular, and nuclear genomic data from 2× and 4× *C. reticulata* to systematically investigate their status as independent species.

Conclusion

This study characterized the chloroplast genomes of 2× and 4× *C. reticulata* and conducted comparative analyses with closely related species. This research reveals diversity in structural characteristics, repetitive sequences, polymorphic hotspots, selection pressures, and codon usage, along with their systematic evolutionary significance. These findings suggest that *C. pitardii*, *C. saluenensis*, and *C. mairei* may have participated in the allopolyploidization of *C. reticulata*, with both 2× and 4× *C. reticulata* potentially classified as independent species. These results enhance our understanding of the chloroplast genomic characteristics of *C. reticulata* and contribute to elucidating the systematic evolutionary status of different ploidy levels of *C. reticulata*.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06349-5.

Supplementary Material 1.

Acknowledgements
We are grateful to Jun Wen from the Department of Botany, Smithsonian Institution, US, for her guidance on taxonomy and data analysis.

Authors' contributions
ZW and FZF conceived and designed the research. WQ collected the samples. XXD performed the experiments. FZF analyzed the data and wrote the manuscript. ZW and XXD revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding
This work was funded by the National Natural Science Foundation of China (32060691 and 31660228) and Natural Science Research Foundation of Kunming University of Science and Technology (KKZ3202456200).

Data availability
The complete chloroplast genome sequences of 2× and 4× *C. reticulata* are available in the GenBank (https://www.ncbi.nlm.nih.gov/) with accession numbers of PQ152957 and PQ152958, respectively.

Declarations

Ethics approval and consent to participate
The authors declared that experimental research works on the plants described in this paper comply with institutional, national and international guidelines. Field investigation were conducted in accordance with local legislation and get permissions from provincial department of forest and grass of Yunnan and Sichuan province.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]City College, Kunming University of Science and Technology, Kunming 650500, China. [2]Laboratory of Landscape Plants, Department of Landscape Architecture, Faculty of Architecture and City Planning, Kunming University of Science and Technology, Kunming 650500, China.

Received: 7 November 2024 Accepted: 4 March 2025
Published online: 12 March 2025

References

1. Yao H, Kato A, Mooney B, Birchler JA. Phenotypic and gene expression analyses of a ploidy series of maize inbred Oh43. Plant Mol Biol. 2011;75(3):237–51.
2. Min TL. A systematic synopsis of the genus *Camellia*. Act Bot Yunnanica. 1999;21(2):149–59.
3. Xu X, Zheng W, Harris AJ, Wang W, Shao W, Wen J. Assessing the maternal origin in the polyploid complex of *Camellia reticulata* based on the chloroplast *rpl*16 intron sequences: implications for camellia cross breeding. Mol Breeding. 2018;38(10):123.
4. Xin T, de Riek J, Guo H, Jarvis D, Ma L, Long C. Impact of traditional culture on *Camellia reticulata* in Yunnan, China. J Ethnobiol Ethnomed. 2015;11(1):74.
5. Yu DJ. A historical review and future development of *Camellia reticulata* in Yunnan. Acta Hortic Sin. 1985;12(2):131–6.
6. Min TL. The Classification, differentiation and distribution of the genus *Camellia* Sect. *Camellia.* Act Bot Yunnanica. 1998;20(2):127–48.
7. Xia LF, Gu ZJ, Wang ZL, Xiao TJ, Wang L, Katsuhiko K. Dawn on the origin of *Camellia reticulata* - the new discovery of its wild diploid in Jinshajiang valley. Act Bot Yunnanica. 1994;16(3):255–62.
8. Gu ZJ. The discovery of tetraploid *Camellia reticulata* and its implication in studies on the origin of this species. Acta Phytotaxonomica Sinica. 1997;35(2):107–16.
9. Liu LQ, Gu ZJ. Genomic in situ hybridization identifies genome donors of *Camellia reticulata* (Theaceae). Plant Sci. 2011;180(3):554–9.
10. Wang BY, Ruan ZY. Genetic diversity and differentiation in *Camellia reticulata* (Theaceae) polyploid complex revealed by ISSR and ploidy. Genet Mol Res. 2012;11(1):503–11.
11. Neuhaus HE, Emes MJ. Nonphotosynthetic metabolism in plastids. Annu Rev Plant Biol. 2000;51:111–40.
12. Raman G, Park S. The complete chloroplast genome sequence of *Ampelopsis*: gene organization, comparative analysis, and phylogenetic relationships to other angiosperms. Front Plant Sci. 2016;7:341.
13. Zhou XJ, Zhang K, Peng ZF, Sun SS, Ya HY, Zhang YZ, Cheng YW. Comparative analysis of chloroplast genome characteristics between *Paeonia jishanensis* and other five species of *Paeonia*. Scientia silvae sinicae. 2020;56(4):82–8.
14. Huang Y, Li X, Yang Z, Yang C, Yang J, Ji Y. Analysis of complete chloroplast genome sequences improves phylogenetic resolution in *Paris* (Melanthiaceae). Front Plant Sci. 2016;7:1797.
15. Hong Z, Wu Z, Zhao K, Yang Z, Zhang N, Guo J, Tembrock LR, Xu D. Comparative analyses of five complete chloroplast genomes from the genus *Pterocarpus* (Fabacaeae). Int J Mol Sci. 2020;21(11):3758.
16. Zhai Y, Yu X, Zhou J, Li J, Tian Z, Wang P, Meng Y, Zhao Q, Lou Q, Du S, et al. Complete chloroplast genome sequencing and comparative analysis reveals changes to the chloroplast genome after allopolyploidization in *Cucumis*. Genome. 2021;64(6):627–38.
17. Li L, Hu Y, He M, Zhang B, Wu W, Cai P, Huo D, Hong Y. Comparative chloroplast genomes: insights into the evolution of the chloroplast genome of *Camellia sinensis* and the phylogeny of *Camellia*. BMC Genomics. 2021;22(1):138.
18. Zhao YM, Yang ZY, Zhao YP, Li XL, Zhao ZX, Zhao GF. Chloroplast genome structural characteristics and phylogenetic relationships of Oleaceae. Chinese Bull Bot. 2019;54(4):441–54.
19. Yu XQ, Gao LM, Soltis DE, Soltis PS, Yang JB, Fang L, Yang SX, Li DZ. Insights into the historical assembly of East Asian subtropical evergreen broadleaved forests revealed by the temporal history of the tea family. New Phytol. 2017;215(3):1235–48.
20. Zhang L, Liu D, Yan Z, Zheng Y. Alien DNA introgression and wheat DNA rearrangements in a stable wheat line derived from the early generation of distant hybridization. Sci China, Ser C Life Sci. 2005;48(5):424–33.
21. Wagner ND, Marinček P, Pittet L, Hörandl E. Insights into the taxonomically challenging hexaploid alpine shrub willows of *Salix* sections *Phylicifoliae* and *Nigricantes* (Salicaceae). Plants. 2023;12(5):1144.
22. Yang G, He H, Cao Z, Fu C, Yang S, Wang B. A study on cold resistance in diploid, tetraploid and hexaploid of *Camellia reticulata* (Theaceae). J Yunnan Univ. 2016;38(2):335–43.
23. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21(1):241.
24. 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(1):59–64.
25. Kurtz Sa, Choudhuri JVa, Ohlebusch Ea, Schleiermacher Cb, Stoye Jc, Giegerich RMa. REPuter:the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29(22):4633–42.
26. Beier S, Thiel T, Munch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33(16):2583–5.
27. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35(6):1547–9.
28. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004;32(Web Server issue):273–9.
29. Amiryousefi A, Hyvonen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics. 2018;34(17):3030–1.
30. Rozas J, Ferrer Mata A, Sanchez DelBarrio JC, Guirao Rico S, Librado P, Ramos Onsins SE, Sanchez GA. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017;34(12):3299–302.
31. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: A fast and versatile genome alignment system. PLoS Comput Biol. 2018;14(1): e1005944.

32. Wang DP, Zhang YB, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics, Proteomics & Bioinformatics. 2010;8(1):77–80.
33. Tzeng YH, Pan R, Li WH. Comparison of three methods for estimating rates of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 2004;21(12):2290–8.
34. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2019;20(4):1160–6.
35. Deng W, Wang Y, Liu Z, Cheng H, Xue Y. HemI: a toolkit for illustrating heatmaps. PLoS ONE. 2014;9(11): e111988.
36. Wright F. The "effective number of codons" used in a gene. Gene. 1990;87:23–9.
37. 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):7737–49.
38. Zhang D, Gao FL, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 2020;20(1):348–55.
39. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–9.
40. Lin P, Yin H, Wang K, Gao H, Liu L, Yao X. Comparative genomic analysis uncovers the chloroplast genome variation and phylogenetic relationships of *Camellia* species. Biomolecules. 2022;12(10):1474.
41. Chen J, Guo Y, Hu X, Zhou K. Comparison of the chloroplast genome sequences of 13 Oil-Tea *Camellia* samples and identification of an undetermined Oil-Tea *Camellia* species from Hainan Province. Front Plant Sci. 2022;12:798581
42. Grover CE, Gallagher JP, Szadkowski EP, Yoo MJ, Flagel LE, Wendel JF. Homoeolog expression bias and expression level dominance in allopolyploids. New Phytol. 2012;196(4):966–71.
43. Li W, Zhang C, Guo X, Liu Q, Wang K. Complete chloroplast genome of *Camellia japonica* genome structures, comparative and phylogenetic analysis. PLoS ONE. 2019;14(5):e0216645.
44. Deng NX, Luo ZY, Sun ZX, Meng J, Zhao Y. Chloroplast genome evolution and phylogenetic analysis of *Medicago* sect. *Spirocarpos*. Acta Agrestia Sin. 2024;32(8):2394–407.
45. Ran Z, Li Z, Xiao X, An M, Yan C. Complete chloroplast genomes of 13 species of sect. *Tuberculata* Chang (*Camellia* L.): genomic features, comparative analysis, and phylogenetic relationships. BMC Genomics. 2024;25(1):108.
46. Nei M, Kumar S. Molecular evolution and phylogenetics. New York: Oxford University Press; 2000.
47. Li P, Lou G, Cai X, Zhang B, Cheng Y, Wang H. Comparison of the complete plastomes and the phylogenetic analysis of *Paulownia* species. Sci Rep. 2020;10(1):2225.
48. Wang B, Gao L, Su Y, Wang T. Adaptive evolutionary analysis of chloroplast genes in euphyllophytes based on complete chloroplast genome sequences. Acta Sci Natur Univ Sunyatseni. 2012;51(3):108–14.
49. Joët T, Cournac L, Horvath EM, Medgyesy P, Peltier G. Increased sensitivity of photosynthesis to antimycin A induced by inactivation of the chloroplast *ndh*B gene. Evidence for a participation of the NADH-dehydrogenase complex to cyclic electron flow around photosystem I. Plant Physiol. 2001;125(4):1919–29.
50. Horváth EM, Peter SO, Joët T, Rumeau D, Cournac L, Horváth GV, Kavanagh TA, Schäfer C, Peltier G, Medgyesy P. Targeted inactivation of the plastid *ndhB* gene intobacco results in an enhanced sensitvity of photosynthsis to moderate stomatal closure. Plant Physiology. 2000;123(4):1337–1349.
51. Enan MR. Low-Stringency Single-Specific-Primer PCR as a tool for detection of mutations in the *mat*K gene of *Phaseolus vulgaris* exposed to paranitrophenol. J Cell Mol Biol. 2012;10(2):71.
52. Abugalieva S, Volkova L, Genievskaya Y, Ivaschenko A, Kotukhov Y, Sakauova G, Turuspekov Y. Taxonomic assessment of *Allium* species from Kazakhstan based on ITS and *mat*K markers. BMC Plant Biol. 2017;17:51–60.
53. Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, Cheng T, Guo J, Zhou S. *ycf*1, the most promising plastid DNA barcode of land plants. Sci Rep. 2015;5:8348.
54. Firetti F, Zuntini AR, Gaiarsa JW, Oliveira RS, Lohmann LG, Van Sluys M-A. Complete chloroplast genome sequences contribute to plant species delimitation: A case study of the *Anemopaegma* species complex. Am J Bot. 2017;104(10):1493–509.
55. Huang H, Shi C, Liu Y, Mao S-Y, Gao L-Z. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. BMC Evol Biol. 2014;14(1):151.
56. Hao B, Xia Y, Zhang Z, Wang D, Ye H, Ma J. Comparative analysis of the complete chloroplast genome sequences of four *Camellia* species. Brazilian J Bot. 2024;47(1):93–103.
57. Xu X, Gao Y, Zheng W. Exploring genetic diversity and phylogenetic relationships in *Camellia reticulata* cultivars using novel low-copy nuclear gene markers. Horticulturae. 2024;10(3):303.
58. Gitzendanner MA, Soltis PS, Yi TS, Li DZ, Soltis DE. Plastome phylogenetics: 30 years of inferences into plant evolution. Adv Bot Res. 2018;85:293–313.
59. Fan ZF, Ma CL. Comparative chloroplast genome and phylogenetic analyses of Chinese *Polyspora*. Sci Rep. 2022;12(1):15984.

Publisher's Note