Ornamental
Plant Research

ARTICLE

Open Access

https://doi.org/10.48130/opr-0025-0017
Ornamental Plant Research **2025**, 5: e019

Assembly and comparative analysis of the complete mitochondrial genome of *Yulania liliiflora*: an ornamental plant with high medicinal value

Yunhan Chen[1], Wei Liu[2], Jing Qiu[3*] and Changwei Bi[1*]

[1] *State Key Laboratory of Tree Genetics and Breeding, Co-Innovation Center for Sustainable Forestry in Southern China, Key Laboratory of Tree Genetics and Biotechnology of Educational Department of China, Key Laboratory of Tree Genetics and Silvicultural Sciences of Jiangsu Province, College of Ecology and Environment, Nanjing Forestry University, Nanjing 210037, China*
[2] *College of Optical, Mechanical and Electrical Engineering, Zhejiang A&F University, Hangzhou 311300, China*
[3] *Information Department, The First Affiliated Hospital of Naval Medical University, Changhai Road 168, Shanghai 200433, China*
* Corresponding authors, E-mail: power_ko@126.com; bichwei@njfu.edu.cn

Abstract

Yulania liliiflora is a highly exploitable plant with medicinal value from its flower buds, leaves, and bark. The species, as an ornamental plant, has significant implications for the phylogeny of land plants and offers promising prospects for scientific research. In this study, the *Y. liliiflora* mitochondrial genome (mitogenome) was assembled into a breach linear chromosome using PacBio HiFi sequencing, with a length of 865,191 bp and a GC content of 46.95%. A total of 69 genes were identified in the *Y. liliiflora* mitogenome. The *Y. liliiflora* mitogenome contains 65 unique genes, including 41 protein-coding genes, 21 tRNA genes, and three rRNA genes. Numerous repetitive sequences have been discovered, revealing the recombination events and the reason for the bigger mitogenome of *Y. liliiflora*. Many mitochondrial plastid sequences were found, with 12 complete chloroplast genes encompassed in these homologous fragments. Mitogenome from other Magnoliids were also used for collinearity comparison, which suggests a high recombination rate between *Y. liliiflora* and Magnoliids. Analysis of phylogeny has demonstrated that *Y. liliiflora* had the strongest genetic relation to *Magnolia biondii* of the Magnoliaceae, and Magnoliids emerged as a sister group to the clade encompassing monocots and eudicots. The results presented in this paper improve the understanding of the existing genetic knowledge of the genus *Yulania*. Simultaneously, they offer a promising avenue for conducting more comprehensive genomic investigations of *Y. liliiflora*.

Citation: Chen Y, Liu W, Qiu J, Bi C. 2025. Assembly and comparative analysis of the complete mitochondrial genome of *Yulania liliiflora*: an ornamental plant with high medicinal value. *Ornamental Plant Research* 5: e019 https://doi.org/10.48130/opr-0025-0017

Introduction

Native to the southern regions of China, *Yulania liliiflora* is often associated with classical Chinese poetry and art, where it is depicted as a metaphor for feminine beauty and grace. It has been increasingly introduced to urban centers in Europe and the United States since 1780. This species is not only of high ornamental value as a landscape plant but also has excellent medicinal properties[1]. The flower buds of *Y. liliiflora* have been used in Chinese traditional medicine for over 2,000 years. Previous research has confirmed that the leaves of *Y. liliiflora* could be considered a highly promising by-product and affordable source of lignans for the pharmaceutical, food, and agricultural industries[2]. Additionally, research has found that the essential oil of *Y. liliiflora* exerts a positive effect in the treatment of pregnant women who have decubitus ulcers[3]. The utilization of essential oil and leaf extracts sourced from *Y. liliiflora* stands as an exceptional strategy for hindering the proliferation of food spoilage microorganisms and foodborne pathogens[4].

Magnoliids are one of the largest clades of flowering plants, second only to the eudicots and monocots, which have a global distribution and are highly valued for their economic, ornamental, and ecological attributes. Magnoliids, positioned as the third-largest branch with a species count exceeding 10,000, represent a minority, contributing to less than 3% of the overall angiosperm species[5]. As a dicotyledonous plant family featuring primitive flowers, Magnoliaceae is among the largest families within the Magnoliids[6]. Within Magnoliaceae, species like *Y. liliiflora* represent early diverging lineages, playing a crucial role in advancing our understanding of plant evolution and phylogeny. This information aids in unraveling the evolution route of present-day angiosperms. Previous studies have provided an overview of the phylogenetic research on Magnoliids and explored the possible reasons for the differences in their evolutionary status, such as diversity in phylogenetic data, diversity of sampling, variations in methodology, Long-Branch attraction, and incomplete lineage sorting[5,7]. Nevertheless, achieving a comprehensive understanding of the underlying molecular mechanisms has proven challenging, mainly attributable to the insufficient availability of critical genetic resources. Specifically, there is a dearth of reported mitogenomes for *Y. liliiflora*.

Higher plant cells inherently possess semi-autonomous genetic systems, specifically, the mitochondrial and chloroplast genomes, which function as vehicles for crucial genetic information[8,9]. Primarily serving as the cellular 'energy factories', mitochondria are responsible for converting biomass energy into chemical energy in living cells[10]. Mitochondrial genomes (mitogenomes), owing to the polymorphism exhibited within their genomes, which encompasses a spectrum of genetic variations such as single nucleotide polymorphisms, insertions, deletions, and gene rearrangements, serve as valuable resources for studying the mechanisms of origin and maintenance of genetic diversity in plants[11]. However, owing to the existence of numerous repetitive sequences and mitochondrial plastid sequences (MTPTs), the assembly of plant mitogenomes is often fragmented, making it difficult to obtain a complete mitogenome. To December, 2023, the National Center for Biotechnology Information (NCBI) database has documented nearly 13,000 complete plastid genomes, and a limited 673 plant mitogenomes[12]. Ongoing research highlights the diverse structural forms of plant

© The Author(s)

mitogenomes, including circular molecules, linear conformations, branched structures, and polycyclic molecules[13–15]. The size of plant mitogenomes varies significantly, spanning from 66 kb (*Viscum scurruloideum*) to ~19 Mb (*Cathaya argyrophylla*)[16,17], with the majority falling within the range of 200–800 kb[18]. The gene contents of angiosperm mitogenomes exhibit considerable variation, with notable conservation observed in 24 core protein-coding genes (PCGs), whereas other variable genes display substantial diversity among diverse species[19]. Spermatophyte mitogenomes are rich in repetitive sequences, which include simple sequence repeats (SSRs), tandem repeats, and dispersed repeats. Large repeats greater than 1,000 base pairs (bp) in length and displaying high sequence similarity are typically more prone to frequent recombination, whereas medium-sized repeats spanning 100 to 1,000 bp recombine intermittently, and small repeats under 100 bp are rare[20–22].

In this study, we assembled the complete mitogenome of *Y. liliiflora* using the HiFi sequencing data. We further annotated and analyzed the *Y. liliiflora* mitogenome. To gain a deeper insight into the evolution and origin of the *Y. liliiflora* mitogenome, available data from the other three Magnoliids mitogenomes were used for comparison. The results presented in this study supplement the existing genetic knowledge relating to the genus *Yulania*. Simultaneously, they offer a promising avenue for conducting more comprehensive genomic investigations of *Y. liliiflora*.

Materials and methods

Plant material processing and genome sequencing

Fresh leaves of *Y. liliiflora* were collected from the campus of Nanjing Forestry University, Nanjing, Jiangsu Province, China (32°04'41" N, 118°48'23" E). The leaves were promptly frozen and preserved in liquid nitrogen at a temperature of −80 °C. Total genomic DNA was extracted using the Hi-DNAsecure Plant Kit (Tiangen DP350). After the integrity and purity of the isolated DNA were checked by agarose gel electrophoresis and a Nanodrop 2000 ultraviolet spectrophotometer (ThermoFisher), a sequencing library was constructed by the SMRTbell Express Template Prep Kit 2.0 (PacBio Biosciences, CA, USA). The PacBio HiFi reads, achieved through the application of Circular Consensus Sequencing, exhibit an error rate of 1% or below, ensuring a high degree of accuracy[23,24]. HiFi sequencing data were obtained from the PacBio Revio platform[25].

Assembly and annotation of the *Y. liliiflora* mitogenome

The mitogenome of *Y. liliiflora* was assembled using two distinct approaches. Initially, the 'autoMito' mode of PMAT v1.3.0 was used for mitogenome assembly, with the parameters '-st HiFi -g 2200M'[26]. An alternative approach involved manually selecting three contigs (contig01680, contig00761, and contig114302) from the output file of the PMAT's 'autoMito' mode, based on their congruence with the depth of the mitogenome sequencing. Validation of the PMAT assembly was performed using oatk with the parameters '-k 1000 -c 5'[27]. Consequently, we acquired a preliminary draft mtDNA, a complex multibranched and closed-loop conformation (Supplementary Fig. S1). For the convenience of description, we manually simplified the graphical representation utilizing bandage v0.8.1[28]. During this simplification process, we excluded the chloroplast nodes. Simultaneously, we retained a subset of chloroplast nodes specifically for the analysis of MTPTs (Supplementary Fig. S1). Finally, we processed *Y. liliiflora* mitogenome into a linear molecule (Supplementary Fig. S1). Of course, we emphasize that the treatment is not the only form because the configuration of mitochondrial DNA is in dynamic transformation mediated by repeats. The mitogenomes of *Y. liliiflora* were initially annotated using the web-based tool IPMGA (https://www.lkmpg.cn/mgavas). Subsequently, the tRNA and rRNA genes were further examined using tRNAscan-SE v2.0[29] and BLASTn[30], respectively. The intron contents of the *Y. liliiflora* mitogenome were identified by PMGmap (http://www.1kmpg.cn/pmgview). Manual verification was performed on all PCGs, tRNA, and rRNA genes, and intron contents using MacVector v18.5 (https://macvector.com/). Finally, the mitogenome map of *Y. liliiflora* was constructed using OGDraw vl.3.1, an online tool provided by Greiner et al.[31].

Analysis of relative synonymous codon usage (RSCU) and repeat elements

The RSCU for PCGs was analyzed and calculated using Mega 11.0 software[32]. To identify the SSRs in the *Y. liliiflora* mitogenome, the MISA web service (https://webblast.ipk-gatersleben.de/misa/) was utilized[33]. For mono-, di-, tri-, tetra-, penta-, and hexa-nucleotides, minimum repetition thresholds were established at 10, 5, 4, 3, 3, and 3 respectively[34]. Furthermore, a maximum sequence length of sequence between two SSRs was set at 100 bp. The tandem Repeats Finder v4.09 software was utilized to detect any tandem repeats present in the assembled mitogenome[35]. The minimum alignment score was established as 60, while the maximum period size was defined as 500. The identification of dispersed repeats within the *Y. liliiflora* mitogenome was conducted using the online program REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer)[36]. The analysis considered a hamming distance of 3, a minimal repeat size of 50, and a maximum computed repeat count of 5,000. The distribution of dispersed repeats in the *Y. liliiflora* mitogenome was visualized using the Advanced Circos module in TBtools vl.132[37].

Identification of the MTPTs

To facilitate the identification of MTPTs, the chloroplast genome of *Y. liliiflora* was extracted from the whole genome sequencing data[38]. BLASTn analysis was subsequently conducted to detect homologous fragments between the plastid and mitogenomes of *Y. liliiflora*[30]. The MTPTs with matching rates > 90% and lengths > 50 bp were selected for further investigation. All MTPTs were manually annotated to examine the genes. Subsequently, the MTPTs were visualized using the Advanced Circos module in TBtools v1.132[37].

Identification of RNA editing sites

RNA sequencing has provided profound insights into the intricate workings of complex biological mechanisms[39]. The diversification of PCGs sequences in plant mitochondria is intricately linked to the crucial role of RNA editing[40]. In the *Y. liliiflora* mitogenome, we utilized RNA sequences to assess the prevalence of RNA editing. Specific steps involved the extraction of coding sequences for each PCG, along with 100 bp flanking regions, to systematically construct reference sequences. Subsequently, we used Deepred-mt for the identification of RNA editing sites[41]. Finally, manual confirmation of all identified RNA editing sites was conducted using the IGV software[42].

Whole-genome collinearity analysis

To compare the mitogenome structure of *Y. liliiflora* with other plant species, we downloaded another four closed mitogenomes from NCBI for collinearity analysis including *M. officinalis* (NC_064401.1), *Liriodendron tulipifera* (NC_042758), *Hernandia nymphaeifolia* (NC_063145.1), and *Cinnamomum chekiangense* (NC_082065.1). The program of BLASTn was used to align the four mitogenomes against the *Y. liliiflora* mitogenome with a minimum identity > 90% and minimum alignment length > 500 bp[30]. The R package

RIdeogram for chromosome data visualization was used to display colinear results generated by BLASTn[43].

Phylogenetic analysis

To establish a comparative framework aligned with the kinship of *Y. liliiflora,* we obtained the mitochondrial PCGs of 16 additional plant mitogenomes from the NCBI Nucleotide database. *Cycas taitungensis* and *Ginkgo biloba* were chosen as outgroups. A total of 22 conserved PCGs (*atp1, atp4, atp6, atp8, atp9, ccmB, ccmC, cob, coxI, cox2, cox3, mttB, nad3, nad4, nad4L, nad5, nad6, nad7, nad9, rps3, rps4,* and *rps12*) were meticulously selected for mitochondrial multiple sequence alignment, executed by Prank v.170427[44]. For chloroplasts multiple sequence alignment, 45 conserved PCGs (*atpA, atpB, atpE, atpF, atpI, ccsA, cloP, matK, ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhI, ndhJ, ndhK, petA, petB, psbC, psbD, rbcL, rpl14, rpl20, rpoA, rpoB, rpoC1, rpoC2, rps2, rps4, rps7, rps8, rps11, rps12, rps18, rps19, ycf3*, and *ycf4*). Finally, IQ-TREE v 2.1.4 were used to construct the phylogenetic tree[45]. The maximum likelihood (ML) method was implemented under the GTR + F + I + G4 model chosen for the chloroplast and mitochondrial genome. The online tool iTOL was used to meticulously edit and decorate the phylogenetic tree[46].

Results

Features of the *Y. liliiflora* mitogenome

The *Y. liliiflora* breach linear mitogenome was assembled, annotated, and submitted to the GenBank nucleotide database with an accession number of NC_085212.1 (Supplementary Fig. S1). The annotated genes' physical arrangements and functional categorizations are presented in Fig. 1a. The *Y. liliiflora* mitogenome is 865,191 bp in length, resembling the size of most other previously assembled Magnoliids plant mitogenomes. In the *Y. liliiflora* mitogenome, the nucleotide composition stands at 26.59% A, 26.45% T, 23.40% G, and 23.56% C, with a GC content of 46.95% (Supplementary Table S1). The GC content is higher than the *Y. liliiflora* chloroplast genome (39.25%)[47]. The total length of PCGs constitutes 4.24% (36,717 bp) of the complete mitogenome, while the lengths of tRNA and rRNA genes account for merely 0.24 and 0.68%, respectively. As illustrated in Fig. 1b, the GC content of PCGs, tRNA, and rRNA genes is 43.14%, 48.63%, and 52.99%, respectively. A total of 65 unique genes were identified in the *Y. liliiflora* mitogenome, including 41 PCGs (24 core genes and 17 variable genes), 21 tRNA, and three rRNA genes (Table 1). It is found that there are two copies of *atp4, nad3*, *rps7*, and *rps12*, respectively (Table 1). In addition, there are 17 *cis*-splicing introns in seven genes (*nad4*, *nad7*, *ccmFc*, *cox2*, *rpl2*, *rps3*, and *rps10*), and five *trans*-splicing introns in *nad1*, *nad2*, and *nad5*, respectively (Supplementary Fig. S2).

Codon usage analysis of PCGs

The total length of PCGs in *Y. liliiflora* is 36,717 bp encoding 12,239 codons excluding termination codons. The analysis of codon usage was performed on 45 PCGs in the mitogenome of *Y. liliiflora*, with the codon usage amino acid displayed in Fig. 2. This analysis reveals preferential codon usage patterns that could aid in understanding the underlying evolutionary mechanism in plants. It sheds light on potential selective pressures and evolutionary influences on mitochondrial protein synthesis in *Y. liliiflora*[48,49]. Previous research has strongly demonstrated that mutational bias, translational selection, and gene functional bias should work on affecting the RSCU of single-exon genes in *Oryza sativa*[49,50]. RSCU values exceeding 1 are indicative of amino acid preferences, while values below 1 suggest underutilization of the codon[51]. Notably the initial codons AUG (Methionine) and UGG (Tryptophan), both with RSCU values of 1. The PCGs of *Y. liliiflora* mitogenome display a common codon preference. GUU (Valine) possesses the highest RSCU value at 1.46, followed by GGA (Glycine) at 1.44 (Supplementary Table S2). However, cysteine (Cysteine), lysine (Lysine), and phenylalanine

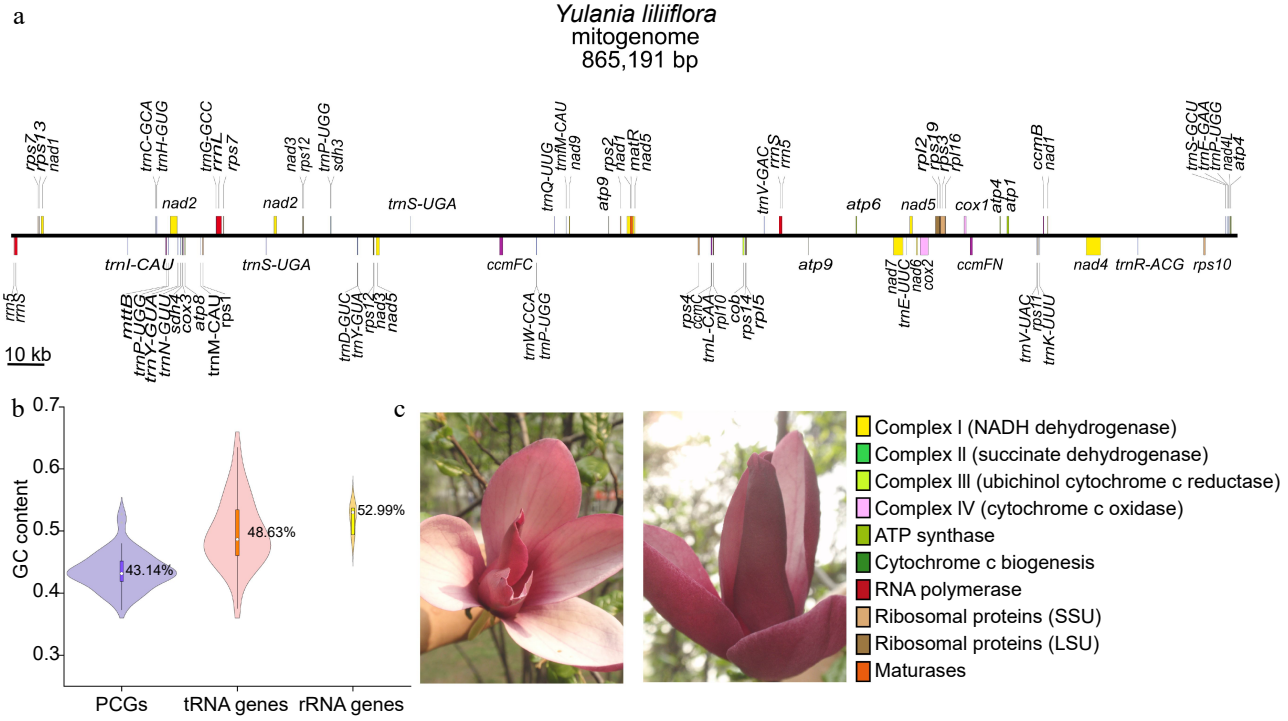

Fig. 1 Summary of the mitogenome of *Y. liliiflora*. (a) Linear map of the mitogenome of *Y. liliiflora*. The genes in the upper and lower linear regions are transcribed clockwise and counterclockwise, respectively. (b) The GC content of PCGs, tRNA, and rRNA genes among *Y. liliiflora* mitogenomes. (c) *Y. liliiflora* in Nanjing Forestry University (Nanjing, China).

Table 1. The physical placements and functional categorizations of the *Y. liliiflora* mitochondrial genome (mitogenome).

Group of genes	Gene name	Length (bp)	Start codon	Stop codon	Amino acids
ATP synthase	*atp1*	1,530	ATG	TGA	509
	atp4 × 2	582	ATG	TAA	193
	atp6	720	ATG	CAA	239
	atp8	480	ATG	TAA	159
	atp9	225	ATG	TAA	74
NADH dehydrogenase	*nad1*$^{**++}$	978	ACG	TAA	325
	nad2^{***+}	1,467	ATG	TAA	488
	nad3 × 2	357	ATG	TAA	118
	nad4***	1,488	ATG	TGA	495
	nad4L	303	ACG	TAA	100
	nad5^{**++}	2,013	ATG	TAA	670
	nad6	735	ATG	TGA	244
	nad7****	1,185	ATG	TAG	394
	nad9	573	ATG	TAA	190
Cytohrome c biogenesis	*ccmB*	621	ATG	TGA	206
	ccmC	960	ATG	TAA	319
	ccmFc*	1,329	ATG	CGA	442
	ccmFn	1,806	ATG	TAG	601
Maturases	*matR*	1,959	ATG	TAG	652
Ubichinol cytochrome c reductase	*cob*	1,182	ATG	TGA	393
Cytochrome c oxidase	*cox1*	1,584	ACG	TAA	527
	cox2**	765	ATG	TAA	254
	cox3	798	ATG	TGA	265
*Trans*port membrance protein	*mttb*	768	ACG	TGA	255
Ribosomal proteins (LSU)	*rpl2**	1,665	ATG	TAG	554
	rpl5	561	ATG	TAA	186
	rpl10	471	ATG	TAA	156
	rpl16	435	GTG	TAA	144
Ribosomal proteins (SSU)	*rps1*	693	ATG	TAG	230
	rps2	657	ATG	TAA	218
	rps3*	1,572	ATG	TAA	523
	rps4	1,071	ACG	TAA	356
	rps7 × 2	450	ATG	TAA	149
	rps10*	360	ACG	TGA	119
	rps11	516	ATG	CAA	171
	rps12 × 2	378	ATG	TGA	125
	rps13	351	ATG	TGA	116
	rps14	303	ATG	TAG	100
	rps19	282	ATG	TAA	93
Succinate dehydrogenase	*sdh3*	330	ATG	TAA	109
	sdh4	447	ATG	TGA	148
Transfer RNAs	*trnC-GCA*	71	–	–	–
	trnD-GUC	74	–	–	–
	trnE-UUC	72	–	–	–
	trnF-GAA	74	–	–	–
	trnfM-CAU	74	–	–	–
	trnG-GCC	72	–	–	–
	trnH-GUG	74	–	–	–
	trnI-CAU	74	–	–	–
	trnK-UUU	73	–	–	–
	trnL-CAA	74	–	–	–
	trnM-CAU	73	–	–	–
	trnN-GUU	72	–	–	–
	trnP-UGG	75	–	–	–
	trnQ-UUG	72	–	–	–
	trnR-ACG	79	–	–	–
	trnS-GCU	88	–	–	–
	trnS-UGA	87	–	–	–
	trnV-GAC	72	–	–	–
	trnV-UAC	73	–	–	–
	trnW-CCA	74	–	–	–
	trnY-GUA	83	–	–	–
Ribosomal RNAs	*rrn5*	118	–	–	–
	rrnL	3,710	–	–	–
	rrnS	2,087	–	–	–

The numbers following the gene names indicate the number of copies. * Indicates the number of *cis*-splicing genes, while $^{+}$ indicates the number of *trans*-splicing genes.

(Phenylalanine) exhibit RSCU values below 1.2, indicating minimal codon preferences (Supplementary Table S2). The results of this study indicate that A/U ending codons are commonly overabundant, while G/C ending codons are frequently under represented. Codon usage analysis of the *Y. liliiflora* mitogenome reveals leucine (Leucine), serine (Serine), and arginine (Arginine) as the most prevalent amino acids, while methionine (Methionine) and tryptophan (Tryptophan) are less common (Fig. 3).

Repetitive sequences of the *Y. liliiflora* mitogenome

Based on sequence characteristics, repetitive sequences are categorized into three types: simple repetitive sequences, tandem, and dispersed repeats. In the *Y. liliiflora* mitogenome, a total of 303 SSRs are found, with monomeric and dimeric SSRs occupying 43.23% of the total SSRs (Fig. 4a, Supplementary Table S3). Among the 68 monomer SSRs, it should be emphasized that adenine (A) and thymine (T) monomer repeats constitute 30.88% and 44.12% respectively, while TA repeats are the most frequent dimeric types, comprising 20.63%. Tetrameric SSRs are the most common, representing 34.65% of the total SSRs. However, only seven hexamer SSRs were observed in the mitogenome. In the mitogenomes of *M. biondii*, *M. officinalis*, and *L. tulipifera*, we identified 273, 327, and 234 SSRs respectively. The prevalent SSR type is tetranucleotide, while the infrequent SSR type is hexanucleotide.

Tandem repeats, are a common phenomenon in eukaryotic genomes, with sporadic instances, also observed in certain prokaryotic genomes. In the mitogenome of *Y. liliiflora*, we identified a total of 31 tandem repeats, spanning a length range of 31 to 127 bp, and exhibiting a high degree of sequence similarity, exceeding 90%. As depicted in Fig. 4b, the *Y. liliiflora* mitogenome exhibits 328 pairs of dispersed repeats, each having a length equivalent to, or exceeding 50 bp. The analysis revealed 164 palindromic repeat pairs and 161 forward repeat pairs, whereas no complement or reverse repeats were detected among them (Supplementary Table S4). Among them, there are 25 dispersed repeats of more than 500 bp (Supplementary Table S5), and the longest forward repeat is 48,270 bp, whereas the longest palindromic repeat is 105,147 bp.

Migration sequences of *Y. liliiflora*

The transfer of genes from chloroplasts to mitochondrial genomes is a prevalent phenomenon observed during the extended evolution of angiosperm plants[52,53]. The mitogenome of *Y. liliiflora* contains 35 DNA fragments homologous to cp genomes (Fig. 5, Supplementary Table S6). The length of these fragments is 37,517 bp and constitutes 4.34% of the total length of the mitogenome. These findings substantiate the inference that mitochondrial DNA can be attributed to the chloroplast genome, constituting

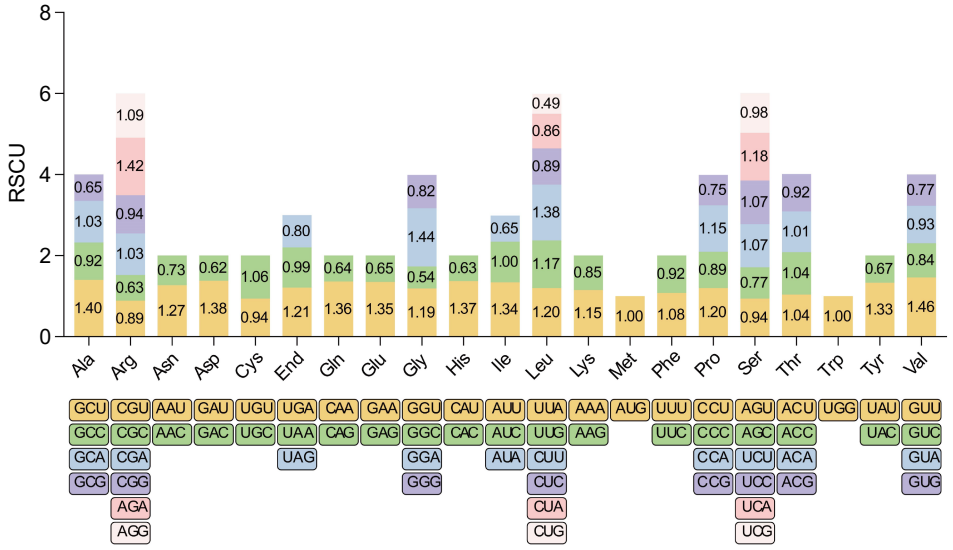

Fig. 2 RSCU in the *Y. liliiflora* mitogenome. The x-axis of the graphical representation depicts codon families, while RSCU values on the y-axis signify the frequency of codons relative to the expected frequency for uniform synonymous codon usage. This analysis reveals preferential codon usage patterns, shedding light on potential selective pressures and evolutionary influences on mitochondrial protein synthesis in *Y. liliiflora*.

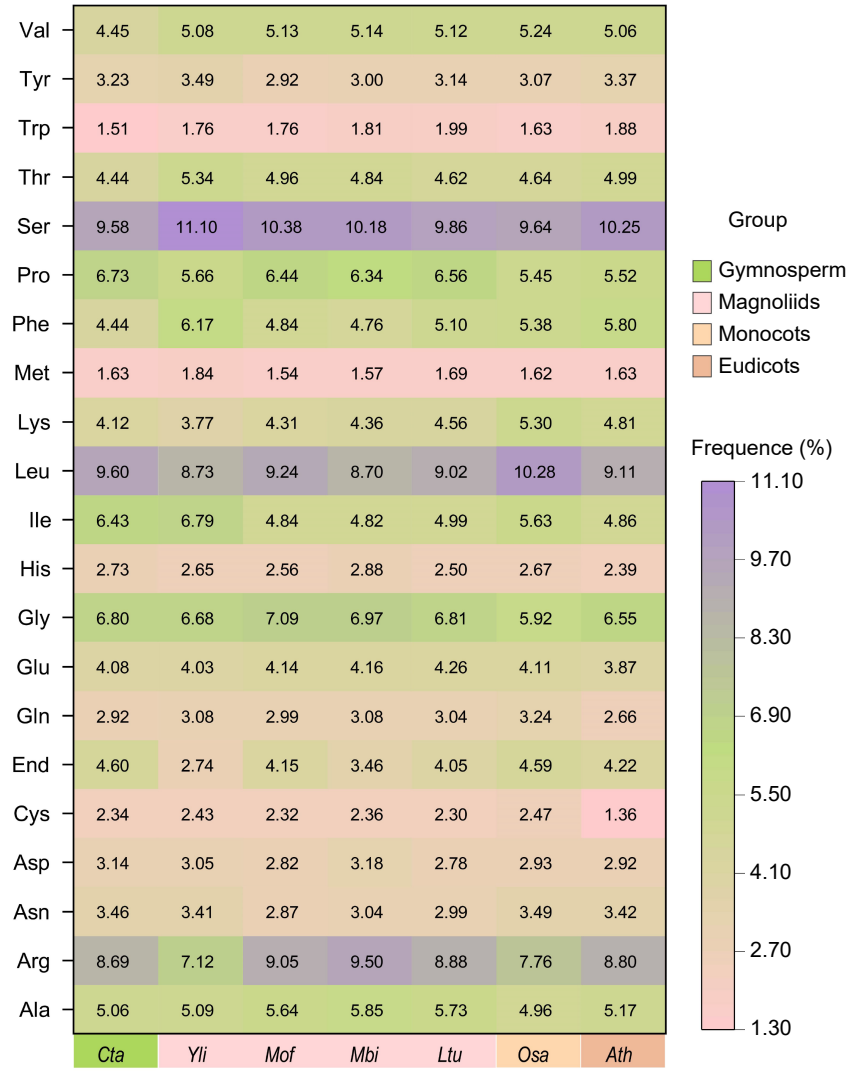

Fig. 3 Codon usage frequence of the *Y. liliiflora* mitogenome compared with *C. taitungensis*, *M. officinalis*, *M. biondii*, *L. tulipifera*, *O. sativa*, and *A. thaliana*. The x-axis represents different species, while the y-axis depicts amino acid families. Color intensity reflects the percentage of each codon relative to the total codons coding for a specific amino acid, ranging from pink (1.30%) to deep purple (11.10%). The gradual color *trans*ition signifies increasing percentages, providing a visual representation of the relative abundance of each codon in the context of all mitochondrial proteins.

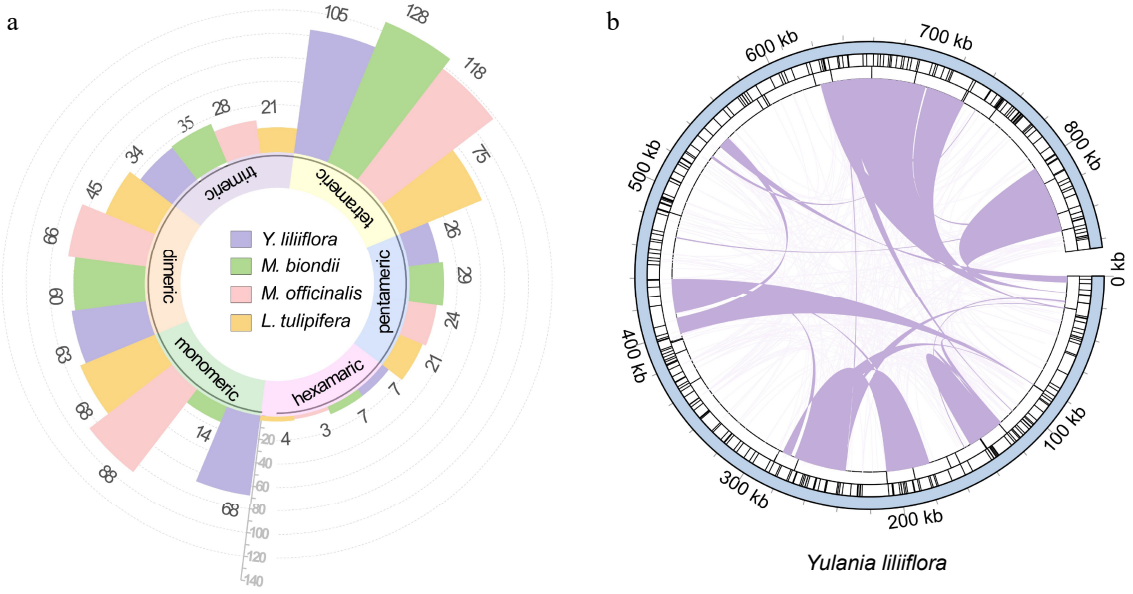

Fig. 4 Repetitive sequences of the *Y. liliiflora* mitogenome. (a) Ribbon bar graphs presenting SSRs. The legend is in purple, green, pink, and yellow for *Y. liliiflora*, *M. biondii*, *M. officinalis*, and *L. tulipifera* respectively, with every four bars clustered into one category starting from the 0 scale, representing monomers, dimers, trimers, tetramers, pentamers, and hexamers. (b) The repeats of *Y. liliiflora* mitogenome. The color line on the inner circle connects two repeated dispersed repeat. The deep purple lines represent ≥ 500 bp repetitive fragments. The tandem repeat is represented by short bars in the intermediate circle, while the SSRs are shown in the outer circle.

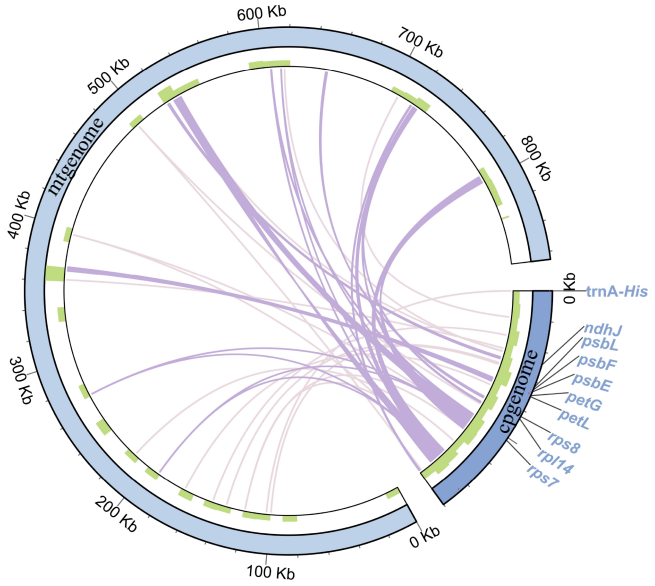

Fig. 5 Migration sequences of the *Y. liliiflora* mitogenome. The pale purple lines on the inner circle represent fragments that transfer from chloroplasts to mitochondria. The diagram displays transfer fragments of ≥ 100 bp as deep purple lines, while the sequence length is shown on the outside circle.

approximately 1%–10.3%[54]. Among them, there are a total of 19 MTPTs that are larger than 100 bp. The longest fragment is 5,866 bp. There are 10 complete chloroplast genes be encompassed in these homologous fragments, including nine PCGs (*ndhJ*, *petG*, *petL*, *psbE*, *psbF*, *psbL*, *rpl14*, *rps7*, and *rps8*), and one tRNA gene (tRNA-*His*).

Prediction of RNA editing sites in PCGs

Within plant systems, RNA editing is crucial for modulating gene expression, specifically through the conversion of cytidine (C) to uridine (U) in mitogenomes[55]. By utilizing the Deepred-mt software to analyze the variations observed between the DNA and RNA sequences of the *Y. liliiflora* mitogenome, we predicted 894 potential C-U RNA editing sites distributed across 45 PCGs. Figure 6 shows that the *nad4* gene has 68 identified editing sites, establishing it as the most extensively edited gene. In close succession, the *nad5* gene displays 48 instances of RNA editing, while the *rpl2* and *rps14* genes exhibit only two editing events each. Most PCGs start with the typical ATG codon, although specific genes such as *nad1*, *nad4L*, *cox1*, *mttB*, *rps4*, and *rps10* initiate with ACG. This phenomenon is plausibly a result of C to U RNA editing taking place at the second site. It should be noted that *rpl16* uses GTG as its start codon, given that the actual start codon has not yet been definitively determined. As for PCGs, four distinct termination codons were identified: TGA, TAA, CAA, and CGA. This can be attributed to the observed phenomenon of C-to-U RNA editing, which is prevalent among the genes *atp6*, *ccmFc*, and *rps11*.

As shown in Fig. 6, it exhibits similar editing site numbers of PCGs in Magnoliids. For example, *Y. liliiflora*, *M. biondii*, and *L. tulipifera* contained 894, 734, and 834 C-U editing sites, respectively. The PCG copies of *M. officinalis* are notably abundant, resulting in the inclusion of 1184 C-U editing sites. The number of RNA editing sites exhibits a gradual decline, progressing from gymnosperms to angiosperms, from Magnoliids to monocots, and further to eudicots. Additionally, specific variable genes exhibit an absence of editing sites in *A. thaliana*, *O. sativa*, and *C. taitungensis*. The RNA editing events in plant mitogenomes result in alterations to start and stop codons, potentially influencing mitochondrial function and cellular metabolism. However, due to the lack of transcriptome data, the specific impact on *Y. liliiflora* remains to be further elucidated.

Phylogenetic and synteny analysis

To elucidate the phylogenetic classification of *Y. liliiflora*, we retrieved 16 plant mitogenomes and chloroplast genomes from GenBank. Phylogenetic trees were constructed using 22 mitochondrial PCGs derived from 17 species, spanning 12 orders of angiosperms (Amborellales, Nymphaeales, Austrobaileyales, Laurales, Magnoliales, Alismatales, Arecales, Poales, Malpighiales, Brassicales,

Fig. 6 Bubble plot of the number of RNA editing sites in mitogenome PCGs of *C. taitungensis*, *Y. liliiflora*, *M. officinalis*, *M. biondii*, *L. tulipifera*, *O. sativa*, and *A. thaliana*.

Solanales, and Apiales) and two orders of Gymnospermae (Cycadales, Ginkgoales). *G. biloba* and *C. taitungensis* are designated as outgroups (Supplementary Tables S7 & S8). As depicted in Fig. 7, every node within the constructed tree boasts bootstrap support values surpassing 80%, and it is noteworthy that 10 nodes, specifically, enjoy a unanimous support of 100%. The phylogenetic trees originated from the chloroplast and mitogenomes align with the latest Angiosperm Phylogenetic Group (APG IV) classification[6]. The phylogenetic results provide robust support for the grouping of *Y. liliiflora*. The phylogenetic tree provides support for situating *Y. liliiflora* within the Magnoliids and delineates its associations with eudicots and monocots.

The study identified and analyzed mitogenome collinearity among five Magnoliid species with a matching degree greater than 90 and a length of more than 500 bp. To analyze the structural collinearity of *Y. liliiflora*, it was compared with *M. officinalis*, *L. tulipifera*, *C. chekiangense*, and *H. nymphaeifolia*. The principal factor driving the evolution of plant mitogenomes is frequent rearrangement. As depicted in Fig. 8, between the mitogenomes of *Y. liliiflora* and *M. officinalis*, a total of 96 local colinear blocks are present, accounting for 95.78% (828,710 bp) of the whole *Y. liliiflora* mitogenome. Between the mitogenomes of *Y. liliiflora* and *L. tulipifera*, we detected 91 local colinear blocks (total length: 443,814 bp; 51.30% of the mitogenome). Comparing the mitogenome of *Y. liliiflora* with those of *C. chekiangense* and *H. nymphaeifolia*, the colinear sequences only accounts for 27.55% and 23.26% of *Y. liliiflora* whole mitogenome, respectively. The colinear length and ratio of *C. chekiangense*, and *H. nymphaeifolia* are lower than those of the other two species (*M. officinalis* and *L. tulipifera*) of Magnoliids. While the degree of collinearity remains significant, it is noteworthy that the complex distribution of collinear segments across the contigs exhibits considerable diversity in sequence orientation among the various species. However, the study detected numerous similar segments between *Y. liliiflora* and the four species, indicating a distinct relationship (Fig. 8).

Discussion

The size of the *Y. liliiflora* mitogenome

The complete mitogenome of *Y. liliiflora* was assembled into a breach linear structure, with 865,191 bp in length (GenBank accession number: NC_085212.1), reflecting the average size of Magnoliids mitogenomes. Notably, within the Magnoliid clade, mitogenome sizes exhibit substantial variability, extending from a minimum of 535,805 bp in *H. ymphaeifolia* to a maximum of 967,100 bp in *M. biondii*[56]. However, the size of most mitogenomes of angiosperms ranges from 200 to 800 kb, with Magnoliid being on the large side[18]. Research into structure and size in plant mitogenomes is important because the evolution of plant mitogenomes involves notable episodes of breakage and rejoining, accompanied by DNA sequence gain and loss, more likely driven by non-adaptive selection[57–59]. During this process, repeats, and MTPTs may occur. Previous studies observed repeats within vascular plants tend to be larger and occur with greater frequency[59,60], such as four larger repeats (63.9, 10.6, 9.1, and 2.5 kb) in *Gossypium raimondii* mitogenome, and intramolecular recombination in the evolution of higher plants has been identified as an active process[61]. Meanwhile, some large repeats (> 1 kb) were identified in *Y. liliiflora* mitogenome, and this phenomenon was also reported in *L. tulipifera* (Magnoliids)[59,62]. Research has indicated that the presence of repeats in vascular plants is related to the size and complexitity of mitogenomes[59]. In addition, MTPT also has a significant impact on the size of the mitogenome. For example, *Amborella trichopoda* has acquired foreign mitochondrial DNA from sources such as green algae, mosses, and other angiosperms[63]. Nevertheless, further examination is necessary to explore their potential influence on the phenotypic traits of *Y. liliiflora*.

Structure and intron contents of the *Y. liliiflora* mitogenome

Ever since the initial discovery of the mitogenome in the land-dwelling *Marchantia polymorpha*[13], we have been intrigued by the classical circular DNA structure. Emerging evidence indicates that, in terms of stoichiometry, multi-chromosomal arrangements comprising circular, overlapping, and linear molecules could outpace the prevalence of the circular mapping architecture exhibited plant mitogenomes[15]. The mitogenome of *Ilex macrocarpa* and *Salix cardiophylla* have typical circular structures, while the mitogenomes of *Lactuca linnaeus* and *Nicotiana tabacum have* complex structures[64–67]. Indeed, the mitogenome sizes exhibit significant variations across the Magnoliids clade, such as *M. biondii*[56], *C.*

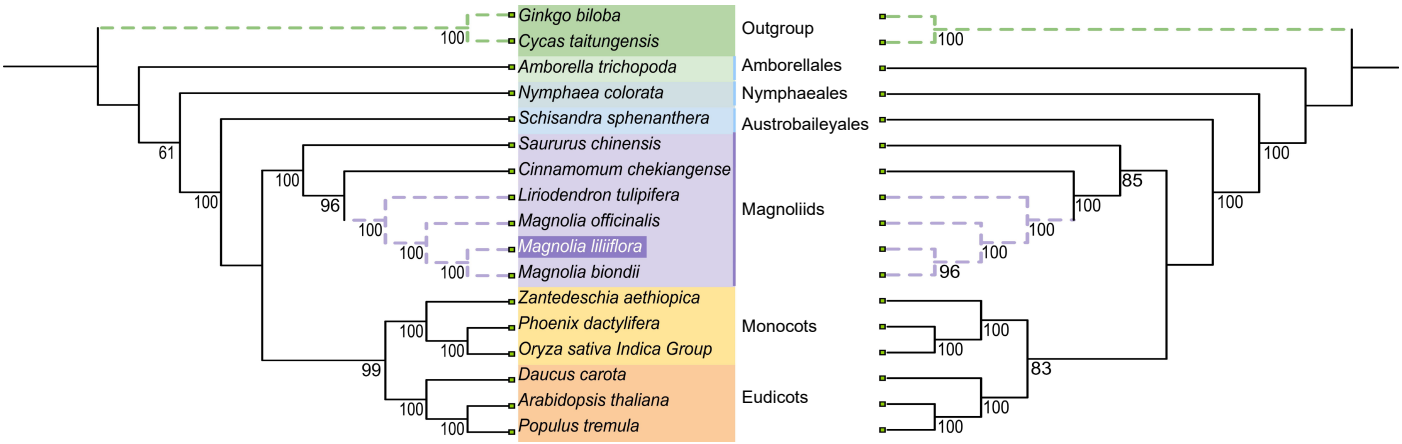

Fig. 7 The relationships of *Y. liliiflora* with the 16 other represented land plants are shown based on the chloroplast (left) and mitochondrial (right) genomes. Bootstrap support values are presented at each node.

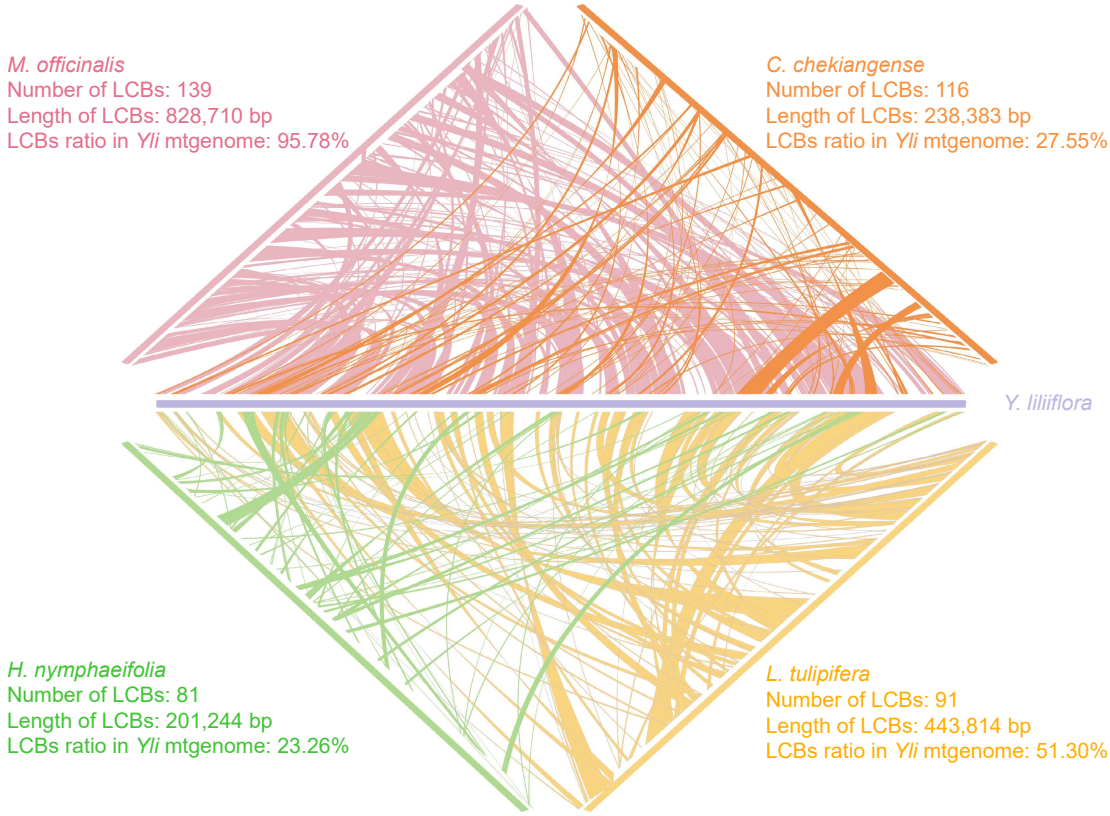

Fig. 8 Mitogenome synteny of five species. The purple, pink, orange, yellow, and green legends represent *Y. liliiflora, M. officinalis, C. chekiangense, L. tulipifera*, and *H. nymphaeifolia,* respectively.

chekiangense[68] were assembled into a single circular molecule. It is significant to point out that the mitogenome structure of *Y. liliiflora* is breach linear in shape.

Mitogenomes have been found to occur in complex structures depending on the stage of plastid maturation[69,70]. The abundance of repetitive sequences can enhance recombination events, facilitate exogenous DNA integration, and contribute to the complexity of mitogenome structure, thereby enabling, for instance, distinctive self-replication of three mitochondrial chromosomes is noted in cucumber plant cells[69,71]. The linear structure also may be attributed to the telomeric recombination's significant role in the formation of breached linear organelle chromosomes[14]. In addition, this study identified colinear regions and rearrangements among five Magnoliids mitogenomes, revealing a high degree of homology as depicted in Fig. 8. The mitogenomes of *Y. liliiflora* and two Magnoliaceae species (*M. officinalis* and *L. lulipifera*) showed the highest collinearity (95.78%, 51.30%), while the collinearity between *Y. liliiflora* and Lauraceae (*C. chekiangense*, *H. nymphaeifolia*) are over 20%. The collinearity analysis results indicate that the mitogenome of Magnoliaceae may be more evolutionarily conserved.

Previous studies have found that the phenomenon of *trans*-splicing has manifested recurrently in numerous lineages of vascular plants, with a significant portion originating from ancestors that employed *cis*-splicing[18,72–74]. According to research by Guo et al. and Mower, across angiosperms and gymnosperms, there is a shared presence of five mitochondrial *trans*-spliced group II introns,

suggesting their probable evolution in the common ancestor through the fragmentation of a *cis*-spliced arrangement[72,73], which confirms the fact that *nad1*, *nad2*, and *nad5* contains *trans*-spliced in our study. This result also strongly suggests that the mitogenome of *Y. liliiflora* is highly conserved.

PCG content of the *Y. liliiflora* mitogenome

In endosymbiotic evolution, organellar gene content declined, yet photosynthetic and respiratory chains, along with protein synthesis components, remained conserved across lineages[73,75]. The core PCGs consist of five ATP synthase genes (*ap1*, *4*, *6*, *8*, and *9*), four cytochrome C biogenesis genes (*ccmB*, *C*, *Fc*, and *Fn*), ubiquinol cytochrome c reductase (*cob*), three cytochrome C oxidase genes (*cox1–3*), 9 NADH dehydrogenase genes (*nadl-7*, *9*, and *4L*), a transport membrane protein (*mttB*/*tatC*), and a maturase (*matR*). The variable PCGs consist of five large subunits of ribosome proteins (*rpl2, 5, 6, 10*, and *16*), 12 small subunits of ribosome proteins (*rps1–4, 7, 8, 10–14,* and *19*), and two respiratory genes (*sdh3–4*). The mitogenome of *Y. liliiflora* mainly consists of non-coding sequences, with PCGs only accounting for 4.24% in the mitogenome. As depicted in Supplementary Fig. S3, all 24 core PCGs and 17 variable PCGs were identified in the *Y. liliiflora*. Almost all of the ancestral 24 core PCGs were identified, such *G. biloba* (gymnosperm)[76], *L. tulipifera* (Magnoliids)[62], *O. sativa* (monocots)[77], and *A. thaliana* (eudicots). During the evolutionary process from bryophytes to the ancestral line of angiosperms, there was a loss of the *rpl6* and *rps8* genes. However, the *rps11* gene, while retained in gymnosperms, seems to have vanished in both monocots and eudicots, presumably during the split of these groups from the gymnosperms. Following the divergence of monocots and eudicots, a distinct pattern of gene loss emerged, with four genes (*sdh3, sdh4, rps10,* and *rps14*) being absent in monocots and *rps2* being lost in eudicots. However, variable genes are more severely missing in monocots and eudicots plants. In sharp contrast to gene loss, gene duplication emerges as one of the principal genetic mechanisms supporting the expansion of gene families in mitogenomes[78,79]. In the mitogenome of *Y. liliiflora*, *atp4, nad3, rps7*, and *rps12* have been duplicated, which is possibly related to the significant physiological functions, but further research is needed to confirm.

Plastid and mitochondrial phylogeny

The phylogenetic relationships of the five major lineages of the Mesangiospermae (Ceratophyllales, Chloranthales, eudicots, Magnoliids, and monocots) remain unclear, despite extensive studies of their diversification. As compared to Magnoliids, the phylogenetic trees constructed from organellar genomes currently suggest a closer evolutionary affiliation between monocots and eudicots[80,81]. Intergeneric relationships based on the mitochondrion and chloroplast dataset are generally well supported. The phylogenetic results obtained from the mitochondrial and chloroplast datasets are consistent, indicating that *Y. liliiflora* is the sister to *M. biondii* (posterior probability [PP] = 0.96). *Y. liliiflora*, *M. biondii*, and *M. officinalis* constitute the *Magnolia* genus, and they are sister groups with *L. tulipifera*. The plastid sequence data, inclusive of our dataset, unambiguously situates *Y. liliiflora* belongs to *Magnoliales* order and is closely related to the *Laurales*. The remaining species form two major clades. Within the updated APG IV classification framework, angiosperms are segregated into two major groups: basal angiosperms and Mesangiospermae[6]. Amborellales, Nymphaeales, and Austrobaileyales constitute the basal angiosperms, known as ANA clade, which is sister to all remaining selected species (PP > 0.99). Another clade formed by eudicots and monocots as successive sister groups to Magnoliids (PP > 0.83). The results show that Magnoliids is sister to the clade consisting of monocots and eudicots, which agrees with previous studies[80,81]. Nevertheless, the phylogenetic associations inferred from nuclear genes demonstrate a notable instability among the five Mesangiospermae clades[81]. This finding points to a potential occurrence of hybridization and introgression during the early evolutionary history of angiosperms.

Conclusions

In this study, we assembled and annotated the mitogenome of *Y. liliiflora*, undertaking comprehensive analyses utilizing both DNA and amino acid sequences derived from the annotated genes. The mitogenome of *Y. liliiflora* exhibits a breach linear structure, encompassing a length of 865,191 bp. We annotated 69 genes, including 45 protein-coding, 21 tRNA, and three rRNA genes. We detected 17 *cis*- and five *trans*-splicing introns in the mitogenome of *Y. liliiflora*. Furthermore, we conducted analyses on codon usage, sequence repeats, RNA editing, and sequence migration within the *Y. liliiflora* mitogenome. Phylogenetic analyses demonstrate that *Y. liliiflora* has the closest genetic relationship with *M. biondii* of Magnoliaceae, and Magnoliids emerge as a closely related group to the clade encompassing monocots and eudicots. The results presented in this paper supplement the existing genetic knowledge pertaining to the genus *Yulania*. Simultaneously, they offer a promising avenue for conducting more comprehensive genomic investigations of *Y. liliiflora*.

Author contributions

The authors confirm contribution to the paper as follows: study design and leading: Bi C; research conducting: Chen Y, Qiu J, Liu W; materials collecting: Bi C; DNA for sequencing preparation and analysis: Bi C; data analysis and manuscript writing: Chen Y; manuscript revision: Bi C, Qiu J. All authors reviewed the results and approved the final version of the manuscript.

Data availability

The sequencing data have been deposited in the NCBI under accession number NC_085212.1.

Acknowledgments

The work is supported by the Natural Science Foundation of Jiangsu Province (BK20220414). We thank Assoc. Prof. Kewang Xu from Nanjing Forestry University for collecting the samples of *Y. liliiflora*. We would also like to express our gratitude to Assoc. Prof. Meng Li from Nanjing Forestry University for providing photos of *Y. liliiflora*.

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary information accompanies this paper at (https://www.maxapress.com/article/doi/10.48130/opr-0025-0017)

Dates

Received 25 November 2024; Revised 10 February 2025; Accepted 3 March 2025; Published online 6 May 2025

References

1. Schühly W, Skarbina J, Kunert O, Nandi OI, Bauer R. 2009. Chemical characterization of *Magnolia biondii* (Flos Magnoliae, Xin Yi). *Natural Product Communications* 4:231–34

2. Wu H, Liu T, Zhang Z, Wang W, Zhu W, et al. 2018. Leaves of *Magnolia liliflora* Desr. as a high-potential by-product: lignans composition, antioxidant, anti-inflammatory, anti-phytopathogenic fungal and phytotoxic activities. *Industrial Crops And Products* 125:416–24
3. Liang Z. 2011. Chemical analysis of *Magnolia liliflora* essential oil and its pharmacological function in nursing pregnant women suffering from decubitus ulcer. *Journal of Medicinal Plants Research* 5:2283–88
4. Bajpai VK, Rahman A, Dung NT, Huh MK, Kang SC. 2008. *In vitro* inhibition of food spoilage and foodborne pathogenic bacteria by essential oil and leaf extracts of *Magnolia liliflora* Desr. *Journal of Food Science* 73:M314–M320
5. Shen Z, Ding X, Cheng J, Wu F, Yin H, et al. 2022. Phylogenetic studies of magnoliids: advances and perspectives. *Frontiers in Plant Science* 13:1100302
6. The Angiosperm Phylogeny Group, Chase MW, Christenhusz MJM, Fay MF, Byng JW, et al. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* 181:1–20
7. Shen Y, Chen K, Gu C, Zheng S, Ma L. 2018. Comparative and phylogenetic analyses of 26 Magnoliaceae species based on complete chloroplast genome sequences. *Canadian Journal of Forest Research* 48:1456–69
8. Wang Z, Kang M, Li J, Zhang Z, Wang Y, et al. 2022. Genomic evidence for homoploid hybrid speciation between ancestors of two different genera. *Nature Communication* 13:1987
9. Sun N, Han F, Wang S, Shen F, Liu W, et al. 2024. Comprehensive analysis of the *Lycopodium japonicum* mitogenome reveals abundant tRNA genes and *cis*-spliced introns in Lycopodiaceae species. *Frontiers in Plant Science* 15:1446015
10. Ma Q, Wang Y, Li S, Wen J, Zhu L, et al. 2022. Assembly and comparative analysis of the first complete mitochondrial genome of *Acer truncatum* Bunge: a woody oil-tree species producing nervonic acid. *BMC Plant Biology* 22:29
11. Ye N, Wang X, Li J, Bi C, Xu Y, et al. 2017. Assembly and comparative analysis of complete mitochondrial genome sequence of an economic plant *Salix suchowensis*. *PeerJ* 5:e3148
12. Wang J, Kan S, Liao X, Zhou J, Tembrock LR, et al. 2024. Plant organellar genomes: much done, much more to do. *Trends in Plant Science* 29:754–69
13. Oda K, Kohchi T, Ohyama K. 1992. Mitochondrial DNA of *Marchantia polymorpha* as a single circular form with no incorporation of foreign DNA. *Bioscience, Biotechnology, and Biochemistry* 56:132–35
14. Smith DR, Keeling PJ. 2013. Gene conversion shapes linear mitochondrial genome architecture. *Genome Biology and Evolution* 5:905–12
15. Xiang Q, Tang J, Yu J, Smith DR, Zhu Y, et al. 2022. The evolution of extremely diverged plastomes in *Selaginellaceae* (lycophyte) is driven by repeat patterns and the underlying DNA maintenance machinery. *The Plant Journal* 111:768–84
16. Skippington E, Barkman TJ, Rice DW, Palmer JD. 2015. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all *nad* genes. *Proceedings of the National Academy of Sciences of the United States of America* 112:E3515–E3524
17. Huang K, Xu W, Hu H, Jiang X, Sun L, et al. 2024. The mitochondrial genome of *Cathaya argyrophylla* reaches 18.99 Mb: analysis of super-large mitochondrial genomes in pinaceae. *arXiv* 00:2410.07006
18. Guo W, Grewe F, Fan W, Young GJ, Knoop V, et al. 2016. *Ginkgo* and *Welwitschia* mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution. *Molecular Biology and Evolution* 33:1448–60
19. Zardoya R. 2020. Recent advances in understanding mitochondrial genome diversity. *F1000Research* 9:270
20. Shearman JR, Sonthirod C, Naktang C, Pootakham W, Yoocha T, et al. 2016. The two chromosomes of the mitochondrial genome of a sugarcane cultivar: assembly and recombination analysis using long PacBio reads. *Scientific Reports* 6:31533
21. Dong S, Zhao C, Chen F, Liu Y, Zhang S, et al. 2018. The complete mitochondrial genome of the early flowering plant *Nymphaea colorata* is highly repetitive with low recombination. *BMC Genomics* 19:614
22. Wang Y, Cui G, He K, Xu K, Liu W, et al. 2024. Assembly and comparative analysis of the complete mitochondrial genome of *Ilex rotunda* Thunb. *Forests* 15:1117
23. Kraft F, Kurth I. 2020. Long-read sequencing to understand genome biology and cell function. *International Journal Of Biochemistry & Cell Biology* 126:105799
24. Lu N, Qiao Y, An P, Luo J, Bi C, et al. 2023. Exploration of whole genome amplification generated chimeric sequences in long-read sequencing data. *Briefings in Bioinformatics* 24:bbad275
25. Wenger AM, Peluso P, Rowell WJ, Chang PC, Hall RJ, et al. 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology* 37:1155–62
26. Bi C, Shen F, Han F, Qu Y, Hou J, et al. 2024. PMAT: an efficient plant mitogenome assembly toolkit using low-coverage HiFi sequencing data. *Horticulture Research* 11:uhae023
27. Zhou C, Brown M, Blaxter M, The Darwin Tree of Life Project Consortium, McCarthy SA, et al. 2024. Oatk: a de novo assembly tool for complex plant organelle genomes. *bioRxiv* Preprint
28. Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics* 31:3350–52
29. Chan PP, Lin BY, Mak AJ, Lowe TM. 2021. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Research* 49:9077–96
30. Chen Y, Ye W, Zhang Y, Xu Y. 2015. High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic Acids Research* 43:7762–68
31. Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* 47:W59–W64
32. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33:1870–74
33. Beier S, Thiel T, Münch T, Scholz U, Mascher M. 2017. MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33:2583–85
34. Jia Y, Bai J, Liu M, Jiang Z, Wu Y, et al. 2019. Transcriptome analysis of the endangered *Notopterygium incisum*: cold-tolerance gene discovery and identification of EST-SSR and SNP markers. *Plant Diversity* 41:1–6
35. Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* 27:573–80
36. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, et al. 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research* 29:4633–42
37. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, et al. 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant* 13:1194–202
38. Li J, Li J, Ma Y, Kou L, Wei J, et al. 2022. The complete mitochondrial genome of okra (*Abelmoschus esculentus*): using nanopore long reads to investigate gene transfer from chloroplast genomes and rearrangements of mitochondrial DNA molecules. *BMC Genomics* 23(1):481
39. Liu Y, Lu N, Bi C, Han T, Guo Z, et al. 2021. FEM: mining biological meaning from cell level in single-cell RNA sequencing data. *PeerJ* 9:e12570
40. Yan J, Zhang Q, Yin P. 2018. RNA editing machinery in plant organelles. *Science China Life Sciences* 61:162–69
41. Edera AA, Small I, Milone DH, Sanchez-Puerta MV. 2021. Deepred-Mt: deep representation learning for predicting C-to-U RNA editing in plant mitochondria. *Computers in Biology and Medicine* 136:104682
42. Robinson JT, Thorvaldsdottir H, Turner D, Mesirov JP. 2023. igv. js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics* 39:btac830
43. Hao Z, Lv D, Ge Y, Shi J, Weijers D, et al. 2020. *RIdeogram*: drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Computer Science* 6:e251
44. Löytynoja A. 2021. Phylogeny-aware alignment with PRANK and PAGAN. *Methods in Molecular Biology* 2231:17–37
45. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32:268–74
46. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* 47:W256–W259

47. Sun L, Jiang Z, Wan X, Zou X, Yao X, et al. 2020. The complete chloroplast genome of *Magnolia polytepala*: comparative analyses offer implication for genetics and phylogeny of *Yulania*. *Gene* 736:144410
48. Hu H, Dong B, Fan X, Wang M, Wang T, et al. 2023. Mutational bias and natural selection driving the synonymous codon usage of single-exon genes in rice (*Oryza sativa* L.). *Rice* 16:11
49. Labella AL, Opulente DA, Steenwyk JL, Hittinger CT, Rokas A. 2019. Variation and selection on codon usage bias across an entire subphylum. *PLoS Genetics* 15:e1008304
50. Almutairi MM. 2021. Analysis of chromosomes and nucleotides in rice to predict gene expression through codon usage pattern. *Saudi Journal of Biological Sciences* 28:4569–74
51. Mohasses FC, Solouki M, Ghareyazie B, Fahmideh L, Mohsenpour M. 2020. Correlation between gene expression levels under drought stress and synonymous codon usage in rice plant by in-silico study. *PLoS One* 15:e0237334
52. Wang X, Chen H, Yang D, Liu C. 2018. Diversity of mitochondrial plastid DNAs (MTPTs) in seed plants. *Mitochondrial DNA Part A* 29:635–42
53. Gui S, Wu Z, Zhang H, Zheng Y, Zhu Z, et al. 2016. The mitochondrial genome map of *Nelumbo nucifera* reveals ancient evolutionary features. *Scientific Reports* 6:30158
54. Sloan DB, Wu Z. 2014. History of plastid DNA insertions reveals weak deletion and AT mutation biases in angiosperm mitochondrial genomes. *Genome Biology and Evolution* 6:3210–21
55. Lukeš J, Kaur B, Speijer D. 2021. RNA editing in mitochondria and plastids: weird and widespread. *Trends in Genetics* 37:99–102
56. Dong S, Chen L, Liu Y, Wang Y, Zhang S, et al. 2020. The draft mitochondrial genome of *Magnolia biondii* and mitochondrial phylogenomics of angiosperms. *PLoS One* 15:e0231020
57. Wang J, Zou Y, Mower JP, Reeve W, Wu Z. 2024. Rethinking the mutation hypotheses of plant organellar DNA. *Genomics Communications* 1(1):e003
58. Bi C, Xu Y, Ye Q, Yin T, Ye N. 2016. Genome-wide identification and characterization of WRKY gene family in *Salix suchowensis*. *PeerJ* 4:e2437
59. Wynn EL, Christensen AC. 2019. Repeats of unusual size in plant mitochondrial genomes: identification, incidence and evolution. *G3 Genes|Genomes|Genetics* 9:549–59
60. Xiong Y, Lei X, Bai S, Xiong Y, Liu W, et al. 2021. Genomic survey sequencing, development and characterization of single- and multilocus genomic SSR markers of *Elymus sibiricus* L. *BMC Plant Biology* 21:3
61. Bi C, Paterson AH, Wang X, Xu Y, Wu D, et al. 2016. Analysis of the complete mitochondrial genome sequence of the diploid cotton *Gossypium raimondii* by comparative genomics approaches. *BioMed Research International* 2019:5040598
62. Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD. 2013. The "fossilized" mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biology* 11:29
63. Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, et al. 2013. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science* 342:1468–73
64. Wang Y, Sun N, Shi W, Ma Q, Sun L, et al. 2023. Assembly and comparative analysis of the complete mitochondrial genome of *Ilex macrocarpa*. *Forests* 14:2372
65. Chen X, Zhang L, Huang Y, Zhao F. 2020. Mitochondrial genome of *Salix cardiophylla* and its implications for infrageneric division of the genus of *Salix*. *Mitochondrial DNA Part B* 5:3485–86
66. Liu M, Liu F, Chen N, Melton JT, Luo M. 2020. Mitochondrial genomes and phylogenomic analysis of *Ulva lactuca* Linnaeus (Ulvophyceae, Chlorophyta). *Mitochondrial DNA Part B* 5:1638–39
67. Li F, Yang A, Lv J, Gong D, Sun Y. 2016. The complete mitochondrial genome sequence of *Sua*-type cytoplasmic male sterility of tobacco (*Nicotiana tabacum*). *Mitochondrial DNA Part A* 27:2929–30
68. Bi C, Sun N, Han F, Xu K, Yang Y, et al. 2024. The first mitogenome of Lauraceae (*Cinnamomum chekiangense*). *Plant Diversity* 46:144–48
69. Wu Z, Liao X, Zhang X, Tembrock LR, Broz A. 2022. Genomic architectural variation of plant mitochondria—a review of multichromosomal structuring. *Journal of Systematics and Evolution* 60:160–68
70. Smith DR, Keeling PJ. 2015. Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. *Proceedings of the National Academy of Sciences of the United States of America* 112:10177–84
71. Alverson AJ, Rice DW, Dickinson S, Barry K, Palmer JD. 2011. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. *The Plant Cell* 23:2499–513
72. Guo W, Zhu A, Fan W, Adams RP, Mower JP. 2020. Extensive shifts from *Cis*- to *Trans*-splicing of gymnosperm mitochondrial introns. *Molecular Biology and Evolution* 37:1615–20
73. Mower JP. 2020. Variation in protein gene and intron content among land plant mitogenomes. *Mitochondrion* 53:203–13
74. Yu R, Sun C, Liu Y, Zhou R. 2021. Shifts from *cis* -to *trans*-splicing of five mitochondrial introns in *Tolypanthus maclurei*. *PeerJ* 9:e12260
75. Sloan DB, Warren JM, Williams AM, Wu Z, Abdel-Ghany SE, et al. 2018. Cytonuclear integration and co-evolution. *Nature Reviews Genetics* 19:635–48
76. He P, Huang S, Xiao G, Zhang Y, Yu J. 2016. Abundant RNA editing sites of chloroplast protein-coding genes in *Ginkgo biloba* and an evolutionary pattern analysis. *BMC Plant Biology* 16:257
77. Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, et al. 2002. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Molecular Genetics and Genomics* 268:434–45
78. Yu J, Chen Q, Ren J, Yang Y, Wang J, et al. 2015. Analysis of the multicopied genes and the impact of the redundant protein coding sequences on gene annotation in prokaryotic genomes. *Journal of Theoretical Biology* 376:8–14
79. De Abreu VAC, Alves RM, Silva SR, Ferro JA, Domingues DS, et al. 2023. Comparative analyses of *Theobroma cacao* and *T. grandiflorum* mitogenomes reveal conserved gene content embedded within complex and plastic structures. *Gene* 849:146904
80. Guo X, Fang D, Sahu SK, Yang S, Guang X, et al. 2021. *Chloranthus* genome provides insights into the early diversification of angiosperms. *Nature Communications* 12:6930
81. Hu H, Sun P, Yang Y, Ma J, Liu J. 2023. Genome-scale angiosperm phylogenies based on nuclear, plastome, and mitochondrial datasets. *Journal of Integrative Plant Biology* 65:1479–89

Copyright: © 2025 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit https://creativecommons.org/licenses/by/4.0/.