BMC Plant Biology

RESEARCH Open Access

Comparative analysis of the mitogenomes of multiple species of Fagaceae, with special focus on *Quercus gilva*

Yu Li[1], Si-Si Zheng[1], Gregor Kozlowski[1,2,3] and Yi-Gang Song[1*]

Abstract

Background *Quercus*, as the most abundant and widely distributed genus within the family Fagaceae, has been extensively studied at nuclear genome and plastome. However, mitogenome studies in *Quercus* remain scarce. In this study, we assemble and annotate the mitogenome of *Quercus gilva* based on Illumina and Nanopore data. Additionally, we explore the structural features of its mitogenome and provide comprehensive analyses of the phylogeny and evolution of Fagaceae.

Results The *Q. gilva* mitogenome consists of four molecules (three circular molecules and one linear molecule) with 490,015 bp in total length and 45.68% in guanine-cytosine (GC) content. The mitogenome encodes 59 genes, including 37 protein-coding genes (PCGs), 19 transfer RNA genes (tRNAs), and three ribosomal RNA genes (rRNAs). We also examine the repeat sequences, codon usage bias, RNA editing sites, and endosymbiotic gene transfer in the mitogenome. The wide distribution of repeat sequences is a key factor in mitogenome rearrangement. These is widespread gene transfer among the mitogenome, plastome, and nuclear genome of *Q. gilva*. Comparative genomic analyses of the 11 Fagaceae mitogenomes reveal significant structural variations in size and gene loss. Synteny analysis further indicates extensive genome rearrangements and inversions within the 11 mitogenomes. However, analyses of nucleotide diversity (Pi) and nonsynonymous and synonymous substitution ratio (Ka/Ks) values reveal a low rate of evolution in the mitogenomes of Fagaceae. Finally, phylogenetic analysis based on 12 conserved mitochondrial PCGs of 40 taxa strongly supports the classification of Fabids.

Conclusions In this study, the mitogenome of *Q. gilva* is newly assembled, providing important genomic resources for the phylogeny, resource conservation and development of *Quercus*. At the same time, the study of structural variation among the mitogenomes of Fagaceae species also help to elucidate the formation mechanism of mitogenome structural diversity.

Keywords *Quercus*, Mitogenome, RNA editing site, Endosymbiotic gene transfer

*Correspondence:
Yi-Gang Song
ygsong@cemps.ac.cn
[1]Eastern China Conservation Centre for Wild Endangered Plant Resources, Shanghai Chenshan Botanical Garden, Shanghai 201602, China
[2]Department of Biology and Botanic Garden, University of Fribourg, Chemin du Musée 10, Fribourg, Fribourg 1700, Switzerland
[3]Natural History Museum Fribourg, Chemin du Musée 6, Fribourg 1700, Switzerland

Introduction

Nuclear genome (nDNA), mitochondrial genome (mitogenome, mtDNA), and plastid genome (plastome, cpDNA) are the three independently inherited genomes in higher plants, and they exhibit different structure, function, and molecular evolutionary rate [1–3]. The evolutionary trajectories of these genomes have long fascinated many evolutionary biologists [4–7]. Unlike the biparentally inherited nuclear genome, organelle genomes (mitogenome and plastome) are uniparentally inherited with endosymbiotic origin [8–11]. Over the course of evolution, the gene content of organelle genomes has been significantly reduced, largely due to gene loss or transfer to nuclear genome [12–14]. Recent advances in sequencing technologies have opened new avenues for exploring the evolution of plant organelle genomes.

Normally, plastomes range from 120 kb to 170 kb in size and are highly conserved with stable circular and quadripartite structure that contains the core genes for photosynthesis [15–18]. In contrast, plant mitogenomes, with the core genes for respiration, exhibit extreme variability in both size (ranging from 66 kb to 12 Mb) and structural topology (master circular molecules, subgenomic circular molecules, linear molecules, and branched multigenomic structures), even among closely related species [19–23]. This variability is further complicated by the presence of highly variable intergenic spacer regions (IGS) and conserved coding sequences (CDSs) [12, 24, 25]. Notably, plant mitogenomes present an intriguing evolutionary paradox: while they have a remarkably low mutation rate, they exhibit a high rate of structural rearrangements, such as gene duplications, inversions, and translocations [2, 26, 27]. Endosymbiotic gene transfer (EGT) is a frequent phenomenon among the mitogenome, plastome, and nuclear genome in plants [28, 29]. Evidence suggests that the gene transfer from plastome to mitogenome occurs in a unidirectional manner, whereas the gene transfer between nuclear genome and mitogenome is bidirectional [30, 31].

Due to these complex properties, the assembly and study of mitogenomes face significant challenges. To date, over 13,000 plastomes have been fully assembled, while fewer than 1,000 plant mitogenomes have been completed [6]. To better understand the unique evolutionary features of plant mitogenomes (such as their mutation rates, structural variations, and gene transfer events) and their contributions to plant phylogeny, further studies on a larger number of mitogenomes are essential in the future.

Quercus, with approximately 450 species, is divided into two subgenera and eight Sect. [32]. Due to its high ecological adaptability, *Quercus* has become a widespread taxon in the Northern Hemisphere [33–35]. Recent developments in high-throughput sequencing have greatly facilitated the assembly of *Quercus* genomes. Numerous studies, focusing on the evolution of nuclear genomes and plastomes (including structure variation, population genomics, introgression, phylogenomics, and so on), have been published in the last decade [36–41]. Compared with the plentiful number of nuclear genomes (16 individuals) and plastomes (120 individuals) available for *Quercus*, only five mitogenomes, belonging to two sections, have been sequenced (*Quercus acutissima*, *Q. cerris*, and *Q. variabilis* in section *Cerris*; and *Q. petraea* and *Q. robur* in section *Quercus*) (Accessed on November 24, 2024) [42, 43]. It will be interesting to explore the structural variation, mutation rate, and phylogenetic relationship of *Quercus* based on mitogenomes from different sections.

Quercus gilva, which belongs to subgenus *Cerris* section *Cyclobalanopsis*, is an endangered species with very low genetic diversity in the subtropical evergreen broadleaved forests of East Asia [44, 45]. As an important local quality timber species, its population size and number have declined dramatically in recent decades [45–47]. Recently, the nuclear genome and plastome of *Q. gilva* have been reported, providing an opportunity to compare their mutation rate and explore their gene transfer events [48–50].

In this study, we successfully assembled the *Q. gilva* mitogenome using Illumina short reads and Nanopore long reads, providing the first mitogenome reference for *Quercus* section *Cyclobalanopsis*. We examined the structural features of the *Q. gilva* mitogenome, evaluated endosymbiotic gene transfer events among the plastome, mitogenome, and nuclear genome of *Q. gilva*, and explored the phylogeny and evolution of *Q. gilva* and related species within the family Fagaceae. The purpose of this study is to provide valuable genomic resources that will support future research on *Quercus*, particularly in the areas of species identification, conservation genetics, and evolutionary biology. Furthermore, our findings lay a theoretical foundation for understanding the role of the mitogenomes in the adaptation and speciation of *Quercus* species, offering new perspectives on their evolutionary trajectories.

Materials and methods

Plant material, DNA extraction, and sequencing

Fresh and healthy plant leaves of *Quercus gilva* were sampled from Fujian Province, China (29.94°N, 116.87°E). The voucher specimen was identified by one of the authors (Yi-Gang Song), and deposited in the Chenshan Herbarium (CSH) at Shanghai Chenshan Botanical Garden. High-quality genomic DNA was extracted from fresh leaves using the modified cetyltrimethylammonium bromide (CTAB) protocol [51]. DNA quality and concentration were tested by 0.75% agarose gel electrophoresis

analysis, NanoDrop One spectrophotometer (Thermo Fisher Scientific), and Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). The genomic DNA was then randomly fragmented using the Covaris ultrasonic disruptor.

Paired-end (PE) sequencing libraries with 350 bp insert fragments were performed using the Nextera DNA Flex Library Prep Kit (Illumina, San Diego, CA, USA). The products were then subject to sequencing (each end 150 bp PE reads) on the Illumina HiSeq XTen platform. For Oxford Nanopore sequencing, the sequencing libraries were prepared using the Ligation Sequencing Kit SQK-LSK109 (Oxford Nanopore Technologies, Cambridge, the UK). The purified library was loaded onto primed R9.4 Spot-On Flow Cells for sequencing using the Oxford Nanopore PromethION sequencer (Oxford Nanopore Technologies, Oxford, UK) at Wuhan Benagen Tech Solutions Company Limited, Wuhan, China.

Genome assembly and annotation

In this study, we *de novo* assembled and annotated organelle genomes of *Q. gilva*. For the plastome, we completely assembled it by employing the "get_organelle_from_reads.py" command of the GetOrganelle v1.7.6.1 software, with the parameter "-F embplant_pt" [52]. Moreover, for the mitogenome, we tested three different assembly strategies: GetOrganelle v1.7.6.1, Unicycler v0.5.1, and GSAT v1.10 [52–54]. A rough mitogenome was initially obtained with the other parameter of "-F embplant_mt" by the GetOrganelle v1.7.6.1 software [52]. Unicycler was also used to perform hybrid assembly [53]. We then further assembled the mitogenome by the GSAT v1.10 software based on the rough mitogenome and Nanopore long reads [54]. After comparing the assembly completeness and structural consistency of the results from all three approaches, we selected the GSAT-assembled mitogenome for downstream analysis due to its higher accuracy and completeness (Fig. S1 and S2A).

Four different assembled graphs, namely OG (organelle graph), MRG (mitochondrial rough graph), MRMG (mitochondrial rough master graph), and MMG (mitochondrial master graph), were obtained through the four steps of graphShort, graphLong, graphSimplification, and graphCorrection respectively. Finally, the final mitogenome was obtained by manually checking and correcting the assembly structure with the Bandage software [55]. Minimap2 was then used to map the raw high-throughput sequencing data, including both Illumina and Nanopore reads, back to the assembled mitogenome to evaluate the coverage distribution [56].

The assembled organelle genomes were annotated using the online annotation tool GeSeq (https://chlorobox.mpimp-golm.mpg.de/geseq.html) [57]. The annotation results were manually reviewed and corrected using the Geneious R9.0.2 software [58]. The plant mitogenomes typically encode 24 core genes (including nine NADH dehydrogenase genes, five ATP synthase genes, four ubiquinol cytochrome c reductase genes, three cytochrome c oxidase genes, one cytochrome c biogenesis gene, one maturase gene, and one transport membrane protein gene) and 17 variable genes (including four large subunit of ribosome genes, eleven small subunit of ribosome genes, and two succinate dehydrogenase genes) [12]. The open reading frames (ORFs) in the mitogenome were annotated using the gene prediction tool ORF Finder (https://www.ncbi.nlm.nih.gov/orffinder/) [59]. Finally, the genomic structures of the organelle genomes were plotted using the organelle genome drawing program OGDRAW (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) [60]. The organelle genome sequences and their annotations were submitted to GenBank with accession numbers PQ862789–PQ862792 for the mitogenome and PQ876385 for the plastome, which are scheduled for release on July 24, 2025.

Repeat sequence identification

Based on the sequence length, repeat sequences can be divided into three types, namely simple sequence repeats (SSRs), tandem repeat sequences (TRSs), and dispersed repeat sequences (DRSs). In this study, we analyzed these repeat sequences in the mitogenome. SSRs were detected using the MISA Perl script (http://pgrc.ipk-gatersleben.de/misa/) [61]. The thresholds of repeat units were set to 10, 5, 4, 3, 3, and 3 for the mono-, di-, tri-, tetra-, penta-, and hexanucleotides, respectively. TRSs were predicted using the Tandem Repeats Finder (TRF) web service (https://tandem.bu.edu/trf/trf.advanced.submit.html) with the following parameters: 2, 7, and 7 for matches, mismatches, and indels, respectively, and 50 and 500 for the minimum alignment score and maximum period size, respectively [62]. Finally, DRSs, including forward (F), reverse (R), complementary (C), and palindromic (P) repeat sequences, were analyzed using the REPuter web service (https://bibiserv.cebitec.uni-bielefeld.de/reputer) [63]. The hamming distance and minimal repeat size were set as 3 and 30 bp, respectively. The visualization of the repeat sequences was accomplished using the "Circlize" package in R.

Codon usage analysis and RNA editing site prediction

The protein-coding genes (PCGs) of the mitogenome were extracted using the Geneious R9.0.2 software [58]. Subsequently, twenty-eight PCG sequences, which had a length exceeding 300 bp and a start codon of ATG, were chosen for codon usage analysis by the CodonW v1.4.2 software [64]. The relative synonymous codon usage (RSCU) values were calculated and drawn the stacked bar plot using the "ggplot2" package in R.

The Deepred-Mt software was used to predict potential cytosine (C) to uracil (U) RNA editing sites in the mitogenome based on the convolutional neural network (CNN) model [65]. The threshold of the probability was set to 0.9.

Endosymbiotic gene transferred fragments identification

To identify the mitochondrial plastid transferred fragments (MTPTs) and nuclear mitochondrial transferred fragments (NUMTs) of *Q. gilva*, we compared the mitogenome with the nuclear genome and plastome using the BLASTn v2.9.0 software [66]. The plastome was assembled from this study. And the nuclear genome was obtained from the previous study of Wang et al., which has been deposited in the National Genomics Data Center (NGDC) (https://ngdc.cncb.ac.cn) under the access ion number GWHDOCW00000000, with a scheduled public release date of June 27, 2025 [48]. The mitogenome was used for building database with makeblastdb, while the nuclear genome and plastome served as query sequences with the criteria of E-value ≤ 1e − 5, identity ≥ 80%, and alignment length ≥ 50 bp. Eventually, the endosymbiotic gene transferred fragments were visualized using the "Circlize" package in R.

Comparative analyses of the mitogenomes of Fagaceae

We downloaded all publicly available mitogenomes of ten Fagaceae species from the National Center of Biotechnology Information (NCBI) database and reannotated them to reduce potential analytical errors. We performed a gene presence-absence analysis across all species and visualized the results using a heatmap generated with the TBtools-II software [67]. BLASTn was used for pairwise comparison of the mitogenomes of these 11 Fagaceae species [66]. Homologous sequences greater than 500 bp as conserved colinear blocks were then retained for collinearity analysis and plotting using the NGenomeSyn v1.41 software [68]. Not only that, we also downloaded the corresponding plastomes of these ten species. The shared PCGs of the organelle genomes (30 of mitogenomes and 79 of plastomes) across the 11 species were aligned using the MAFFT v7.487 software [69]. These PCGs sequences were then used for nucleotide diversity analysis and Ka/Ks analysis. Nucleotide diversity (Pi) values of each PCG were calculated using the DnaSP v6.12.03 software with step size of 10 bp and window length of 100 bp [70]. Finally, the calculation of the nonsynonymous (Ka) and synonymous (Ks) substitution ratio (Ka/Ks) for each shared PCG was performed for pairwise species comparisons using the KaKs_Calculator v3.0 software [71], employing the MLWL method [72]. A gene was considered to be under positive selection if at least one pairwise comparison yielded a Ka/Ks value greater than 1.

Phylogenetic analysis

We downloaded the mitogenomes of 38 Fabid taxa from the NCBI database, including seven Fagales species, seven Cucurbitales species, eleven Rosales species, and thirteen Fabales species, and performed phylogenetic analyses with the new assembly mitogenome of *Q. gilva* (Table S1). The complete mitogenome of *Arabidopsis thaliana* (NC_037304) was used to root the tree as an outgroup. We extracted the CDSs of 12 shared PCGs (*atp4*, *atp8*, *nad2*, *nad4*, *nad6*, *nad7*, *nad9*, *rps3*, *matR*, *cox3*, *ccmB*, and *ccmC*) from 40 species using the PhyloSuite v1.2.3 software [73]. The CDSs were aligned using the MAFFT v7.450 software with default parameters [69]. The phylogenetic tree was constructed using the maximum-likelihood (ML) method implemented in the IQ-tree v2.1.3 software with 1,000 ultrafast bootstrap replicates [74]. Finally, the phylogenetic tree was visualized and beautified in the FigTree v.1.4.4 software (http://tree.bio.ed.ac.uk/software/figtree/).

Results

Assembly of the mitogenome of *Q. gilva*

Based on Illumina short reads and Nanopore long reads, we successfully assembled the master graphs of the mitogenome with the GSAT software. A total of 615.7, 482.4, and 481.1 kb sequences with 39, 27, and nine contigs and 52, 36, and 12 links were obtained from each step of OG, MRG, and MRMG, respectively (Table S2). The final MMG was 481,113 bp in length, with nine contigs and 12 links (Figure S2A and Table S2). All contigs exhibited high coverage depths, with short-read coverage ranging from 1257.14 X to 3015.83 X and long-read coverage ranging from 400.06 X to 1054.17 X (Table S3). This indicated that all contigs were strongly supported by both short and long reads, further confirming the high quality and accuracy of the mitogenome assembly. Notably, the long-read coverage depths of contig 1 (1054.17 X) and contig 3 (802.43 X) were approximately twice as high as those of the other contigs (Figure S2B and Table S3), suggesting that they likely represent large repeats with two copies.

To evaluate the homologous recombination events between subgenomes mediated by these two large repeat sequences, we systematically analyzed the possible recombination pathways in Bandage by integrating the contig graph structure and sequencing read support. After evaluating all possible pathways in the MMG, it was finally found that the mitogenome showed a complex multi-branched structure through repeat-mediated recombination (Figure S2C). Specifically, molecule 1 consists of contig 8; molecule 2 consists of contig 1 (and its repeat copy), contigs 2, 5, and 6; molecule 3 contains contigs 3, 4, and 7; and molecule 4 includes the duplicated contig 3 and contig 9 (Table 1). These rearrangements

reflect the dynamic nature of the mitogenome structure, which likely exists as a mixture of alternative conformations in vivo.

General features of the mitogenome and plastome

The mitogenome consisted of four molecules (molecules 1–4) with 490,015 bp in total length and 45.68% in guanine-cytosine (GC) content. The lengths of circular molecule 1, linear molecule 2, circular molecule 3, and circular molecule 4 were 96,997, 125,848, 133,477, and 133,693 bp, respectively, and the GC contents were 47.01, 45.19, 45.98, and 44.88%, respectively (Figure S2C and Table 1).

The mitogenome annotated a total of 37 protein-coding genes (PCGs) (*atp9* were multi-copy), 19 transfer RNA genes (tRNAs) (*trnM-CAU* and *trnP-UGG* were multi-copy), and three ribosomal RNA genes (rRNAs) (Fig. 1 and Table 2). Among the 36 unique PCGs, 24 PCGs were identified as core genes, and the other 12 PCGs were classified as variable genes. The 12 variable genes comprised four large subunits of ribosome genes, six small subunits of ribosome genes, and two succinate dehydrogenase gene. Seven genes were cut by introns: *rpl2* and *ccmFc* contained a single intron, *nad4* harbored three introns, while *nad1*, *nad2*, *nad5*, and *nad7* had four introns. Notably, three PCGs (*nad1*, *nad2*, and *nad5*) were identified as special trans-splicing genes. Detailed information on these genes was provided in Table S4. In addition, a total of 193 open reading frames (ORFs) larger than 300 bp were predicted by the ORF Finder in the *Q. gilva* mitogenome (Table S5).

The plastome assembled in this study was 160,757 bp in length, with a typical circular tetrad structure consisting of a large single copy (LSC) region (90,217 bp) and a small single copy (SSC) region (18,870 bp), separated by a pair of inverted repeats (IR) regions (25,835 bp). A total of 86 PCGs, 37 tRNAs and eight rRNAs were annotated in the plastome (Figure S3).

Repeat sequences

The mitogenome contained a large number of repeat sequences, with their quantity and specific distribution shown in Fig. 2C. Using the MISA Perl script, a total of 20, 40, 52, and 43 SSRs were identified in the four molecules, respectively (Fig. 2A and Table S6). SSRs exhibited diverse types; however, hexanucleotide repeats were not detected in any of the four molecules (Fig. 2A). The tetranucleotide repeats were the most abundant, accounting for 65 loci (40.37%), followed by mono- (40, 24.84%), di- (35, 21.74%), tri- (15, 9.32%), and pentanucleotide repeats (6, 3.73%). Most SSRs were composed of two complementary bases, adenine (A) and thymine (T). Additionally, two, five, four, and five TRSs were detected in the four molecules, respectively (Fig. 2A and Table S7). The copy number of these TRSs was primarily between two and four, with the size of the basic repeat units ranging from 15 to 45 bp (Table S7). Among DRSs detected by the REPuter, no reverse or complementary repeats in the four molecules, while a total of 69 forward repeats and 50 palindromic repeats were identified respectively (Fig. 2A and Table S8). The lengths of DRSs were predominantly between 30 and 60 bp (Fig. 2B). Statistical analysis of the distribution of all repeats showed that most were located in the intergenic spacer regions (IGS), with only a small fraction found within the gene regions (Fig. 2D).

Codon usage bias

The codon usage analysis of 28 selected PCGs revealed a universal codon usage preference in the mitogenome. Among the RSCU values calculated for the 64 codons, 30 codons had RSCU values greater than 1 (Table S9). Of these, only three codons (UCC, ACC, and UUG) ended in G/C, while the remaining 27 codons ended in A/U, indicating a strong preference for A/U ending codons (Figure S4 and Table S9). Histidine (His) exhibited the strongest bias, favoring the CAU codon (RSCU = 1.51), followed by Alanine (Ala), which preferred the GCU codon (RSCU = 1.50). Additionally, the most frequently used codons in the mitogenome were UUU (Phe), AUU (Ile), UUC (Phe), and GAA (Glu), each appearing more than 250 times.

Prediction of RNA editing sites

Based on the deep representation learning method (Deepred-Mt), we predicted 491 potential C to U RNA editing sites in the 36 PCGs of the *Q. gilva* mitogenome (Table S10). Among the PCGs, *nad4* had the highest number of predicted RNA editing sites, followed by *ccmB*, while no RNA editing sites were detected in *rpl2* (Fig. 3A). Of the 491 RNA editing sites, 461 (93.90%) sites resulted in nonsynonymous mutation, while 30 (6.10%) sites were synonymous mutation. The nonsynonymous editing

Table 1 Branched conformation of the *Quercus gilva* mitogenome

Molecule	Path	Type	Length (bp)	GC content (%)	PCGs	tRNAs	rRNAs	ORFs
molecule 1	contig8	Circular	96,997	47.01	7(2)	2	1	43
molecule 2	contig2_1_5_1(copy)_6	linear	125,848	45.19	11(3)	4	0	51
molecule 3	contig7_3_4	Circular	133,477	45.98	6(3)	8	2	52
molecule 4	contig3(copy)_9	Circular	133,693	44.88	10	5	0	47
Total		Branched	490,015	45.68	37	19	3	193

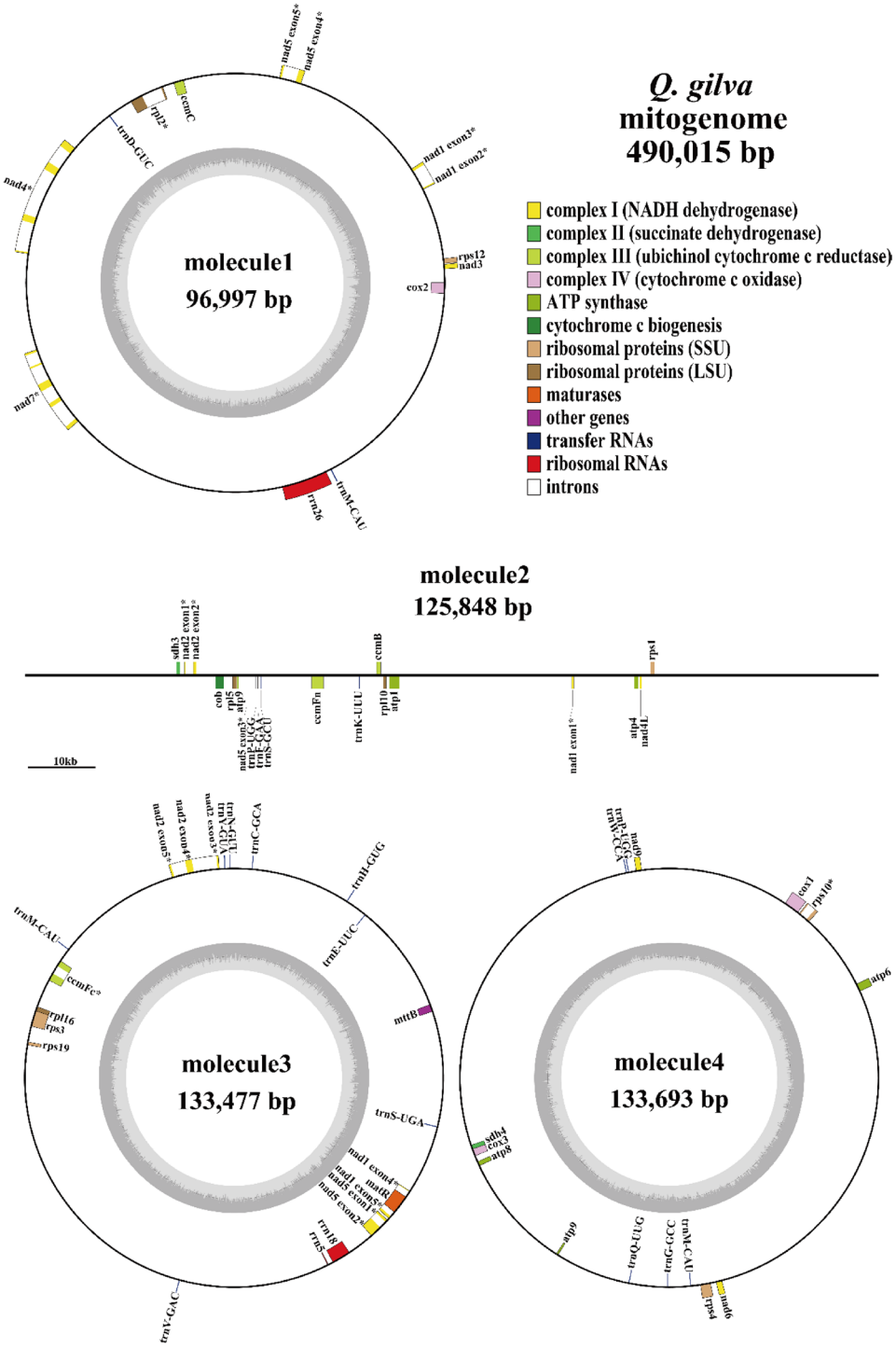

Fig. 1 Genome map of the *Quercus gilva* mitogenome. Genes outside the circle and on the line are transcribed in the clockwise direction, while genes inside the circle and below the line are transcribed in the counterclockwise direction. The genes belonging to different functional groups are identified by different colors

events caused 12 distinct types of amino acid changes. The most frequent mutation was the replacement of serine (Ser) and proline (Pro) with leucine (Leu), occurring 112 and 105 times, respectively (Fig. 3B). Furthermore, 227 (46.23%) sites converted hydrophilic amino acids into hydrophobic ones, while 50 (10.18%) reversed this trend, changing hydrophobic amino acids to hydrophilic ones. Notably, three RNA editing sites at *ccmFc*–1315, *nad1*–308, and *rps10*–391 introduced stop codons (Arg→Ter) (Table S10).

Table 2 Gene composition of the *Q. gilva* mitogenome

Group of genes	Number	Gene name
ATP synthase	6	atp1, atp4, atp6, atp8, atp9 (×2)
Ubiquinol cytochrome c reductase	4	ccmB, ccmC, ccmFc*, ccmFn
Cytochrome c biogenesis	1	cob
Cytochrome c oxidase	3	cox1, cox2, cox3
Maturases	1	matR
Transport membrane protein	1	mttB
NADH dehydrogenase	9	nad1****, nad2****, nad3, nad4***, nad4L, nad5****, nad6, nad7****, nad9
Large subunit of ribosome (LSU)	4	rpl2*, rpl5, rpl10, rpl16
Small subunit of ribosome (SSU)	6	rps1, rps3, rps4, rps10*, rps12, rps19
Succinate dehydrogenase	2	sdh3, sdh4
Ribosomal RNAs	3	rrn5, rrn18, rrn26
Transfer RNAs	19	trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnH-GUG, trnK-UUU, trnM-CAU (×3), trnN-GUU, trnP-UGG (×2), trnQ-UUG, trnS-GCU, trnS-UGA, trnV-GAC, trnW-CCA, trnY-GUA

Gene (×) Number of copies of multi-copy genes, *Bold* trans-splicing genes that fragment across multiple branches

*one intron

***three introns

****four introns

Endosymbiotic gene transfer

A total of 11 MTPTs between the mitogenome and plastome of *Q. gilva* were identified by conducting sequence similarity calculations (Fig. 4A and Table S11). The total length of the MTPTs was 6,512 bp, representing approximately 1.33% of the mitogenome and 4.05% of the plastome (Table S11). Among them, two MTPTs exceeded 1,000 bp in length, namely MTPT3 (1,322 bp) and MTPT7 (1,149 bp). Five MTPTs were observed to have two copies localized within the IR regions of the plastome, while the remaining MTPTs were located in the LSC region. Annotation results of the MTPTs revealed that these homologous fragments were derived from 14 plastid genes, including six complete genes (*trnD-GUC*, *trnI-CAU*, *trnV-GAC*, *trnH-GUG*, *trnN-GUU*, *and trnM-CAU*) and eight partial genes. Although all PCGs were pseudogenized, the tRNA genes retain functional roles in the mitogenome. Compared with the 11 MTPTs, we also found 1824 NUMTs between the mitogenome and nuclear genome (Fig. 4B and Table S12). The matched lengths ranged from 50 to 133,698 bp (Table S12). Notably, extensive sequence transfer was observed on chromosome 5 in the nuclear genome, which encompassed nearly the entire mitogenome (Fig. 4B). However, the origin of these fragments remained to be determined. Thus, these findings indicated a widespread gene transfer among the mitogenome, plastome, and nuclear genome of *Q. gilva*.

Comparison and collinearity of the mitogenomes in Fagaceae

The mitogenomes of eleven Fagaceae species, including five genera (*Quercus*, *Fagus*, *Castanea*, *Castanopsis*, and *Lithocarpus*), exhibited notable structure variation. Genome sizes ranged from 349,290 bp (*Q. petraea*) to 592,702 bp (*C. carlesii*), representing a significant disparity (Table S13). The length of *Q. gilva* mitogenome was the longest among the existing mitogenomes of *Quercus*. In contrast, GC content varied within a narrow range of 45.5% (*C. carlesii*) to 46.1% (*Q. petraea*), indicating low variability in base composition. Additionally, during the evolutionary process, the mitogenomes had also undergone the loss of PCGs. Our results showed that a total of 30 PCGs were shared among these 11 species (Figure S5). *Quercus petraea* exhibited the highest number of gene deletions, including two core genes. The number of large and small subunits of ribosome genes varied the most among all mitogenomes. More notably, *rps14* was unique to *Fagus sylvatica* and had been lost in all other species.

A synteny analysis was also conducted for these 11 species for deeper comparative insights. The results revealed a significant number of homologous collinear blocks among these Fagaceae species, alongside substantial genome rearrangements and inversions (Fig. 5). Notably, the largest collinear blocks were identified between *Q. petraea* and *Q. robur* within the same section, followed by *C. mollissima* and *C. henryi* within the same genus. This suggested that closely related species were more likely to retain larger syntenic blocks.

Pi values and Ka/Ks ratio of organelle genomes from Fagaceae

Pi values were calculated for the PCGs shared by the 11 organelle genomes of Fagaceae, respectively, including 30 PCGs of mitogenomes and 79 PCGs of plastomes (Fig. 6A and Figure S6A). On average, the Pi values of chloroplast genes were higher than those of mitochondrial genes. Moreover, the difference of Pi values of chloroplast genes was greater than that of mitochondrial genes. The *rpl22* and *ycf1* in the plastome and *atp1*, *atp9*, and *cox2* in the mitogenome showed relatively high level of Pi values, reflecting their high sequence diversity within the family Fagaceae.

By calculating the Ka/Ks ratio of common PCGs, we compared the evolutionary rates in the organelle genomes of Fagaceae to assess the selection pressure of each PCG (Fig. 6B and Figure S6B). In general, most PCGs had Ka/Ks values of less than 1 in the organelle

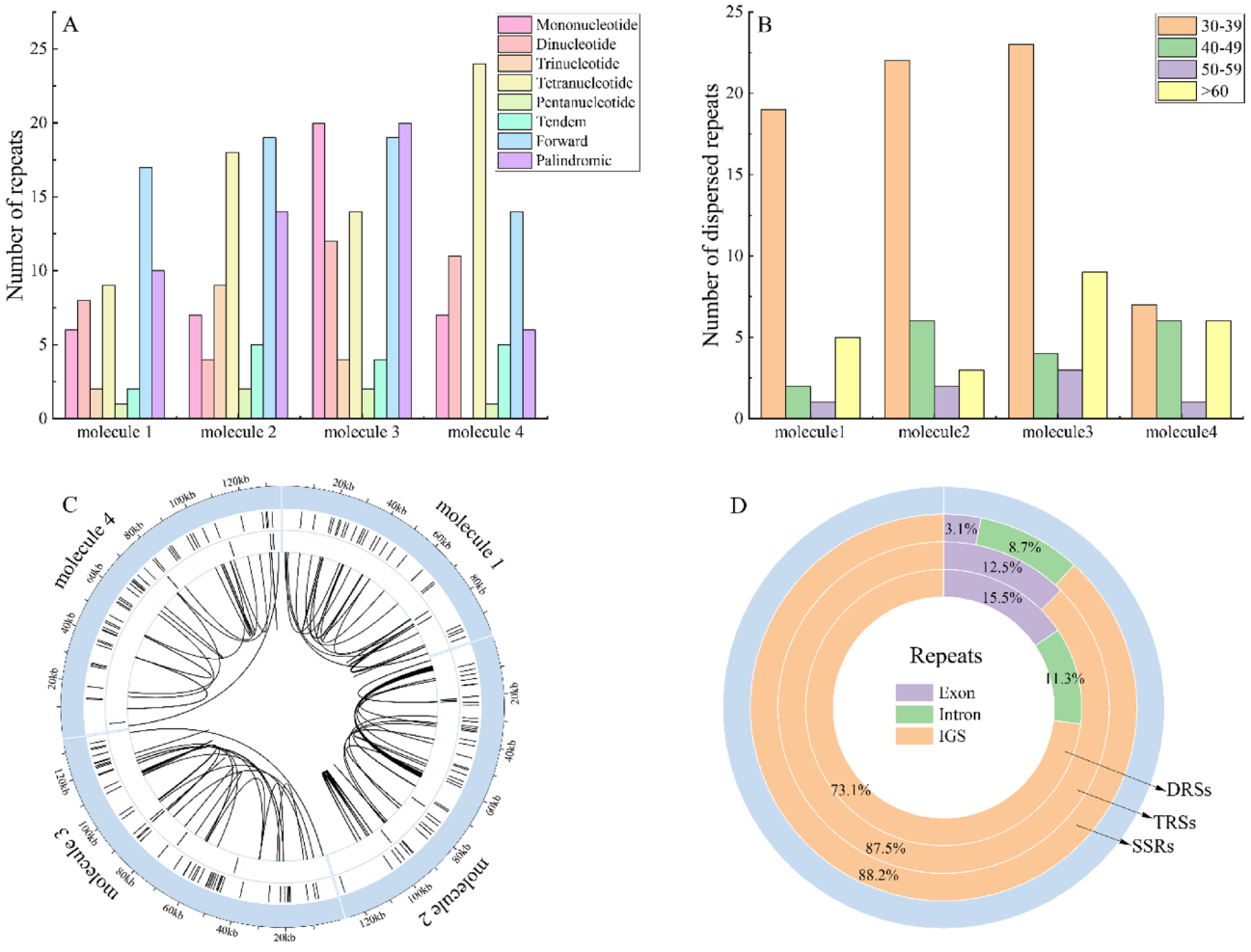

Fig. 2 The repeats of the *Q. gilva* mitogenome. (**A**) Total number of repeats across various types; (**B**) Distribution of dispersed repeat sequences by length categories (30–39 bp, 40–49 bp, 50–59 bp, and > 60 bp); (**C**) Circos plot showing the distribution of all repeats. The outermost circle represents the *Q. gilva* mitogenome, while the layers from the outside to the inside illustrate the distribution of SSRs, TRSs, and DRSs, respectively; (**D**) Proportion of all repeats located in different genomic regions. SSRs, simple sequence repeats; TRSs, tandem repeat sequences; DRSs, dispersed repeat sequences; IGS, intergenic spacer regions

genomes. However, there were ten chloroplast PCGs (*accD*, *ccmA*, *infA*, *matK*, *nadH*, *petA*, *rpl14*, *rpl22*, *rpoC2*, and *ycf2*) and seven mitochondrial PCGs (*atp6*, *atp9*, *ccmC*, *ccmFN*, *matR*, *nad1*, and *rps3*) in which at least one pairwise comparison had a Ka/Ks value greater than 1, indicating potential positive selection pressure for these PCGs (Table S14 and Table S15). These PCGs with high Pi and Ka/Ks values can be used as potential markers for phylogenetic studies of Fagaceae.

Phylogeny

In order to determine the phylogenetic position of *Q. gilva*, the phylogenetic tree was reconstructed after sequence alignment based on 12 conserved mitochondrial PCGs from the selected 40 angiosperm species (Fig. 7). The results showed that the bootstrap support (BS) values of most nodes in the phylogenetic tree were greater than 80, which strongly supported the divisions of the four orders of Fabids. The phylogenetic tree was consistent with the angiosperm phylogeny group IV (APG IV) and plastid phylogenomic angiosperm (PPA II) classification system. In the *Quercus* genus, the *Q. gilva* of section *Cyclobalanopsis* in this study differentiated first at the base of the phylogenetic tree.

Discussion

The mitogenome of *Quercus gilva*

In this study, we successfully assembled the mitogenome of *Q. gilva* by integrating short-read and long-read sequencing data, providing valuable information for mitogenome research within *Quercus* and Fagaceae. Compared to other published mitogenomes from closely related species within the family, the size, GC content and gene content of the mitogenome of *Q. gilva* were in the medium range, indicating a high level of assembly completeness and quality [42, 43, 75–79]. The mitogenome of

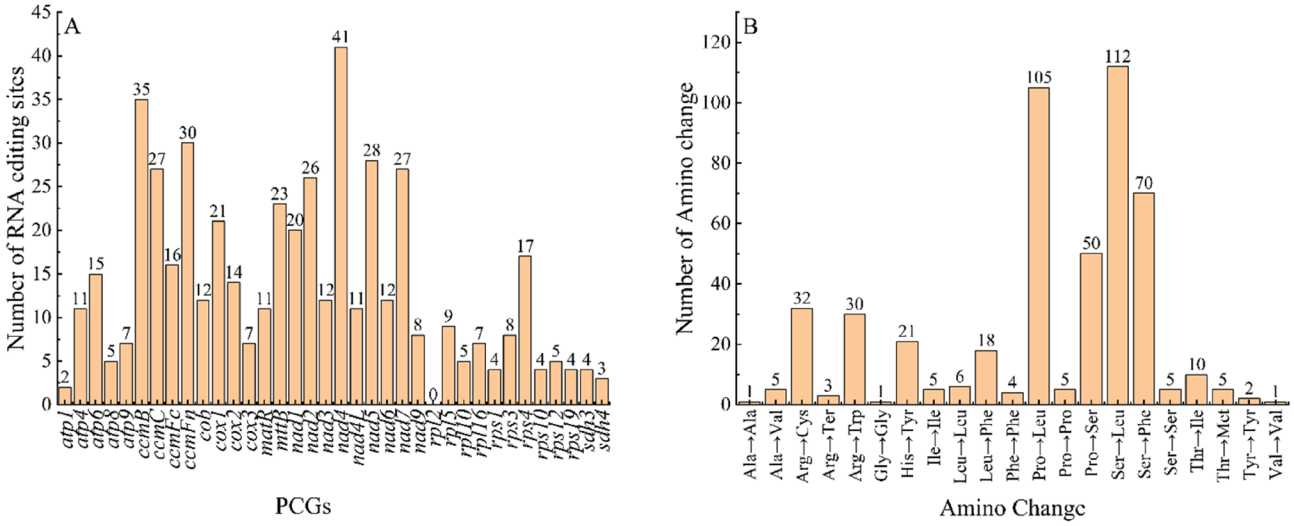

Fig. 3 Characteristics of RNA editing sites predicted in the 36 PCGs of the *Q. gilva* mitogenome. (**A**) The number of RNA editing sites predicted in the PCGs. (**B**) The distribution of amino acid changes resulting from RNA editing events

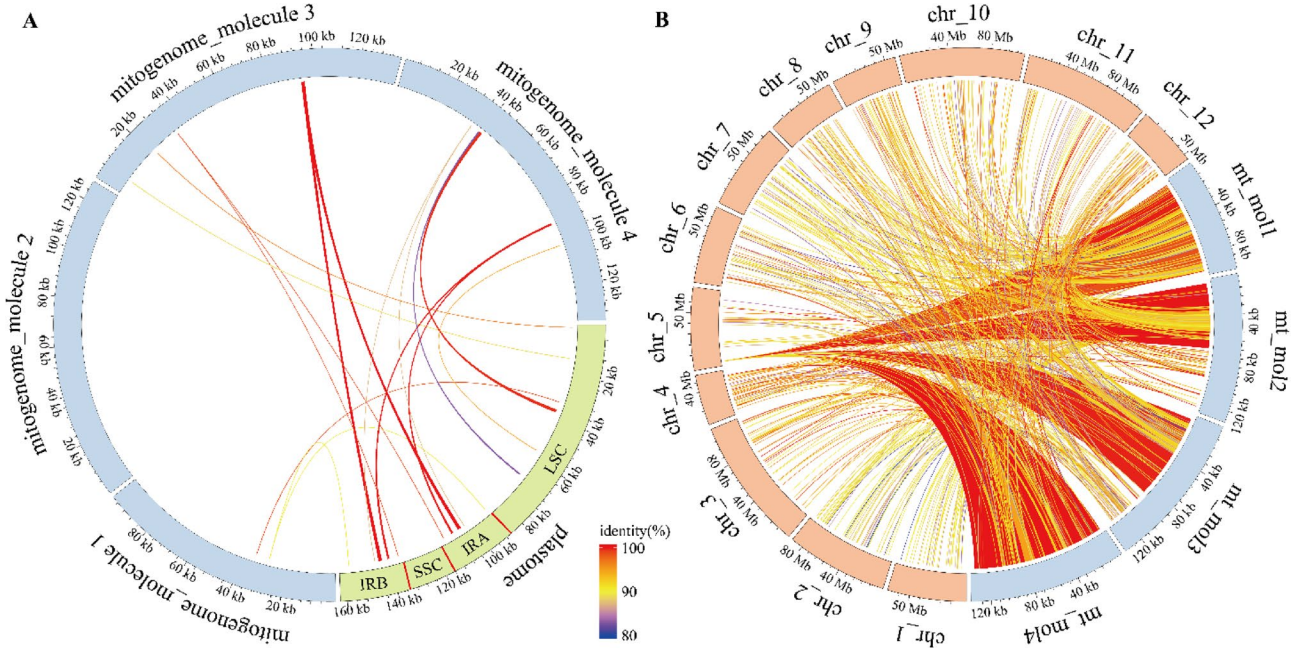

Fig. 4 Schematic representations of MTPTs and NUMTs of *Q. gilva*. (**A**) Endosymbiotic gene transfer between the mitogenome and plastome. (**B**) Endosymbiotic gene transfer between the mitogenome and nuclear genome. The green, blue, and orange arcs represent the plastome, mitogenome, and nuclear genome, respectively. And the inner colored lines connecting the arcs indicate MTPTs and NUMTs, with different colors representing the level of the sequence identity. MTPT, mitochondrial plastid transferred fragments; NUMT, nuclear mitochondrial transferred fragments

Q. gilva had a branched structure, which was similar to that of the mitogenome of *Q. acutissima* [43]. The widespread distribution of repeats in this mitogenome was likely to contribute to genome rearrangement, which was a key factor in the existence of mitogenome isoforms [80–83]. Dispersed repeat sequences were predominantly composed of forward and palindromic repeats, a pattern that was also observed in the mitogenomes of *Saposhnikovia divaricate* (Apiaceae) and *Calophyllum soulattri* (Calophyllaceae) [84, 85]. Statistical analysis revealed that most of these repeats were located in the IGS, which were generally considered hotspots for genome rearrangement. The presence of repeats within IGS likely enhanced their recombination potential, further contributing to the dynamic evolution and plasticity of mitogenomes [86].

RNA editing was widespread in the mitogenomes of higher plants, where it regulated gene expression by altering transcriptional sequences and played a crucial

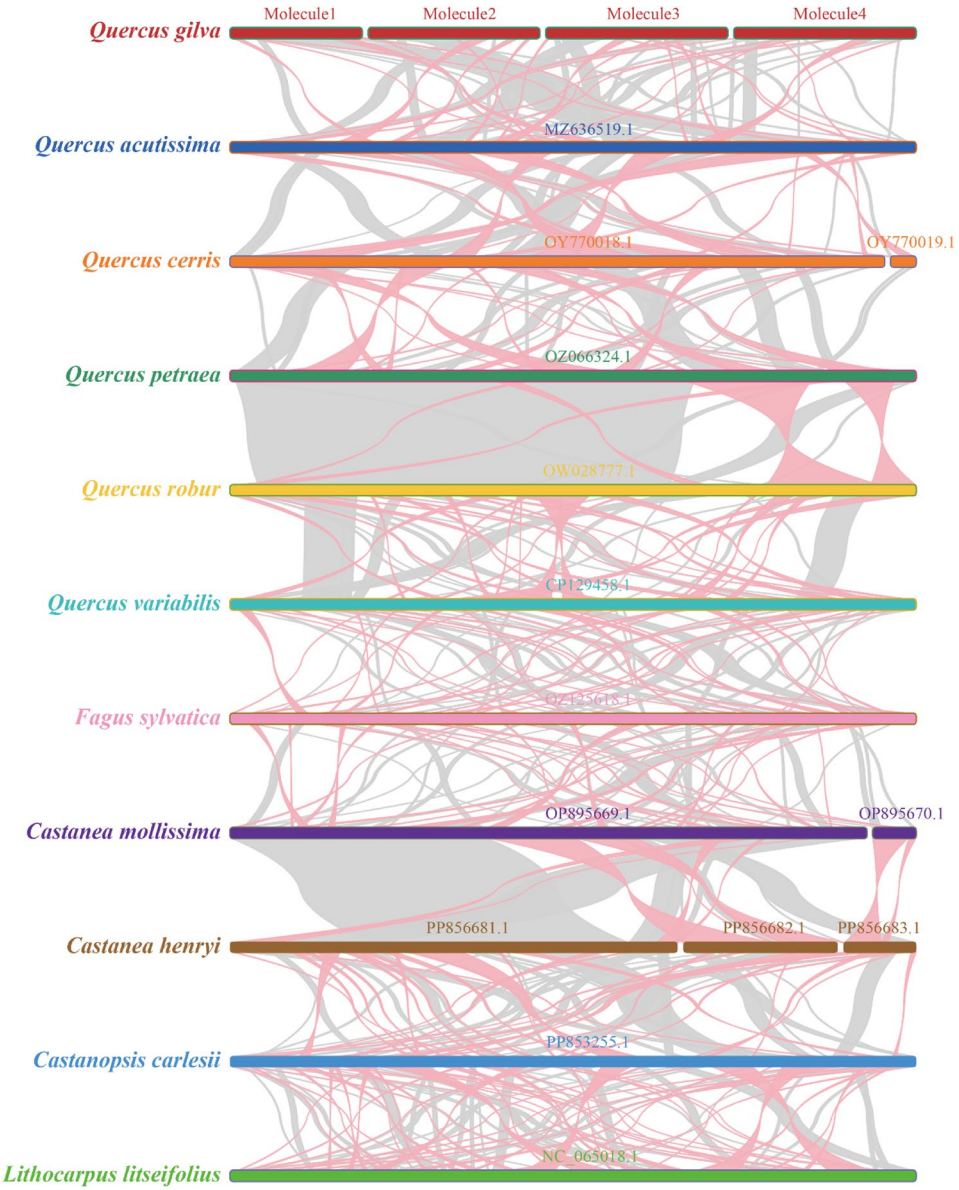

Fig. 5 Collinearity analysis of the mitogenomes of 11 Fagaceae species. The red arc area implies inversion, whereas the gray area represents high homology

role in the mitogenomes [87–89]. The most common form of RNA editing was a deamination reaction that converted C to U at a specific site [90]. In this study, we predicted a total of 491 potential RNA editing sites in 36 PCGs in the *Q. gilva* mitogenome, all of which exhibited the conversion of C to U. These RNA editing sites were unevenly distributed among different genes, with *nad4* exhibiting the highest number of editing sites, followed by *ccmB*. This distribution pattern was consistent with previous findings in *Q. acutissima* [43]. Furthermore, the RNA editing sites were mainly located at the second codon position, which was also observed in other plant mitogenomes [91, 92]. These RNA editing sites provided valuable insights into the regulation of mitochondrial gene expression and function in plants.

Endosymbiotic gene transfer in plant cells had been recognized as a significant evolutionary mechanism, facilitating the movement of genetic material between organelle genomes and nuclear genome [6]. Three main types of EGT events were commonly observed: between mitochondrion and nucleus (NUMT), between chloroplast and nucleus (NUPT), and between mitochondrion and chloroplast (MTPT) [9]. We identified 11 MTPTs in the *Q. gilva* mitogenome, which included six complete chloroplast genes. MTPTs had also been reported in other angiosperms, such as *Ipomoea batatas*

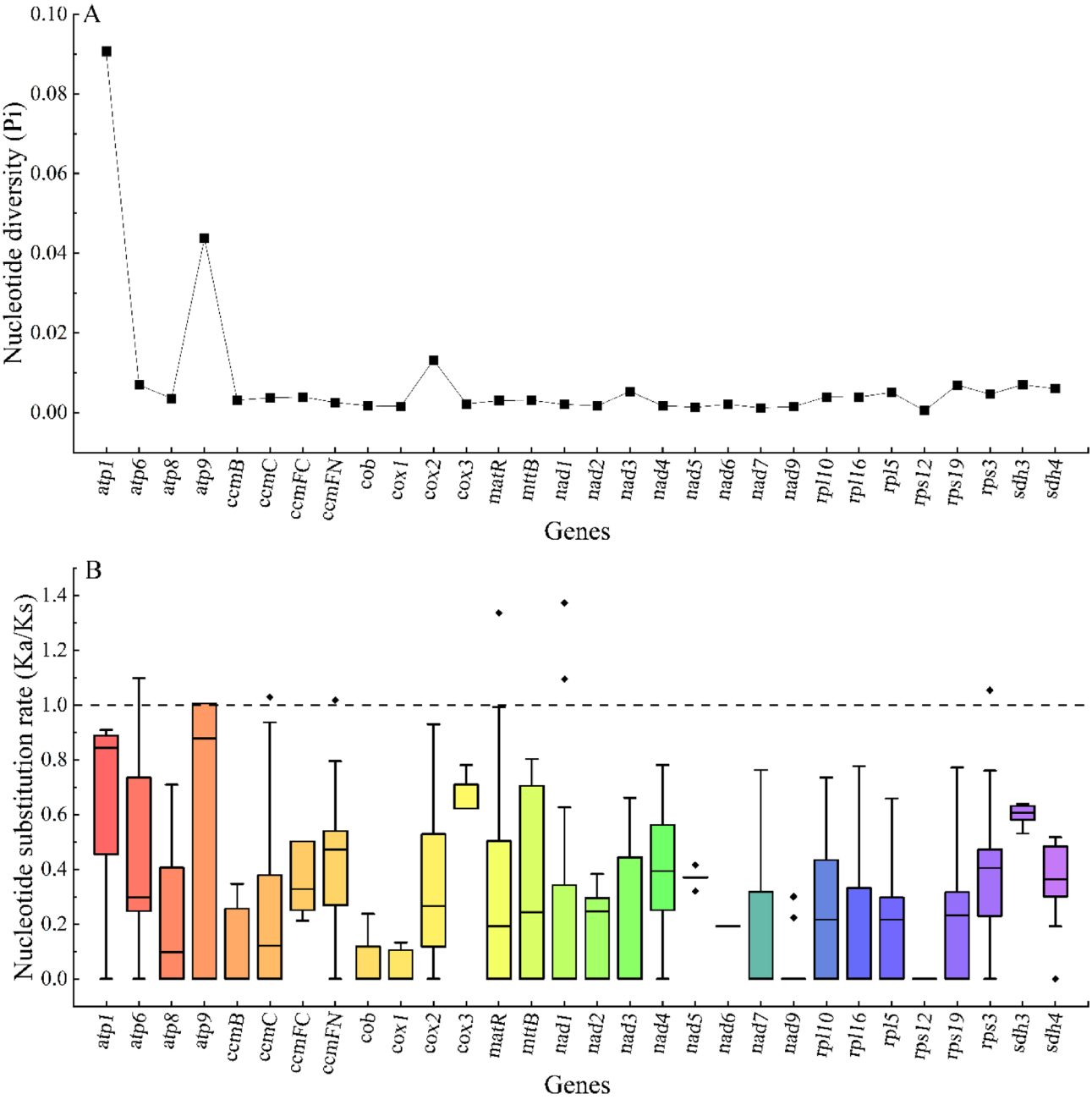

Fig. 6 The Pi values (**A**) and Ka/Ks ratio (**B**) of 30 common mitogenome PCGs among Fagaceae. The box plot represents the distribution of Ka/Ks values for all pairwise comparison in the data. Pi, nucleotide diversity; Ka/Ks, nonsynonymous and synonymous substitution ratio

(Convolvulaceae), *Cannabis sativa* (Cannabaceae), and *Paeonia lactiflora* (Paeoniaceae) [93–95]. Notably, most of the identified MTPTs were tRNA genes, which were commonly transferred between organelle genomes in angiosperms [96]. In addition to MTPTs, we also identified 1824 NUMTs in the *Q. gilva*, further emphasizing the widespread nature of gene transfer. Notably, extensive sequence transfer was observed on chromosome 5, which encompassed nearly the entire mitogenome.

These EGT events had profound implications for the evolution of angiosperms [94]. On the one hand, by integrating genetic fragments from one genome into another, EGT events increased genomic complexity and genetic diversity [97]. On the other hand, some transferred genes were retained and subsequently functionalized, acquiring new biological roles in the recipient genome and enhancing the plant's adaptability to environmental changes [98]. Although the successful establishment of functional

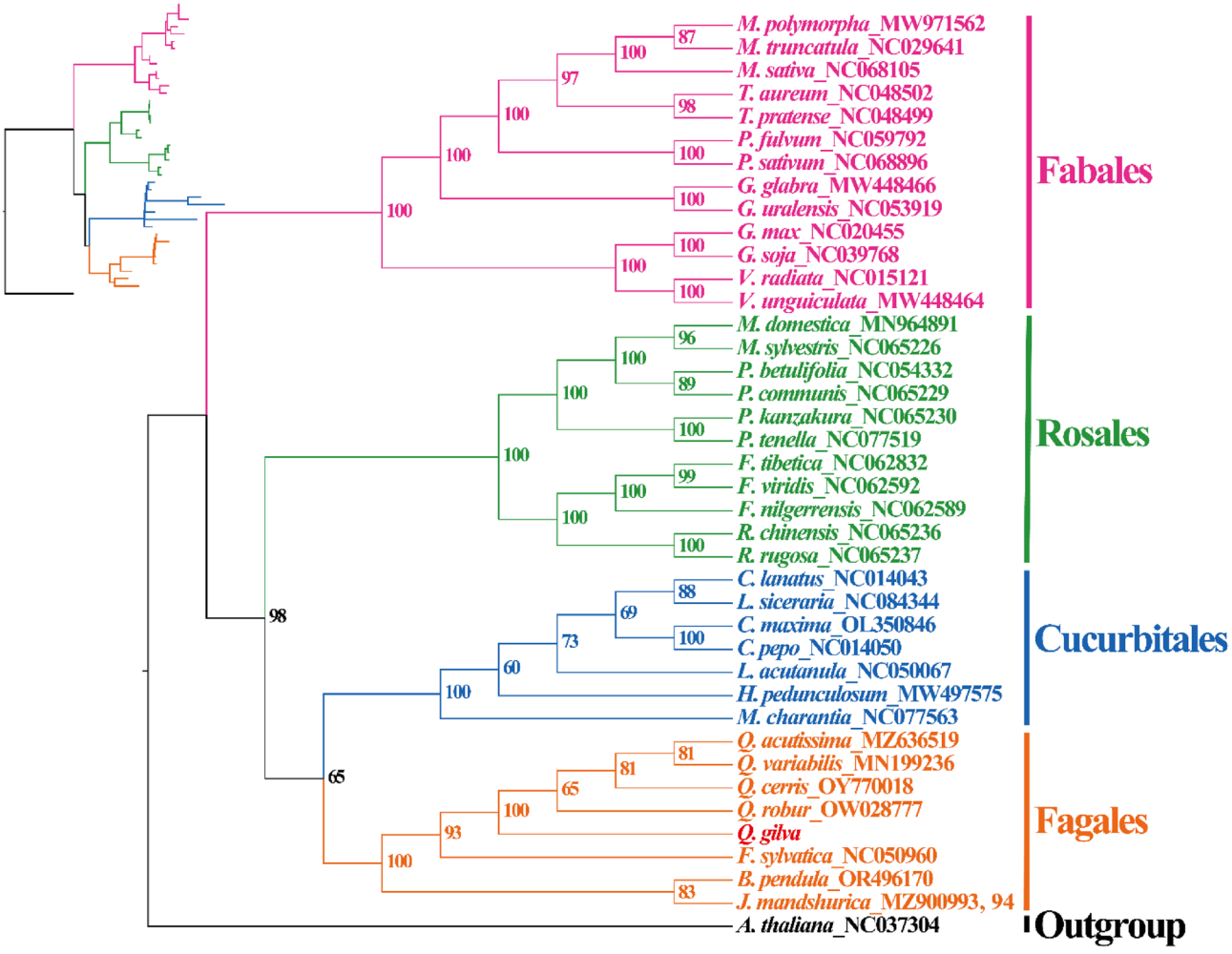

Fig. 7 Phylogenetic tree of 40 angiosperms based on 12 conserved mitochondrial PCGs. The number at each node is the bootstrap support values

genes in the nuclear genome was relatively rare, a large number of non-functional DNA transfer events appeared to be widespread, suggesting that such gene transfers accumulated over evolutionary timescales through neutral processes [99]. Therefore, the combination of functional genomics and expression-based validation will be essential to further elucidate the specific roles of EGT events in plant adaptive evolution.

Significant structural variations in the mitogenomes of Fagaceae

The mitogenomes of plants exhibited complex structures and remarkable diversity, which had significantly hindered the progress of mitogenome research [19, 100, 101]. This trend was also evident in Fagaceae, where the number of published plastomes was nearly 20 times greater than that of mitogenomes. The mitogenome size in Fagaceae ranged from 349,290 bp to 592,702 bp, highlighting the substantial size variation not only within the same family but also within the same genus [102, 103]. The number of genes in the mitogenomes was also highly diverse [21, 85]. However, changes in mitogenome size did not correlate with the gene number, as the genome size was primarily influenced by the accumulation of non-coding regions [104]. The number of large and small subunits of ribosome genes varied most among all mitogenomes. In addition, three rRNA genes were consistently present: *rrn5*, *rrn18* (*rrnS*), and *rrn26* (*rrnL*), which were essential for mitochondrial protein synthesis. These findings strengthened the understanding of mitogenome diversity in Fagaceae.

Many studies had traditionally presented plant mitogenomes as having a single circular structure [92, 105, 106]. However, with the advancements in sequencing technologies and assembly methods, increasing evidence had shown that mitogenomes exhibited structural polymorphism [22, 107]. These structures included branching linear structures [108, 109], polycyclic structures [110, 111], and linear structures [112]. The mitogenome of *Q. gilva* assembled in this study also displayed a branching linear

structure, consisting of three circular molecules and one linear molecule. This assembly structure was similar to that of the mitogenome of *Q. acutissima* [43]. However, the mitogenomes of *F. sylvatica* and *Lithocarpus litseifolius* in Fagaceae were simplified circular structure [76, 78]. And the mitogenomes of *Castanea mollissima*, *Castanopsis carlesii*, and *C. henryi* possessed complex multi-branched structures [75, 79]. These diverse structures were crucial for understanding the evolution of mitogenome architecture in Fagaceae, though the coexistence of multiple molecular structures needed further investigation.

A comprehensive genome-wide collinearity analysis of 11 Fagaceae species was performed. The results revealed the complex structural evolution of Fagaceae. The significant number of homologous syntenic blocks highlighted the conserved nature of mitogenomes, while the observed genome rearrangements and inversions underscored the dynamic evolutionary processes of genome structure. Notably, the largest syntenic blocks were identified between *Q. petraea* and *Q. robur* within the same section, as well as between *C. mollissima* and *C. henryi* within the same genus. These findings strongly suggested that closely related species were more likely to retain larger syntenic regions. Similar patterns had also been observed in the mitogenomes of other plant species [113, 114].

Phylogeny and evolution in the mitogenomes of Fagaceae

We analyzed the Pi values and Ka/Ks ratios of shared PCGs in the mitogenomes and plastomes of Fagaceae, providing valuable insights into the evolution of different species [115]. The results from both analyses indicated that the evolutionary rate of plastid genes was higher than that of mitochondrial genes, highlighting the relatively lower evolutionary rate of mitogenome [2, 6]. Genes exhibiting higher Pi and Ka/Ks values (Ka/Ks > 1) were likely involved in adaptive evolutionary processes within these Fagaceae species. Moreover, these genes could serve as potential molecular markers for phylogenetic studies and species identification within Fagaceae. In the mitogenome, *atp9* was a gene co-screened by the two analyses. The ATP9 protein played a crucial role in proton transport ATPase activity and proton transfer [116]. The evolution of this gene may influence the growth and development of Fagaceae plants.

The mitogenomes of plants exhibited a relatively slow evolutionary rate, making them valuable tools for phylogenetic studies [25]. Our phylogenetic analysis grouped the selected 40 species into distinct taxa corresponding to their respective families and orders. Furthermore, the phylogenetic tree also revealed concordance with both the APG IV and PPA II classification systems [117, 118], demonstrating the reliability of using mitogenomes for constructing phylogenetic relationships. However, within the *Quercus* clade, *Quercus gilva* from section *Cyclobalanopsis* differentiated first at the base, a finding that contrasted with results derived from nuclear genome (RAD-seq) and plastome [38, 50, 119]. The different results may be due to the lack of data on the one hand, and may be due to the serious hybridization of *Quercus* on the other hand [120]. As the dominant species in the subtropical evergreen broad-leaved forests of East Asia (EBLFs), the early divergence of section *Cyclobalanopsis* may also reflect ancient lineages associated with early ecological events, which provides a new evolutionary context for our mitochondrial genomics findings [121]. Therefore, in order to further explore the phylogenetic relationships and evolutionary biology of this genus, we need to sequence and analyze the mitogenomes of more species of *Quercus*.

Abbreviations

BS	Bootstrap support
CDS	Coding sequences
CTAB	Cetyltrimethylammonium bromide
DRS	Dispersed repeat sequence
EGT	Endosymbiotic gene transfer
IGS	Intergenic spacer region
IR	Inverted repeats
Ka/Ks	Nonsynonymous and synonymous substitution ratio
LSC	Large single copy
MTPT	Mitochondrial plastid transferred fragment
NCBI	National center of biotechnology information
NGDC	National genomics data center
NUMT	Nuclear mitochondrial transferred fragment
ORF	Open reading frame
PCG	Protein-coding gene
Pi	Nucleotide diversity
rRNA	Ribosomal RNA
RSCU	Relative synonymous codon usage
SSC	Small single copy
SSR	Simple sequence repeat
tRNA	Transfer RNA
TRS	Tandem repeat sequence

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-07072-x.

Supplementary Material 1: Figure S1. Structural graph comparison of the *Q. gilva* mitogenome assembled using different tools. (A) Mitochondrial genome graph assembled with GetOrganelle v1.7.6.1, with a total length of 388,738 bp; (B) Mitochondrial genome graph assembled with Unicycler v0.5.1, with a total length of 479,114 bp. Figure S2. Schematic representation of the assembly results of the *Q. gilva* mitogenome. (A) The final MMG graph assembled using the GSAT software; (B) The circular diagram of all contigs, with the green and blue rings representing the coverage depths of short and long reads, respectively; (C) The multi-branched structure graph resulted from resolving the structure of (A). Figure S3. Genome map of the *Q. gilva* plastome. Genes outside the circle are transcribed in the counter-clockwise direction, whereas those inside the circle are transcribed in the clockwise direction. The genes belonging to different functional groups are identified by different colors. The length and boundary of the LSC, SSC, and the two IR regions are indicated in the inner circle. The dark gray area indicates GC content while the lighter gray corresponds to the AT content of the genomes. LSC, large single copy; SSC, small single copy; IR, inverted repeat. Figure S4. Relative synonymous codon usage (RSCU) values of the *Q. gilva* mitogenome. Figure S5. Distribution of the PCGs in the mitogenomes of eleven Fagaceae species. Red boxes indicate

that the PCG is present while blue boxes indicate that the PCG is absent. PCGs are categorized into core genes and variable genes following the classification of Adams et al. (2003). PCG, protein-coding genes. Figure S6. The Pi values (A) and Ka/Ks ratio (B) of 79 common plastome PCGs among Fagaceae. The box plot represents the distribution of Ka/Ks values for all pairwise comparison in the data. Pi, nucleotide diversity; Ka/Ks, nonsynonymous and synonymous substitution ratio

Supplementary Material 2: Table S1. The basic information of selected angiosperms for constructing phylogenetic trees in this study. Table S2. Assembly statistics of different graphs of the *Q. gilva* mitogenome. Table S3. Contigs of the MMGs of the *Q. gilva* mitogenome. Table S4. Basic information statistics of genes of the *Q. gilva* mitogenome. Table S5. Basic information statistics of open reading frames (ORFs) of the *Q. gilva* mitogenome. Table S6. Simple sequence repeats (SSRs) of the *Q. gilva* mitogenome. Table S7. Tandem repeat sequences (TRSs) of the *Q. gilva* mitogenome. Table S8. Dispersed repeat sequences (DRSs) of the *Q. gilva* mitogenome. Table S9. Relative synonymous codon usage (RSCU) value and codon number of 61 codons of the *Q. gilva* mitogenome. Table S10. The predicted C to U RNA editing sites in the PCGs of the *Q. gilva* mitogenome. Table S11. The homologous mitochondrial plastid transferred fragments (MTPTs) identified between the mitogenome and plastome of *Q. gilva*. Table S12. The homologous nuclear mitochondrial transferred fragments (NUMTs) identified between the mitogenome and nuclear genome of *Q. gilva*. Table S13. Basic information on the eleven assembled mitogenomes in Fagaceae (access on 2024.11.24). Table S14. Ka/Ks values for pairwise comparisons of mitogenomes across 11 Fagaceae species. Table S15. Ka/Ks values for pairwise comparisons of plastomes across 11 Fagaceae species

References

Acknowledgements
We want to thank Wuhan Benagen Technology Co., Ltd. for their help on sequencing.

Authors' contributions
Y.G. S. conceived and conceptualized the study; Y. L. analyzed the mitogenome and wrote the manuscript; S.S. Z. and Y.G. S. collected the sample; G. K. and Y.G. S. reviewed and revised the manuscript. All authors have read and approved the final manuscript.

Funding
This work was supported by the National Natural Science Foundation of China (grant number 31901217), the Special Fund for Scientific Research of Shanghai Landscaping and City Appearance Administrative Bureau (grant numbers G192422, G242414, and G242416), and the Science and Technology Development Center, National Forestry and Grassland Administration (grant number KJZXSA202214).

Data availability
The assembled complete organelle genomes of *Quercus gilva* has been deposited in GenBank (NCBI), under accession numbers PQ862789–PQ862792 for the mitogenome and PQ876385 for the plastome.

Declarations

Ethics approval and consent to participate
The collection and usage of plant specimens in current study complied with the IUCN Policy Statement on Research Involving Species at Risk of Extinction and the Convention on the Trade in Endangered Species of Wild Fauna and Flora. Ethical approval was not applicable for this study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 16 January 2025 / Accepted: 15 July 2025
Published online: 19 August 2025

References

1. Drouin G, Daoud H, Xia JN. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. Mol Phylogenet Evol. 2008;49(3):827–31.
2. Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci U S A. 1987;84(24):9054–8.
3. Muse SV. Examining rates and patterns of nucleotide substitution in plants. Plant Mol Biol. 2000;42(1):25–43.
4. Bi CW, Sun N, Han FC, Xu KW, Yang Y, Ferguson DK. The first mitogenome of Lauraceae (*Cinnamomum chekiangense*). Plant Divers. 2023;46(1):144–8.
5. Hu HY, Sun PC, Yang YZ, Ma JX, Liu JQ. Genome-scale angiosperm phylogenies based on nuclear, plastome, and mitochondrial datasets. J Integr Plant Biol. 2023;65(6):1479–89.
6. Wang J, Kan SL, Liao XZ, Zhou JW, Tembrock LR, Daniell H, Jin SX, Wu ZQ. Plant organellar genomes: much done, much more to do. Trends Plant Sci. 2024;29(7):754–69.
7. Xiang KL, Wu SD, Lian L, He WC, Peng D, Peng HW, Zhang XN, Li HL, et al. Genomic data and ecological niche modeling reveal an unusually slow rate of molecular evolution in the cretaceous Eupteleaceae. Sci China Life Sci. 2023;67(4):803–16.
8. Hartman H, Fedorov A. The origin of the eukaryotic cell: a genomic investigation. Proc Natl Acad Sci U S A. 2002;99(3):1420–5.
9. Timmis JN, Ayliffe MA, Huang CY, Martin W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet. 2004;5(2):123–35.
10. Lang BF, Burger G. Mitochondrial and eukaryotic origins: a critical review. Adv Bot Res. 2012;63:1–20.
11. Martin WF, Garg S, Zimorski V. Endosymbiotic theories for eukaryote origin. Phil Trans R Soc Lond B Biol Sci. 2015;370(1678): 20140330.
12. Adams KL, Palmer JD. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol Phylogenet Evol. 2003;29(3):380–95.
13. Chen YM, Guo YW, Xie XM, Wang ZH, Miao LF, Yang ZZ, Jiao YN, Xie CJ, et al. Pangenome-based trajectories of intracellular gene transfers in Poaceae unveil high cumulation in triticeae. Plant Physiol. 2023;193(1):578–94.
14. Gualberto JM, Newton KJ. Plant mitochondrial genomes: dynamics and mechanisms of mutation. Annu Rev Plant Biol. 2017;68:225–52.
15. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17(1):134.
16. Palmer JD. Comparative organization of Chloroplast genomes. Annu Rev Genet. 1985;19:325–54.
17. Shaw J, Lickey EB, Schilling EE, Small RL. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am J Bot. 2007;94(3):275–88.
18. Wakasugi T, Tsudzuki T, Sugiura M. The genomics of land plant chloroplasts: gene content and alteration of genomic information by RNA editing. Photosynth Res. 2001;70(1):107–18.
19. Alverson AJ, Wei XX, Rice DW, Stern DB, Barry K, Palmer JD. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). Mol Biol Evol. 2010;27(6):1436–48.
20. Putintseva YA, Bondar EI, Simonov EP, Sharov VV, Oreshkova NV, Kuzmin DA, Konstantinov YM, Shmakov VN, et al. Siberian larch (*Larix sibirica* Ledeb.) mitochondrial genome assembled using both short and long nucleotide sequence reads is currently the largest known mitogenome. BMC Genomics. 2020;21(1):654.
21. Skippington E, Barkman TJ, Rice DW, Palmer JD. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all Nad genes. Proc Natl Acad Sci U S A. 2015;112(27):E3515-24.
22. Sloan DB. One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. New Phytol. 2013;200(4):978–85.
23. Zhang S, Wang J, He WC, Kan SL, Liao XZ, Jordan DR, Mace ES, Tao YF, et al. Variation in mitogenome structural conformation in wild and cultivated lineages of sorghum corresponds with domestication history and plastome evolution. BMC Plant Biol. 2023;23(1):91.
24. Kubo T, Newton KJ. Angiosperm mitochondrial genomes and mutations. Mitochondrion. 2008;8(1):5–14.
25. Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu YL, Song KM. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. Proc Natl Acad Sci U S A. 2000;97(13):6960–6.

26. Christensen AC. Plant mitochondrial genome evolution can be explained by DNA repair mechanisms. Genome Biol Evol. 2013;5(6):1079–86.
27. Knoop V. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. Curr Genet. 2004;46(3):123–39.
28. Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, et al. Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA: a primitive form of plant mitochondrial genome. J Mol Biol. 1992;223(1):1–7.
29. Keeling PJ. Horizontal gene transfer in eukaryotes: aligning theory with data. Nat Rev Genet. 2024;25:416–30.
30. Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. Mol Genet Genomics. 2002;268(4):434–45.
31. Stupar RM, Lilly JW, Town CD, Cheng ZK, Kaul S, Buell CR, Jiang JM. Complex MtDNA constitutes an approximate 620-kb insertion on *Arabidopsis thaliana* chromosome 2: implication of potential sequencing errors caused by large-unit repeats. Proc Natl Acad Sci U S A. 2001;98(9):5099–103.
32. Denk T, Grimm GW, Manos PS, Deng M, Hipp AL et al. An updated infrageneric classification of the oaks: review of previous taxonomic schemes and synthesis of evolutionary patterns. In: Gil-Pelegrín E, editors. Oaks Physiological Ecology. Exploring the Functional Diversity of Genus *Quercus* L. Tree Physiol. 2017;13–38.
33. Cavender-Bares J. Diversification, adaptation, and community assembly of the American oaks (*Quercus*), a model clade for integrating ecology and evolution. New Phytol. 2019;221(2):669–92.
34. Jin DM, Yuan Q, Dai XL, Kozlowski G, Song YG. Enhanced precipitation drove the evolution of subtropical evergreen broad-leaved forests in eastern China since the early miocene: evidence from ring-cupped oaks. J Syst Evol. 2023;62:677–86.
35. Kremer A. Microevolution of European temperate Oaks in response to environmental changes. C R Biol. 2016;339(7–8):263–7.
36. Fu RR, Zhu YX, Liu Y, Feng Y, Lu RS, Li Y, Li P, Kremer A, et al. Genome-wide analyses of introgression between two sympatric Asian oak species. Nat Ecol Evol. 2022;6(7):924–35.
37. Kapoor B, Jenkins J, Schmutz J, Zhebentyayeva T, Kuelheim C, Coggeshall M, Heim C, Lasky JR, et al. A haplotype-resolved chromosome-scale genome for *Quercus rubra* L. provides insights into the genetics of adaptive traits for red oak species. (Bethesda). 2023;G3(11):jkad209.
38. Li Y, Zheng SS, Wang TR, Liu MH, Kozlowski G, Yi LT, Song YG. New insights on the phylogeny, evolutionary history, and ecological adaptation mechanism in cycle-cup oaks based on chloroplast genomes. Ecol Evol. 2024;14(9): e70318.
39. Liu X, Zhang WX, Zhang YT, Yang J, Zeng P, Tian ZZ, Sun WB, Cai J. Chromosome-scale genomes of *Quercus sichourensis* and *Quercus Rex* provide insights into the evolution and adaptation of Fagaceae. J Genet Genomics. 2024;51(5):554–65.
40. Wang LL, Li Y, Zheng SS, Kozlowski G, Xu J, Song YG. Complete chloroplast genomes of four oaks from the section *Cyclobalanopsis* improve the phylogenetic analysis and understanding of evolutionary processes in the genus *Quercus*. Genes. 2024;15(2): 230.
41. Zhou YR, Li Y, Yang LH, Kozlowski G, Yi LT, Liu MH, Zheng SS, Song YG. The adaptive evolution of *Quercus* section *Ilex* using the Chloroplast genomes of two threatened species. Sci Rep. 2024;14:20577.
42. Bi QX, Li DX, Zhao Y, Wang MK, Li YC, Liu XJ, Wang LB, Yu HY. Complete mitochondrial genome of *Quercus variabilis* (Fagales, Fagaceae). Mitochondrial DNA B. 2019;4(2):3927–8.
43. Liu D, Guo HL, Zhu JL, Qu K, Chen Y, Guo YT, Ding P, Yang HP, et al. Complex physical structure of complete mitochondrial genome of *Quercus acutissima* (Fagaceae): a significant energy plant. Genes. 2022;13(8): 1321.
44. Huang CJ, Zhang YT, Bruce B. In: Chun WY, Huang CJ, editors. Fagaceae. Flora of China, Science Press and Missouri Botanical Garden; 1999. p. 380–400.
45. Song YG, Wang TR, Lu ZJ, Ge BJ, Zhong X, Li XC, Jin DM, Yuan Q, et al. Population survey combined with genomic-wide genetic variation unravels the endangered status of *Quercus gilva*. Diversity. 2023;15:230.
46. Carrero C, Jerome D, Beckman E, Byrne A, Coombes AJ, Deng M, Rodríguez AG, Sam HV, et al. The Red List of Oaks 2020. The Morton Arboretum. Lisle, IL: 2020.
47. Sugiura N, Tang DQ, Kurokochi H, Saito Y, Ide Y. Genetic structure of *Quercus Gilva* Blume in Japan as revealed by chloroplast DNA sequences. Botany. 2015;93(12):870–80.
48. Wang TR, Ning X, Zheng SS, Li Y, Lu ZJ, Meng HH, Ge BJ, Kozlowski G, et al. Genomic insights into ecological adaptation of Oaks revealed by phylogenomic analysis of multiple species. Plant Divers. 2025;47(1):53–67.
49. Zeng QM, Liu B, Lin RQ, Jiang YT, Liu ZJ, Chen SP. The complete Chloroplast genome sequence of *Quercus Gilva* (Fagaceae). Mitochondrial DNA B. 2019;4(2):2493–4.
50. Yang YC, Zhou T, Qian ZQ, Zhao GF. Phylogenetic relationships in Chinese oaks (Fagaceae, *Quercus*): evidence from plastid genome using low-coverage whole genome sequencing. Genomics. 2021;113(3):1438–47.
51. Doyle JJ. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 1987;19(1):11–5.
52. 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.
53. Wick RR, Judd LM, Gorrie CL, Holt KE, Unicycler. Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13(6):e1005595.
54. He WC, Xiang KL, Chen CJ, Wang J, Wu ZQ. Master graph: an essential integrated assembly model for the plant mitogenome based on a graph-based framework. Brief Bioinform. 2023;24(1): bbac522.
55. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 2015;31(20):3350–2.
56. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–100.
57. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. Geseq–versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45(W1):W6-11.
58. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647–9.
59. Rombel IT, Sykes KF, Rayner S, Johnston SA. ORF-finder: a vector for high-throughput gene identification. Gene. 2002;282(1–2):33–41.
60. Greiner S, Lehwark P, Bock R. Organellargenomedraw (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47(W1):W59-64.
61. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33(16):2583–5.
62. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80.
63. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. Reputer: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29(22):4633–42.
64. Wright F. The 'effective number of codons' used in a gene. Gene. 1990;87(1):23–9.
65. Edera AA, Small I, Milone DH, Sanchez-Puerta MV. Deepred-Mt: deep representation learning for predicting C-to-U RNA editing in plant mitochondria. Comput Biol Med. 2021;136: 104682.
66. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer L, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10: 421.
67. Chen CJ, Wu Y, Li JW, Wang X, Zeng ZH, Xu J, Liu YL, Feng JT, et al. TBtools-II: a one for all, all for one bioinformatics platform for biological big-data mining. Mol Plant. 2023;16(11):1733–42.
68. He WM, Yang J, Jing Y, Xu L, Yu K, Fang XD. NGenomeSyn: an easy-to-use and flexible tool for publication-ready visualization of syntenic relationships across multiple genomes. Bioinformatics. 2023;39(3): btad121.
69. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
70. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017;34(12):3299–302.
71. Zhang Z. KaKs_Calculator 3.0: calculating selective pressure on coding and non-coding sequences. Genom Proteom Bioinf. 2022;20(3):536–40.
72. Yong YS, Hu SX, Zhong MY, Wen Y, Zhou Y, Ma RX, Jiang XY, Zhang QS. Horizontal gene transfer from Chloroplast to mitochondria of seagrasses in the yellow-Bohai seas. Genomics. 2024;116(5): 110940.
73. Xiang CY, Gao FL, Jakovlić I, Lei HP, Hu Y, Zhang H, Zou H, Wang GT, et al. Using phylosuite for molecular phylogeny and tree-based analyses. Imeta. 2023;2(1): e87.

74. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-tree 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37(5):1530–4.
75. Guo HL, Liu Q, Chen Y, Niu HY, Zhao QR, Song H, Pang RD, Huang XL, et al. Comprehensive assembly and comparative examination of the full mitochondrial genome in *Castanea mollissima* Blume. Genomics. 2023;115(6): 110740.
76. Mader M, Schroeder H, Schott T, Schöning-Stierand K, Leite Montalvão AP, Liesebach H, Liesebach M, Fussi B, et al. Mitochondrial genome of *Fagus sylvatica* L. as a source for taxonomic marker development in the Fagales. Plants (Basel). 2020;9(10): 1274.
77. Mishra B, Ulaszewski B, Meger J, Ploch S, Burczyk J, Thines M. A comparison of three circular mitochondrial genomes of *Fagus sylvatica* from Germany and Poland reveals low variation and complete identity of the gene space. Forests. 2021;12(5): 12:571.
78. Qiu XY, Tian YQ, Li ZQ, Wu XJ, Xiang ZW, Wang YQ, Li J, Xiao SG. Assembly and characterization analysis of the complete mitochondrial genome of *Lithocarpus litseifolius* (Hance) Chun. Genet Resour Crop Evol. 2025;72(1):295–313.
79. Tu XD, Xin YX, Fu HH, Zhou CY, Liu QL, Tang XH, Zou LH, Liu ZJ, et al. The complete mitochondrial genome of *Castanopsis carlesii* and *Castanea henryi* reveals the rearrangement and size differences of mitochondrial DNA molecules. BMC Plant Biol. 2024;24(1):988.
80. Han Y, Feng YL, Wang J, Zhu SS, Jin XJ, Wu ZQ, Zhang YH. Comprehensive analysis of the complete mitochondrial genome of *Rehmannia chingii*: an autotrophic species in the Orobanchaceae family. Genes. 2024;15(1): 98.
81. Mower JP, Case AL, Floro ER, Willis JH. Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the mitochondrial genome from a Monkeyflower (*Mimulus guttatus*) lineage with cryptic CMS. Genome Biol Evol. 2012;4(5):670–86.
82. Wynn EL, Christensen AC. Repeats of unusual size in plant mitochondrial genomes: identification, incidence and evolution. G3 Genes|Genomes|Genetics. 2019;9(2):549–59.
83. Xia CY, Li JL, Zuo YW, He P, Zhang H, Zhang XX, Wang BR, Zhang JB, et al. Complete mitochondrial genome of *Thuja sutchuenensis* and its implications on evolutionary analysis of complex mitogenome architecture in Cupressaceae. BMC Plant Biol. 2023;23(1):84.
84. Ni Y, Li JL, Chen HM, Yue JW, Chen PH, Liu C. Comparative analysis of the chloroplast and mitochondrial genomes of *Saposhnikovia divaricata* revealed the possible transfer of plastome repeat regions into the mitogenome. BMC Genomics. 2022;23(1):570.
85. Cadorna CAE, Pahayo DG, Rey JD. The first mitochondrial genome of *Calophyllum soulattri* burm.f. Sci Rep. 2024;14(1):5112.
86. Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol. 2012;10(1):e1001241.
87. Binder S, Brennicke A. Gene expression in plant mitochondria: transcriptional and post-transcriptional control. Phil Trans R Soc Lond B. 2023;358(1429):181–9.
88. Hao W, Liu GX, Wang WP, Shen W, Zhao YP, Sun JL, Yang QY, Zhang YX, et al. RNA editing and its roles in plant organelles. Front Genet. 2021;12: 757109.
89. Small ID, Schallenberg-Rüdinger M, Takenaka M, Mireau H, Ostersetzer-Biran O. Plant organellar RNA editing: what 30 years of research has revealed. Plant J. 2020;101(5):1040–56.
90. Ichinose M, Sugita M. RNA editing and its molecular mechanism in plant organelles. Genes (Basel). 2017;8(1):5.
91. Guo LJ, Lao GR, He LF, Xiao D, Zhan J, Wang AQ. De novo assembly and comparative analysis of mitochondrial genomes of two *Pueraria montana* varieties. Int J Mol Sci. 2024;25(11): 5656.
92. Liu XY, Zhang DD, Yu ZF, Zeng B. Assembly and analysis of the complete mitochondrial genome of the Chinese wild dwarf almond (*Prunus tenella*). Front Genet. 2024;14: 1329060.
93. Li GL, Zhang H, Lin ZM, Li HW, Xu GC, Xu YQ, Ji RC, Luo WB, et al. Comparative analysis of chloroplast and mitochondrial genomes of sweet potato provides evidence of gene transfer. Sci Rep. 2024;14(1):4547.
94. Liu JT, Ni Y, Liu C. Polymeric structure of the *Cannabis sativa* L. mitochondrial genome identified with an assembly graph model. Gene. 2023;853: 147081.
95. Tang P, Ni Y, Li JL, Lu QQ, Liu C, Guo JL. The complete mitochondrial genome of *Paeonia lactiflora* Pall. (Saxifragales: Paeoniaceae): evidence of gene transfer from chloroplast to mitochondrial genome. Genes. 2024;15(2): 239.
96. Bi CW, Paterson AH, Wang XL, Xu YQ, Wu DY, Qu YS, Jiang AN, Ye QL, et al. Analysis of the complete mitochondrial genome sequence of the diploid cotton *Gossypium raimondii* by comparative genomics approaches. Biomed Res Int. 2016;2016:5040598.
97. Portugez S, Martin WF, Hazkani-Covo E. Mosaic mitochondrial-plastid insertions into the nuclear genome show evidence of both non-homologous end joining and homologous recombination. BMC Evol Biol. 2018;18(1):162–71.
98. Kleine T, Maier UG, Leister D. DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. Annu Rev Plant Biol. 2009;60:115–38.
99. Huang CY, Ayliffe MA, Timmis JN. Direct measurement of the transfer rate of Chloroplast DNA into the nucleus. Nature. 2003;422:72–6.
100. Kubo T, Mikami T. Organization and variation of angiosperm mitochondrial genome. Physiol Plant. 2007;129:6–13.
101. Palmer JD, Herbon LA. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. J Mol Evol. 1988;28(1–2):87–97.
102. Bi CW, Lu N, Xu YQ, He CP, Lu ZH. Characterization and analysis of the mitochondrial genome of common bean (*Phaseolus vulgaris*) by comparative genomic approaches. Int J Mol Sci. 2020;21(11): 3778.
103. Wu ZQ, Liao XZ, Zhang XN, Tembrock LR, Broz A. Genomic architectural variation of plant mitochondria—a review of multichromosomal structuring. J Syst Evol. 2020;60(1):160–8.
104. Wu ZQ, Stone JD, Štorchová H, Sloan DB. High transcript abundance, RNA editing, and small RNAs in intergenic regions within the massive mitochondrial genome of the angiosperm *Silene noctiflora*. BMC Genomics. 2015;16(1): 938.
105. Liu X, You Q, Liu MM, Bo C, Zhu YF, Duan YB, Xue JP, Wang DX, et al. Assembly and comparative analysis of the complete mitochondrial genome of *Pinellia ternata*. Funct Plant Biol. 2024;51(2): FP23256.
106. Yu XX, Feng YL, Zhang J. Characterization of the complete mitochondrial genome of wintersweet (*Chimonanthus praecox*) and comparative analysis within magnoliids. Life. 2024;14(2): 182.
107. Kozik A, Rowan BA, Lavelle D, Berke L, Schranz ME, Michelmore RW, Christensen AC. The alternative reality of plant mitochondrial DNA: one ring does not rule them all. PLoS Genet. 2019;15(8): e1008373.
108. Jackman SD, Coombe L, Warren RL, Kirk H, Trinh E, MacLeod T, Pleasance S, Pandoh P, et al. Complete mitochondrial genome of a gymnosperm, Sitka spruce (*Picea sitchensis*), indicates a complex physical structure. Genome Biol Evol. 2020;12(7):1174–9.
109. Guo S, Li ZY, Li CL, Liu Y, Liang XL, Qin YM. Assembly and characterization of the complete mitochondrial genome of *Ventilago Leiocarpa*. Plant Cell Rep. 2024;43(3):77.
110. Shan YY, Li JL, Zhang X, Yu J. The complete mitochondrial genome of *Amorphophallus albus* and development of molecular markers for five *Amorphophallus* species based on mitochondrial DNA. Front Plant Sci. 2023;14: 1180417.
111. Wang L, Liu X, Xu YJ, Zhang ZW, Wei YS, Hu Y, Zheng CB, Qu XY. Assembly and comparative analysis of the first complete mitochondrial genome of a traditional Chinese medicine *Angelica biserrata* (Shan et Yuan) Yuan et Shan. Int J Biol Macromol. 2024;257(1):128571.
112. Szandar K, Krawczyk K, Myszczyński K, Ślipiko M, Sawicki J, Szczecińska M. Breaking the limits - multichromosomal structure of an early eudicot *Pulsatilla patens* mitogenome reveals extensive RNA-editing, longest repeats and Chloroplast derived regions among sequenced land plant mitogenomes. BMC Plant Biol. 2022;22(1): 109.
113. Hao ZG, Zhang ZP, Zhang JN, Cui XF, Li JQ, Luo LX, Li YB. The complete mitochondrial genome of *Aglaia odorata*, insights into its genomic structure and RNA editing sites. Front Plant Sci. 2024;15: 1362045.
114. Lu GL, Zhang K, Que YX, Li YF. Assembly and analysis of the first complete mitochondrial genome of *Punica granatum* and the gene transfer from Chloroplast genome. Front Plant Sci. 2023;14: 1132551.
115. Hurst LD. The ka/ks ratio: diagnosing the form of sequence evolution. Trends Genet. 2002;18(9):486.
116. Zancani M, Braidot E, Filippi A, Lippe G. Structural and functional properties of plant mitochondrial F-ATP synthase. Mitochondrion. 2020;53:178–93.
117. Li HT, Luo Y, Gan L, Ma PF, Gao LM, Yang JB, Cai J, Gitzendanner MA, et al. Plastid phylogenomic insights into relationships of all flowering plant families. BMC Biol. 2021;19(1): 232.
118. The Angiosperm Phylogeny Group, Chase MW, Christenhusz MJM, Fay MF, Byng JW, Judd WS, et al. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc. 2016;181(1):1–20.

119. Deng M, Jiang XL, Hipp AL, Manos PS, Hahn M. Phylogeny and biogeography of East Asian evergreen oaks (*Quercus* section *Cyclobalanopsis*; Fagaceae): insights into the Cenozoic history of evergreen broad-leaved forests in subtropical Asia. Mol Phylogenet Evol. 2018;119:170–81.
120. Ortego J, Gugger PF, Riordan EC, Sork VL. Influence of climatic niche suitability and geographical overlap on hybridization patterns among Southern Californian oaks. J Biogeogr. 2014;41:1895–908.
121. Meng HH, Song YG, Hu GX, Huang PH, Li M, Fang OY, Su RP, Cao GL, et al. Evolution of East Asian subtropical evergreen broad-leaved forests: when and how. J Syst Evol. 2025. https://doi.org/10.1111/jse.70001.

Publisher's Note