BMC Plant Biology

RESEARCH

Open Access

The first complete mitochondrial genome of *Eucommia ulmoides*: a multi-chromosomal architecture and controversial phylogenetic relationship in asterids

Meng Li[1], Weijie Zhao[1], Jing Qiu[2*] and Changwei Bi[1*]

Abstract

Background *Eucommia ulmoides* (*E. ulmoides*) is the sole representative of the Eucommiaceae family, exhibiting abundant medical application as traditional Chinese medicine and high commercial application as a hardy rubber tree. Although some high-quality nuclear genomes and plastid genomes (plastome) have been reported, a comprehensive analysis of its complete mitochondrial genome (mitogenome) remains lacking.

Results In this study, we successfully assembled the first complete mitogenome of *E. ulmoides* based on PacBio HiFi sequencing data. The architecture of *E. ulmoides* mitogenome exhibited two circular chromosomes (MC1 and MC2) with a combined length of 541.7 kb. A total of 61 genes were annotated in the *E. ulmoides* mitogenome, including 38 unique protein-coding genes (PCGs), 21 tRNA genes, and 2 rRNA genes. We found that the dispersed repetitive sequences comprised 21.67% of the mitogenome with the largest repeat spanning > 16 kb. The collinearity of *E. ulmoides* mitogenome with other species was relatively low with the longest homologous sequence being only 5,960 bp. The phylogenetic trees constructed based on mitogenomes and plastomes, respectively, consistently confirmed that *E. ulmoides* belongs to the basal lamiid clade, which is congruent with the APG IV classification system. The PCG composition of mitogenomes across four major lineages of land plants (liverworts, ferns, gymnosperms, and angiosperms) demonstrated the convergent reduction in mitogenomic evolution.

Conclusions We reported the latest mitogenome of *E. ulmoides* with a two-circular chromosomal architecture obtained through PacBio HiFi reads, which was the first mitogenome in the genus *Eucommia* and the order Garryales to be described in detail. The mitogenome of *E. ulmoides* provided a solid foundation for comprehending the evolutionary trajectories and elucidating the underlying causes of cytonuclear discordance in *E. ulmoides*.

Keywords *Eucommia ulmoides*, Mitochondrial genome, Multi-chromosomal architecture, Phylogeny

*Correspondence:
Jing Qiu
power_ko@126.com
Changwei Bi
bichwei@njfu.edu.cn
[1]State Key Laboratory of Tree Genetics and Breeding, Key Laboratory of Tree Genetics and Biotechnology of Educational Department of China, Key Laboratory of Tree Genetics and Silvicultural Sciences of Jiangsu Province, Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, China
[2]Information Department, The First Affiliated Hospital of Naval Medical University, Shanghai 200433, China

Background

Mitochondria are essential semi-autonomous organelles in eukaryotic cells involved in metabolic processes such as energy conversion and cellular respiration [1]. Mitochondria originated from endosymbiotic Alphaproteobacteria that resided in archaea-derived host cells and eventually evolved into an organelle in eukaryotic cells [1–3]. The mitochondrial genomes (mitogenomes) of angiosperms present greater assembly challenges compared to their plastid genomes (plastomes) due to their variable size, complex genomic rearrangements, and gene loss and acquisition patterns, which hinder the progress of plant mitogenome research [4].

The mitogenomes of land plants vary dramatically in structure, size, reorganization, intergenomic gene transfer, and gene loss during evolution [5–8]. Traditionally, the predominant structure of plant mitogenome has been determined as a single master circle (MC), constrained by short reads of the second-generation sequencing technology [5]. Recently, the advancement of third-generation sequencing technology has substantially enhanced the accessibility of plant mitogenomes [9, 10], revealing their in vivo architecture to be more intricate in plants compared to a MC, featured by variable chromosome numbers and complex configurations [6, 7, 11–13]. The architecture of multi-chromosomal mitogenomes has evolved independently in diverse lineages, including *Cucumis* [3], *Silene* [6], *Ajuga* [14], and *Rhopalocnemis phalloides* [15], exhibiting significant intra- and interspecies variability in chromosome number and size. The size of plant mitogenomes exhibits considerable variation ranging from 66 kb of *Viscum scurruloideum* [16] to 11.3 Mb of *Silene conica* [6], whereas the mitogenome size of most angiosperms ranges from 200 to 750 kb [17].

Plant mitogenomes are rich in repetitive sequences, which may be responsible for the variable size of plant mitogenomes and increased copies of functional genes [6, 18]. Large repeats (>1 kb) are more active than short repeats in shaping the co-occurrence of multi-chromosomal mitogenomes by intergenomic homologous recombination, thus establishing the foundation for evolutionary shifts in the organizational and structural polymorphism observed in plant mitogenomes [3, 19]. Frequent sequence transfers observed within plants between organellar and nuclear genomes introduce multiple gene copies, significantly enhancing the diversity of mitochondrial genomes [20]. The mitogenomes of liverworts and mosses have preserved most of ancestral land plant repertoire by encompassing nearly 43 functionally diverse protein-coding genes (PCGs), including 24 core PCGs (*atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFc*, *ccmFn*, *cob*, *cox1*, *cox2*, *cox3*, *matR*, *mttB*, *nad1*, *nad2*, *nad3*, *nad4*, *nad5*, *nad6*, *nad7*, *nad8*, and *nad9*) and 19 variable PCGs (*rpl2*, *rpl5*, *rpl6*, *rpl10*, *rpl16*, *rps1*, *rps2*, *rps3*, *rps4*, *rps7*, *rps8*, *rps10*, *rps11*, *rps12*, *rps13*, *rps14*, *rps19*, *sdh3*, and *sdh4*) [21]. However, the content of PCGs varies significantly among and within other land plant lineages due to the convergent reduction of genes in mitogenomic evolution. For instance, *Haplomitrium mnioides* is the sole plant retaining 42 unique PCGs in its mitogenome [22], whereas *Viscum scurruloideum* [16] from eudicots contains a mere 19 PCGs in its mitogenome.

Eucommia ulmoides (*E. ulmoides*) is a tertiary relic perennial tree and the sole representative species of the Eucommiaceae family. It is highly valued for its medicinal properties as a traditional Chinese medicine according to the records of *Shen Nong Ben Cao Jing* [23]. Nowadays, it is widely commercialized and distributed in various countries including China, Russia, Japan, Britain, France, and the United States for the production of trans-polyisoprene rubber [24]. To date, the majority of studies have concentrated on the pharmacology and phytochemistry of *E. ulmoides* [25–28], yet it still lacks in-depth molecular genetic research of its many commercial traits. The complete plastome and nuclear genome of *E. ulmoides* have been sequenced and characterized in recent studies [29, 30], greatly increasing our comprehension of molecular biology and phylogeny for *E. ulmoides.* Although two complete mitogenomes of *E. ulmoides* have been reported in the NCBI database, the accuracy of their assembly results remains questionable due to the lack of detailed analysis. Therefore, further in-depth research on the mitogenome of *E. ulmoides* is of great significance for its effective utilization and genetic enhancement.

Here we reported the latest assembly mitogenome of *E. ulmoides* with a two-circular chromosomal architecture obtained by precise PacBio HiFi reads and provided the first detailed description of the mitogenome within the order Garryales. We analyzed the characteristics of the *E. ulmoides* mitogenome and performed a phylogenetic analysis. This study is anticipated to provide insights into the evolutionary trajectories of *E. ulmoides* and enhance our comprehension of cytonuclear discordance in the phylogenetic position of this genus.

Results

Features of the *E. ulmoides* mitogenome

For the PacBio Revio sequencing platform, a total of 39,643 PacBio HiFi reads representing ~1.9 Gb were generated, with an average read length of 13,618 bp, and the longest read length was 29,313 bp. Assemblies of the mitogenome generated by both PMAT and Oatk software were consistent, which confirmed the mitogenome of *E. ulmoides* with two circular chromosomes (Fig. S1). To validate the assembly of the *E. ulmoides* mitogenome, we performed the comparison of read coverage depth between the nuclear genome, plastome, and mitogenome

of *E. ulmoides*. The mean coverage of *E. ulmoides* mitogenome was approximately 43.1×, excluding anomalies caused by highly similar and repetitive sequences (Fig. S2). The observed coverage depth displayed substantial variation when compared to that of the plastome (mean coverage: 454.4×) and nuclear genome (mean coverage: 5.6×), further supporting the high-quality assembly of the *E. ulmoides* mitogenome. Thus, we confirmed that the *E. ulmoides* mitogenome was assembled into two circular chromosomes and a collective size of 541,669 bp with a GC content of 46.28%. The lengths of MC1 and MC2 were 298,869 bp and 242,800 bp, respectively. The mitogenome of *E. ulmoides* was annotated with a total of 38 unique protein-coding genes (PCGs), 3 rRNA, and 21 tRNA genes (Fig. 1; Table 1), most of which were distributed in the larger chromosome, MC1, indicating higher stability of large chromosomes in gene replication and recombination processes in mitogenomic evolution.

Most PCGs have the same start codon: ATG, while *nad4L* and *rps4* use ACG and ACA (RNA editing in the third site) as their start codons (Table 1). Additionally, the start codons of *mttB* and *rpl16* are not determined as studies reported in the mitogenomes of *Brassica napus* [31] and *Gossypium raimondii* [18] (Table 1). Four distinct types of stop codons were identified among 38 PCGs: TGA, TAA, TAG, and CGA, with *atp9* and *ccmFc* using CGA as their stop codon owing to RNA editing in the first site (Table 1). The diversity of content in functional genes and introns within the mitogenomes of major plant lineages has been shaped by extensive convergent evolution in mitogenomes [21]. The 38 PCGs of *E. ulmoides* contained complete 24 core PCGs (*atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFc*, *ccmFn*, *cob*, *cox1*, *cox2*, *cox3*, *matR*, *mttB*, *nad1*, *nad2*, *nad3*, *nad4*, *nad5*, *nad6*, *nad7*, *nad8*, and *nad9*) and 14 of 19 variable PCGs (*rpl5*, *rpl10*, *rpl16*, *rps1*, *rps3*, *rps4*, *rps7*, *rps10*, *rps12*, *rps13*, *rps14*, *rps19*, *sdh3*, and *sdh4*). Among the 38 PCGs in the mitogenome of *E. ulmoides*, 9 PCGs were identified harboring 33 exons and 24 introns. Specifically, three genes (*ccmFc*, *rps3*, and *rps10*) each contained two exons, four genes (*nad1*, *nad2*, *nad5*, and *nad7*) each possessed five exons, and *cox2* and *nad4* contained three and four exons, respectively. In contrast to the six PCGs (*ccmFc*, *cox2*, *nad4*, *nad7*, *rps3*, and *rps10*) that only contained *cis*-splicing introns, three PCGs (*nad1*, *nad2*, and *nad5*) contained both *cis*-splicing and *trans*-splicing introns, illustrating the variation in intron splicing mechanisms within angiosperm mitogenomes (Fig. S3).

We took our mitogenome of *E. ulmoides* to conduct a comparative analysis with existing mitogenomes of *E. ulmoides* in the NCBI database, including a single-chromosomal mitogenome (NC_082539.1) and a three-chromosomal mitogenome (OQ101613.1, OQ101614.1, OQ101615.1). The mitogenome of the single-circular structure, with a size of only 298.96 kb containing 23 distinct PCGs, was significantly smaller than the other two multi-chromosomal mitogenomes (530–550 kb in length) that both possessed 38 unique PCGs (Table 2). The three mitogenomes of *E. ulmoides* are nearly identical in nucleotide composition illustrated by 46.2–46.3% GC content. Similarly, most of the PCGs were located on the larger chromosome in the multi-chromosomal mitogenomes, and *nad1* was separated by two chromosomes. The results indicated that the assembly solely relying on second-generation sequencing, for instance, Illumina reads, is insufficient to capture the intricate architecture of plant mitogenomes, leading to a substantial loss of conserved genes and an incomplete mitogenome assembly.

Repeat sequences in the *E. ulmoides* mitogenome

To explore the potential function of repeats in the mitogenome of *E. ulmoides*, three different sorts of repeats were identified, including SSRs, tandem repeats, and dispersed repeats. A total of 209 SSRs were found in the mitogenome of *E. ulmoides* with 113 in MC1 and 96 in MC2, respectively (Fig. 2A, Table S1). The most common types of SSRs in both MC1 and MC2 were tetranucleotides, accounting for 47.79% and 47.92%, respectively (Fig. 2A, Table S1). A total of 50 tandem repeats were identified in the mitogenome of *E. ulmoides* with a match greater than 60% and ranging from 30 to 159 bp in length (Fig. 2C and D, Table S2). Tandem repeats were predominantly located in intergenic spacers, consistent with previous findings [32], with seven of the 50 tandem repeats identified within the exonic regions of *cox2*, *matR*, *nad2*, and *nad5* (Table S2).

In addition, we detected 437 pairs of dispersed repeats in MC1, ranging from 50 to 308 bp, including 231 forward repeats and 206 pairs of palindromic repeats (Fig. 2C and D, Table S3). A total of 141 pairs of dispersed repeats were identified in MC2, ranging from 50 to 16,227 bp, including 70 pairs of forward repeats and 71 pairs of palindromic repeats (Fig. 2C and D, Table S3). The lengths of most dispersed repeats ranged from 50 to 150 bp, with only one long repeat (LR1; 16,227 bp) exceeding 1 kb in length in the mitogenome of *E. ulmoides* (Fig. 2B, Table S3 and S4). Previous studies have demonstrated that subgenomic structures in plant mitogenomes are primarily formed through repeat-mediated recombination [33]. These assembly results and the distribution of LR1 within the mitogenome indicated that LR1 may be related to the genomic recombination of MC2. The total length of dispersed repeats within the *E. ulmoides* mitogenome was 113,195 bp in length, accounting for 21.76% of the *E. ulmoides* mitogenome, which revealed that repeats constituted a significant portion of the plant mitogenome, thereby contributing to genomic expansion.

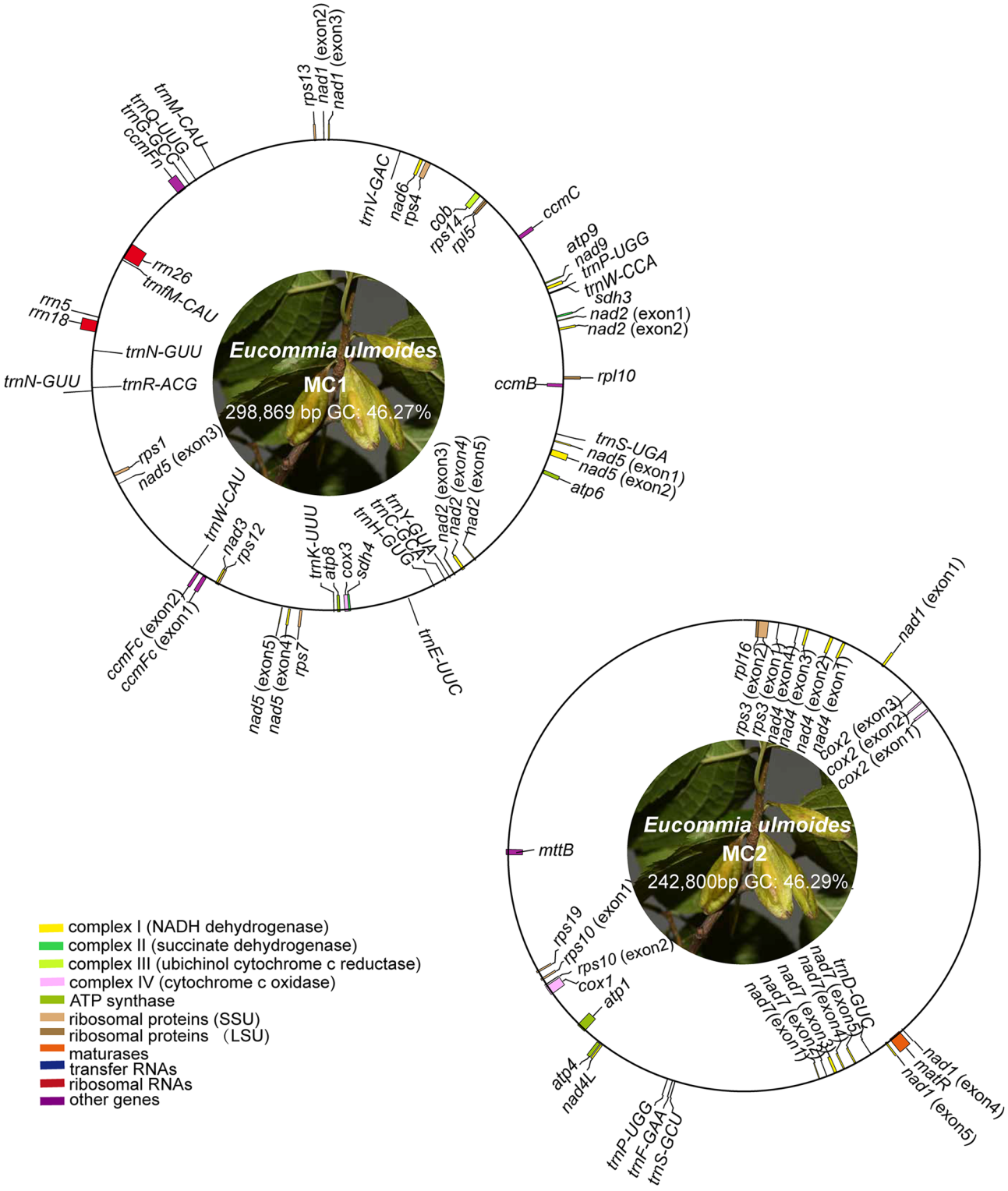

Fig. 1 The Circular maps of the *E. ulmoides* mitogenome. (**A**) MC1 and (**B**) MC2. Genes were shown in different colors in the circle based on different functions of genes. Genes demonstrated on the outside were positive strands but those on the inside were negative strands

Table 1 Gene profile of the *E. ulmoides* mitogenome

Group of genes	Name	Length (bp)	Start codon	Stop codon
ATP synthase	*atp1*	1,530	ATG	TGA
	atp4	579	ATG	TAA
	atp6	720	ATG	CAA
	atp8	525	ATG	TAA
	atp9	225	ATG	CGA (RNA editing)
Cytohrome c biogenesis	*ccmB*	621	ATG	TGA
	ccmC	753	ATG	TGA
	*ccmFc**	1,326	ATG	CGA (RNA editing)
	ccmFn	1,797	ATG	TAG
Cytochrome c oxidase	*cob*	1,182	ATG	TGA
	cox1	1,584	ATG	TGA
	*cox2***	783	ATG	TAA
	cox3	798	ATG	TGA
Maturases	*matR*	1,983	ATG	TAA
	mttB	885	CCG (not determined)	TAG
NADH dehydrogenase	*nad1*####	892	ATG	TAA
	*nad2***#	1,467	ATG	TAA
	nad3	357	ATG	TAA
	*nad4****	1,488	ATG	TGA
	nad4L	303	ACG (RNA editing)	TAA
	*nad5***##	2016	ATG	TAA
	nad6	624	ATG	TAA
	*nad7*****	1,185	ATG	TAG
	nad9	573	ATG	TAA
Ribosomal proteins (LSU)	*rpl5*	492	ATG	TAA
	rpl10	435	ATG	TAA
	rpl16	489	GTG (not determined)	TAA
Ribosomal proteins (SSU)	*rps1*	603	ATG	TAA
	*rps3**	351	ATG	TGA
	rps4	1,059	ACA (RNA editing)	TAA
	rpS6	447	ATG	TAA
	*rps10**	393	ATG	TGA
	rps12	378	ATG	TGA
	rps13	1,650	ATG	TGA
	rps14	303	ATG	TAG
	rps19	285	ATG	TAA
Succinate dehydrogenase	*sdh3*	318	ATG	TGA
	sdh4	456	ATG	TAG
rRNAs	*rrn5*	115	\	\
	rrn18	2,187	\	\
	rrn26	3,296	\	\
tRNAs	*trnC-GCA*	71	\	\
	trnD-GUC	74	\	\
	trnE-UUC	72	\	\
	trnF-GAA	74	\	\
	trnfM-CAU	74	\	\
	trnG-GCC	72	\	\
	trnH-GUG	74	\	\
	trnK-UUU	73	\	\
	trnM-CAU	73	\	\
	trnM-CAU	77	\	\
	trnN-GUU	72	\	\
	trnN-GUU	72	\	\

Table 1 (continued)

Group of genes	Name	Length (bp)	Start codon	Stop codon
	trnP-UGG	71	\	\
	trnP-UGG	75	\	\
	trnQ-UUG	72	\	\
	trnR-ACG	74	\	\
	trnS-GCU	88	\	\
	trnS-UGA	87	\	\
	trnV-GAC	72	\	\
	trnW-CCA	74	\	\
	trnY-GUA	83	\	\

*means the number of *cis*-splicing introns in genes. #means the number of *trans*-splicing introns in genes

Table 2 General mitogenomic features of the three mitogenomes of *E. ulmoides*

Feature	***E. ulmoides* mitogenome**		
GenBank number	OR296708.1–OR296709.1 (sequenced)	NC_082539.1	OQ101613.1–OQ101615.1
Sequencing platform	PacBio HiFi	Illumina	Illumina, Oxford Nanopore
Genome structure	multi-chromosomal	single-chromosomal	multi-chromosomal
Genome size (bp)	541,669	298,961	534,748
Chromosome size (bp)	Chr1: 298,869; Chr2: 242,800	Chr: 298,961	Chr1: 345,526; Chr2: 134,112; Chr3: 55,110
GC (%)	46.28	46.25	46.31
No. of PCGs*	38	23	38
No. of introns	24	9	24
No. of exons	33	12	33
No. of tRNA genes	21	19	22
No. of rRNA genes	3	3	3
Coding regions	39,558 (7.30%)	25,230 (8.44%)	39,555 (7.40%)
tRNA	1,944	1,633	1,941
rRNA	5,598	5,594	55,98
PCGs	32,016	18,003	32,016
Intergenic regions (bp)	502,111 (92.70%)	273,731 (81.56%)	495,193 (82.60%)

*Duplicate genes/introns are included in length statistics but excluded from reported counts

Analysis of RSCU

The codon usage analysis was conducted for 38 unique PCGs in the mitogenome of *E. ulmoides*. Genes with relative synonymous codon usage (RSCU) values greater than 1 were considered to demonstrate a preferential usage of codons. A total of 10,612 codons were identified across the 38 PCGs in the *E. ulmoides* mitogenome excluding termination codons (End). Except for the start codon AUG and the tryptophan codon UGG, both of which showed identical RSCU values of 1, all PCGs in the mitogenome demonstrated significant codon usage bias (Fig. 3A). UUU (Phe), AUU (Ile), and GAA (Glu) are the three most frequent codons in the mitogenome of *E. ulmoides* (Fig. 3B). Apart from the End, Cys is the least frequently used amino acid residue in the mitogenome of *E. ulmoides*, while Leu has the highest usage (Fig. 3B).

Analysis of collinearity

To investigate the evolutionary changes in the mitogenome of *E. ulmoides* with relatives, we aligned sequences of mitogenomes from *E. ulmoides* and two related species *Ipomoea nil (I. nil*; NC_031158.1) and *Scyphiphora hydrophyllacea* (*S. hydrophyllacea*; NC_057654.1) using local BLASTn [34], all of these species belonging to lamiids clade. Abundant homologous local collinear blocks (LCBs) were identified, with the longest LCB not exceeding 6,000 bp (Fig. 4A and B, Table S5), indicating a relatively low level of conservation among their mitogenomes. Between the mitogenomes of *I. nil* and *E. ulmoides*, a total of 71 LCBs (total length: 80,453 bp) were identified accounting for 14.85% of the *E. ulmoides* mitogenome with an average length and identity of 1,133 bp and 92.15%, respectively (Table S5). We detected 68 LCBs (total length: 83,433 bp) between the mitogenomes of *S. hydrophyllacea* and *E. ulmoides*, comprising 15.40% of the *E. ulmoides* mitogenome (Table S5). The average length and identity of these LCBs were 1,227 bp and 91.05%. Comparative analysis of the mitogenomes of *I. nil* and *S. hydrophyllacea* exhibited that the cumulative length of all 65 LCBs was 91,472 bp (average length and identity: 1,407 bp; 91.84%), accounting for 34.42% and 25.83% of their mitogenomes, respectively (Table S5). The mitogenome of *E. ulmoides* may have undergone extensive genomic rearrangements in comparison to the mitogenomes of *I. nil* and *S. hydrophyllacea*, which

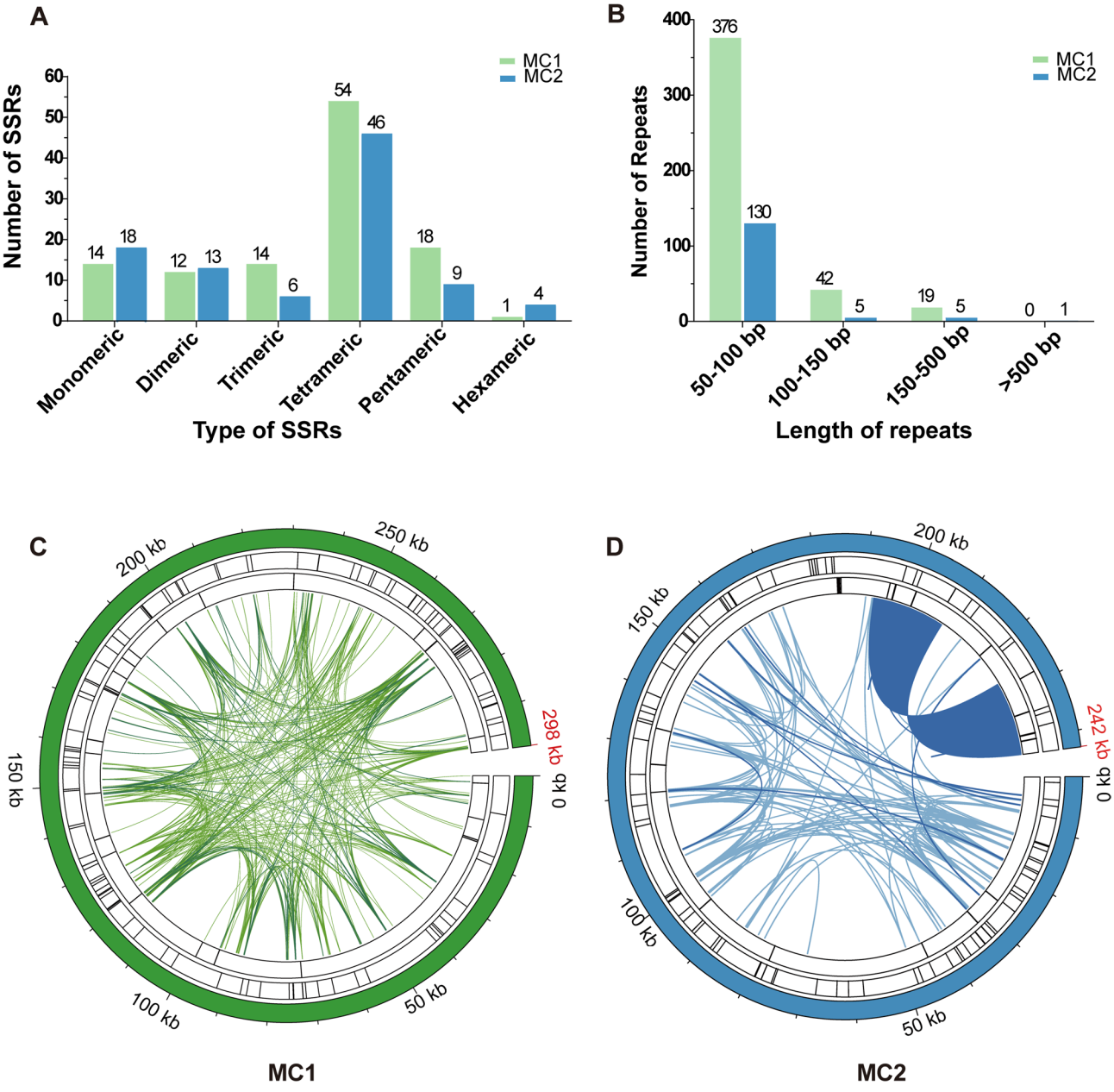

Fig. 2 Repeats identified in the *E. ulmoides* mitogenome. (**A**) SSR numbers of different types. (**B**) Length distribution of dispersed repeats. (**C**) and (**D**) show repeats distribution of MC1 and MC2, respectively. The two outer circles demonstrate SSRs and tandem repeats, respectively, and lines in the inner circle show the distribution of dispersed repeats (> 50 bp)

demonstrated a lower proportion of LCBs throughout their mitogenomes.

Phylogenetic analysis

The classification of *E. ulmoides* into the lamiid lineage remains controversial in recent studies [30, 35–37]. To infer the phylogenetic position of *E. ulmoides* in asterids, two ML phylogenetic trees were constructed based on conserved PCGs of 20 mitogenomes and plastomes, respectively. Both the mitogenome and plastome phylogenetic trees were congruent with the APG IV botanical classification system, confirming *E. ulmoides* in the basal branch of the lamiids lineage, as a sister group to the core lamiids (i.e., Boraginales, Gentianales, Lamiales, and Solanales) [38] (Fig. 5 and Fig. S5, Table S6). However, the phylogenetic position inferred from the organelle genome exhibited a discrepancy with the findings based on nuclear genes, which suggested that *E. ulmoides* belongs to the sister group of the lamiids + campanulids lineage [30], rather than being part of the lamiids lineage clade. Furthermore, the results demonstrated a well-supported core lamiids clade with high bootstrap scores, yet

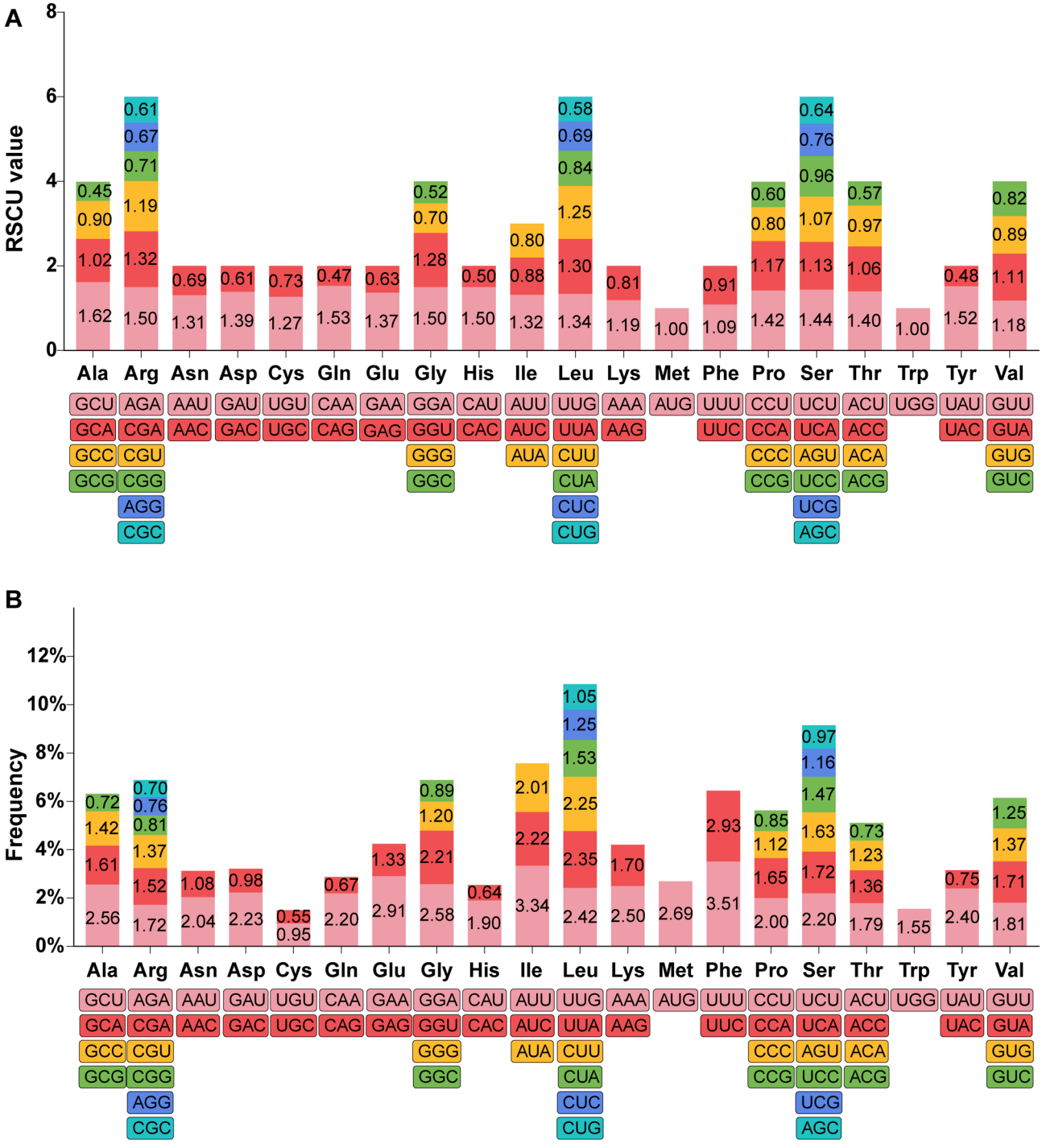

Fig. 3 The Codon usage in the *E. ulmoides* mitogenome. (**A**) RSCU value in the *E. ulmoides* mitogenome. Excluding the stop codons, other codon families are on the X-axis. (**B**) Frequency of each codon and amino acid residues composed of them in usage

the relationships among the major core lineages remain uncertain given lower bootstrap values and inconsistencies between the mitogenome and plastome phylogenetic trees.

Variations in PCG composition among land plant mitogenomes

To demonstrate variations in PCG composition with the evolutionary history of plant mitogenomes, a comparative analysis of 24 core PCGs and 19 variable PCGs among 22 distinct plant mitogenomes was conducted,

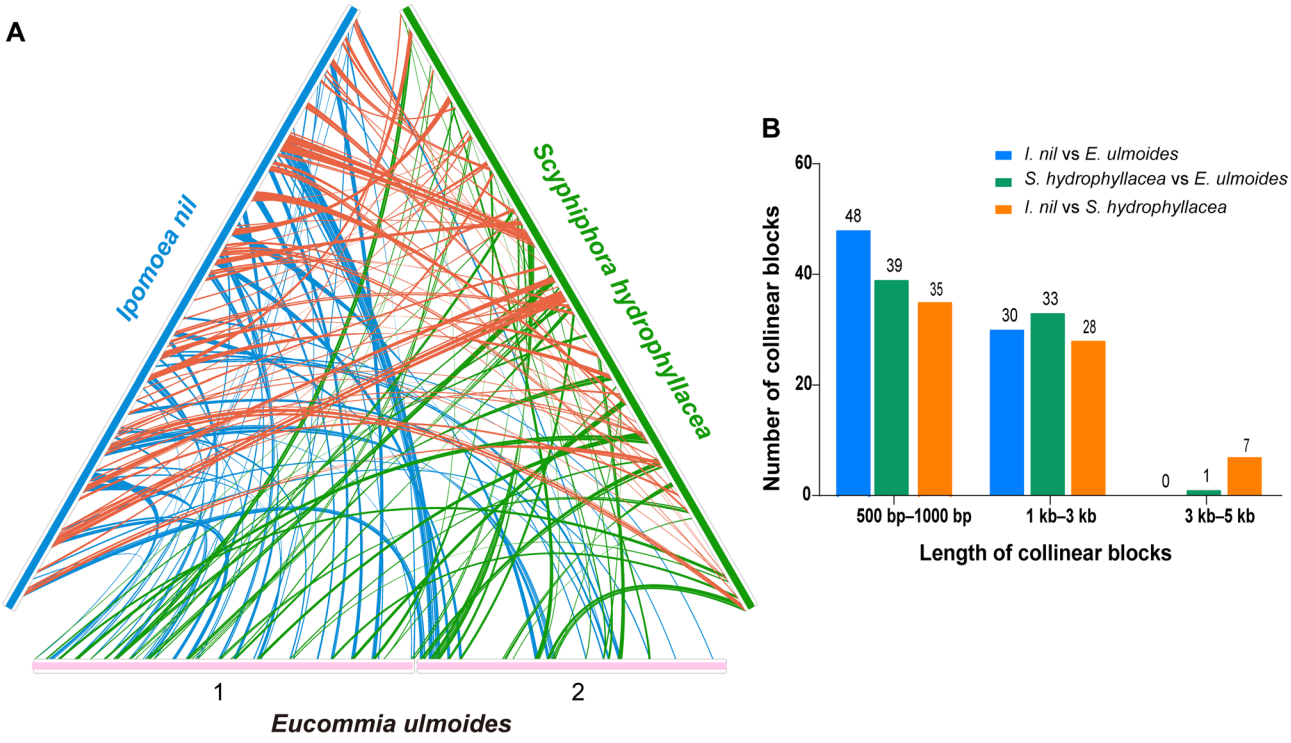

Fig. 4 Local collinear blocks (LCBs) identified between *E. ulmoides*, *Ipomoea nil*, and *Scyphiphora hydrophyllacea* mitogenomes. (**A**) LCBs distribution between these three mitogenomes. The filled bars indicate the mitogenomes of the three species. The lines in each two bars show the LCBs in the two mitogenomes. (**B**) Length distribution of LCBs in each of two mitogenomes

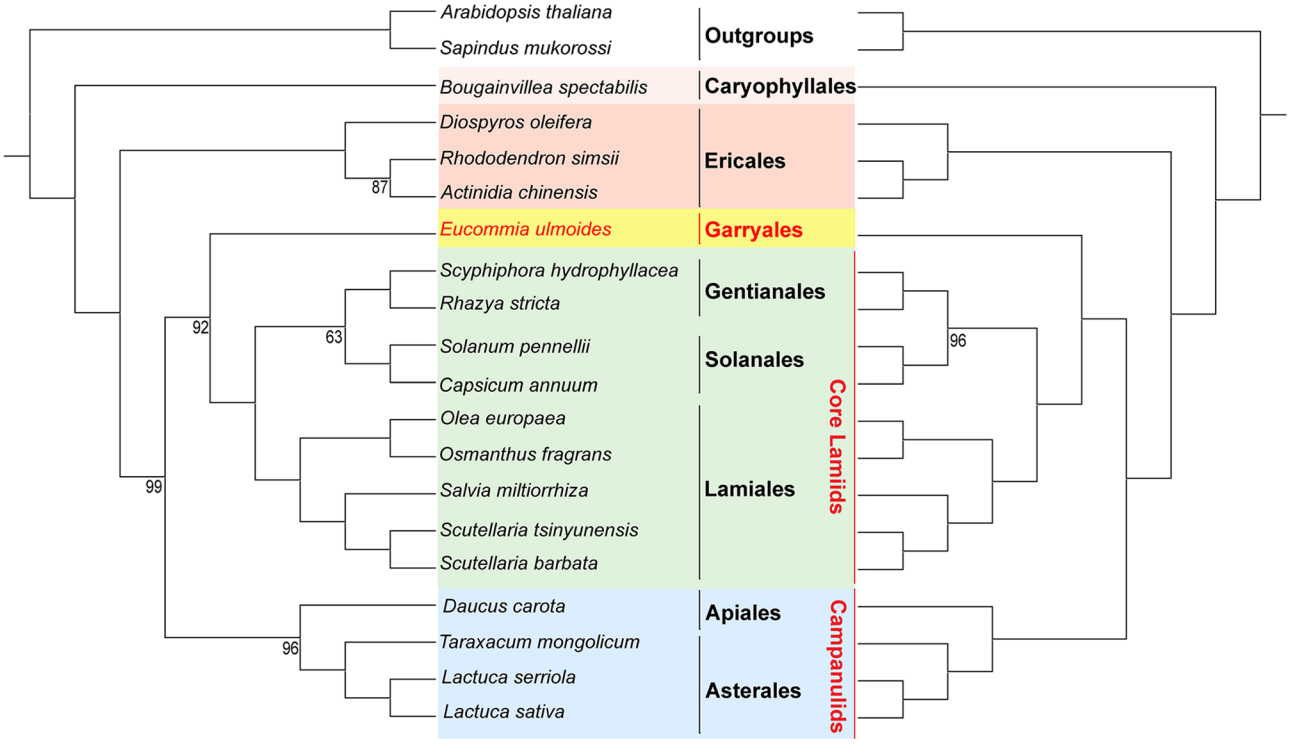

Fig. 5 The phylogenetic relationships of 20 plant species from 7 orders of asterids based on shared PCGs of mitogenomes (left) and plastomes (right). Bootstraps less than 100 were shown in the figure, while equal to 100 were omitted

including four mostly lineages of land plants (liverworts, ferns, gymnosperms, and angiosperms) (Table S7). Unsurprisingly, liverworts mitogenomes were identified to possess the most abundant PCGs, in which *Haplomitrium mnioides* is the only species so far sequenced that contains 42 complete unique PCGs [22] (Fig. 6), while the composition of PCG exhibited significant variation among and within other lineages mitogenomes due to the convergent loss of PCGs. Genes encoding ribosomal proteins, succinate dehydrogenase subunits, and cytochrome c maturation subunits are susceptible to being lost, whereas those encoding oxidative phosphorylation subunits are more likely to be retained. Thus, most land plants have retained between 30 and 42 PCGs in their mitogenomes. The absence of *rps2*, *rps8*, and *rps11* in most of eudicot mitogenomes indicated ancient gene losses in lineage, which is in contrast with the recent losses of *rps14* and *rps19* as evidenced by phylogenetic distribution patterns [6].

Discussion

Multi-chromosomal architecture in plant mitogenomes

The in vivo architecture of plant mitogenomes exhibits rapid transitions among linear, circular, and branched conformations, along with variable chromosome counts, as reported in previous studies [3, 13, 39–41]. Owing to new approaches to whole-genome sequencing and the increasing availability of complete plant mitogenomes, a large number of plant mitogenomes with intricate structures were assembled and reported [9, 10]. Although most eudicots mitogenomes with a multi-chromosomal architecture only contain 2 to 5 molecules [42], the biggest mitogenome in the *Silene* genus possesses more than 128 chromosomes [6]. Recent studies have demonstrated that the multi-chromosomal structure of plant mitogenomes is closely associated with the uneven distribution of PCGs, which are predominantly clustered in the largest chromosome. Surprisingly, certain PCGs like *nad1* or *nad5* were detected spanning across two chromosomes [3, 39], and their functions in plant mitogenome remain unclear.

In this study, we sequenced and assembled the complete mitogenome of *E. ulmoides* based on precise PacBio HiFi sequencing reads into two circular chromosomes, MC1 and MC2, as well as provided the first detailed description of the mitogenome within the order Garryales. Through comparative analysis of three *E. ulmoides* mitogenomes, including two retrieved from the NCBI database, we confirmed the presence of a highly complete multi-chromosome structure and determined the middle-ranking size of the *E. ulmoides* mitogenome among angiosperms. Consistent with previous studies on the gene *nad1*, the mitogenome of *E. ulmoides* revealed that this gene is fragmented across two distinct chromosomes, which necessitates further transcriptome analysis to determine whether the *nad1* gene can be properly expressed and functionally active within the *E. ulmoides* mitochondrion.

Repeat sequences in the *E. ulmoides* mitogenome

Repeats are common in the land plant mitogenomes, demonstrating high-level polymorphism. Repeats enhance the mitogenome's size, and an extensive length of repeats significantly increases the likelihood of mediating genomic recombination [41]. The species within the *Silene* genus are distinguished by their remarkable multi-chromosomal architecture and extensive gene expansion in their mitogenomes, with repeats comprising a significant portion of these genomes. Notably, the largest mitogenome of *Silene* contains 4.6 Mb of dispersed repeats, surpassing all other sequenced plant mitogenomes in both absolute and percentage terms (40.8%) [6]. Repetitive sequences play a crucial role in the replication and repair of plant organelle genomes by mediating DNA recombination, thereby maintaining organelle genome stability [33]. Previous research in the *Phaseolus vulgaris* mitogenome [4] and *Salvia officinalis* mitogenome [41] both revealed that large repeats mediated genomic recombination. In this study, numerous repeats were identified in the *E. ulmoides* mitogenome with dispersed repeats comprising 21.76% of its length, which may explain the middle-ranking size of the mitogenome of *E. ulmoides* among mostly angiosperms.

Phylogenetic position inference of *E. ulmoides*

Previous studies primarily relied on sequence variations from nuclear genomes or plastomes, often neglecting mitogenomes due to the complexities involved in assembling plant mitogenomes. Consequently, reconstructing phylogenetic trees based on both mitogenome and plastome sequences in *E. ulmoides* holds particular significance for elucidating its phylogenetic position. Our paired comparisons indicated *E. ulmoides* was classified into the basal group of lamiids, closely related to core lamiids (Fig. 5 and Fig. S5). The classification was upheld by phylogenetic trees based on plastomes [43, 44], yet it exhibited discrepancies when compared to previous research that relied on nuclear genomes [36]. The inconsistency demonstrated between nuclear genes and organelle genes in their phylogenetic patterns was named cytonuclear discordance [45]. It is widely observed in plants and can be attributed to various factors, such as incomplete lineage sorting, hybridization between species, cytoplasmic introgression, as well as gene duplication or extinction [46–48].

Additionally, we have identified some evolutionary inconsistencies within core lamiids for three orders (Gentianales, Lamiales, and Solanales) in the phylogenetic

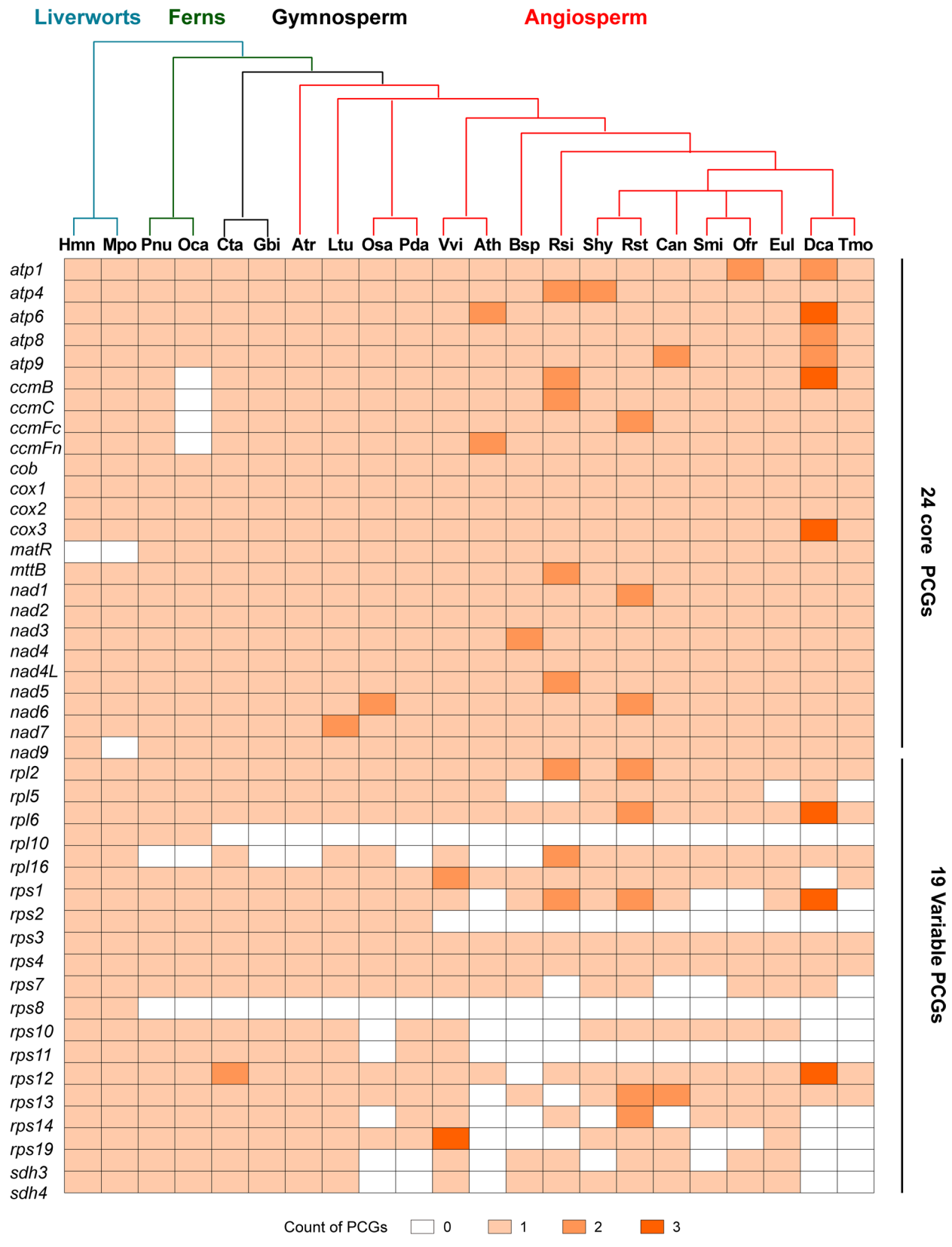

Fig. 6 Comparison of PCG contents among 22 representative land plant mitogenomes

trees based on sequence variation in mitogenome to plastome. These orders with relative orders in asterids may have originated through rapid radiation or undergone complex evolutionary diversifications, which have obscured the lineage sorting of their three types of genomes. Further studies in whole genome sequencing with new methods, such as simulation and selection, are needed to disentangle the cause of such cytonuclear discordance in *E. ulmoides*. The expansion of species in taxa like the orders Icacinales, Boraginales, and Cornales, and the increasing availability of plant mitogenomes are anticipated to elucidate the phylogenetic relationships among species within the asterids.

Variation of protein-coding genes among plant mitogenomes

Mitochondrial protein-coding gene content is highly conserved in mosses and liverworts, yet exhibits considerable variation across and within other land plant lineages [21]. Previous studies demonstrated that *Haplomitrium mnioides* is the sole plant retaining 42 unique PCGs in its mitogenome, whereas most plant mitogenomes retained the number of unique PCGs ranging from 30 to 42 (Fig. 6). This variation is primarily driven by the convergent loss of genes encoding ribosomal proteins, and to a lesser extent, those involved in cytochrome c maturation and oxidative phosphorylation. In contrast with the gene loss in plant mitogenomes, previous studies have investigated the factor related to genes retention. For instance, plant mitochondrial genes with high GC content and strong hydrophobicity are preferentially retained during evolution [49]. Our analysis of mitochondrial gene content across plant lineages revealed plant mitogenomes across different lineages encode almost complete set of PCGs which similar to the common ancestor. Notably, ribosomal subunit (*rps/rpl*) and succinate dehydrogenase (*sdh*) genes exhibited significantly higher frequencies of loss or transfer [21, 50, 51] (Fig. 6). A total of 38 unique PCGs were identified in the *E. ulmoides* mitogenome, lacking *rpl5*, *rps2*, *rps8*, and *rps11*. The absence of *rps2*, *rps8*, and *rps11* in most eudicot mitogenomes suggests ancient gene losses in this lineage, contrasting with the more recent losses of *rps14* and *rps19* as evidenced by phylogenetic distribution patterns [6, 50, 51]. These findings demonstrated extensive convergent evolution in mitogenome PCG composition across land plants.

Conclusions

In this study, we sequenced and assembled the complete mitogenome of *E. ulmoides* and provided the first exhaustive description of its mitogenome, which is composed of two circular chromosomes, MC1 and MC2, with a total length of 541.7 kb. Analysis of gene content, we found that one gene (*nad1*) was separated by MC1 and MC2. The *E. ulmoides* mitogenome contained numerous repeats with dispersed repeats accounting for 21.76% of its size. Phylogenetic trees based on conserved PCGs of mitogenomes and plastomes provided mitogenomic clues for the phylogenetic position of *E. ulmoides* supporting that *E. ulmoides* belongs to the basal group of lamiids. In conclusion, deciphering the mitogenome of *E. ulmoides* will not only provide invaluable material for future studies on genomic evolution and organelle mutation mechanisms, but also foster efforts to improve medical and industrial applications through genetic engineering.

Materials and methods

Plant materials, DNA extracting, and sequencing

Fresh leaves of *E. ulmoides* were collected from the campus of Nanjing Forestry University (N 32°04′, E 118°48′), Nanjing, China and were stored at -80℃. The plant material was identified by Assoc. Prof. Kewang Xu and the voucher specimen was deposited at Herbarium of Nanjing Forestry University (NF: 2036781). The total genomic DNA was extracted from the specimen using the Hi-DNAsecure Plant Kit (Tiangen DP350). The purity and integrity of the DNA were assessed through agarose gel electrophoresis and evaluated using a Nanodrop 2000 ultraviolet spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). SMRTbell Express Template Prep Kit 2.0 was then used to construct sequencing libraries utilizing high-integrity genomic DNA. The PacBio HiFi sequencing data was obtained from the PacBio Revio platform.

Mitogenome assembly and annotation

Two different kinds of software were applied to assemble the mitogenome of *E. ulmoides* using PacBio HiFi sequencing data. Firstly, the "autoMito" mode in the PMAT v1.20 [9] was utilized to assemble the mitogenome of *E. ulmoides* with parameters '-st hifi -g 947.9 M -ml 40 -mi 90'. The genome of *E. ulmoides* (ASM1664770v1) downloaded from the NCBI database was selected as the reference to evaluate the genome size [37]. Bandage [52] was utilized to visualize the assembly graph of the *E. ulmoides* mitogenome by removing full path contigs from the plastome and excluding tip contigs. The mitogenome of *E. ulmoides* was assembled into two circular chromosomes, MC1 and MC2 (Fig. S1). Secondly, to verify the accuracy of the assembly mitogenome of PMAT, we used Oatk v1.0 to assemble the mitogenome of *E. ulmoides* with the following parameters "oatk -k 1001 -c 10 -t 8 -m angiosperm_mito.fam -p angiosperm_pltd.fam -o oatk Pacbio_hifi_reads" [53]. The assembly results produced by Oatk were consistent with those of PMAT, revealing the accuracy of the assembly result, as evidenced by identical mitochondrial contig structures and comparable genome sizes, with a minimal size discrepancy

of less than 2 kb. Additionally, to assess the sequencing quality of PacBio HiFi reads, minimap2 [54] and SAMtools [55] were used to align reads to the nuclear genome (ASM1664770v1), plastome (NC_037948.1), and mitogenome of *E. ulmoides* and calculated corresponding coverage depth (Fig. S2). Finally, all the coverage depth diagrams were visualized using R v4.3.2.

We used the online program IPMGA (http://www.1kmpg.cn/mgavas/) to annotate three mitogenomes of *E. ulmoides* including one we assembled and two retrieved from the NCBI database (NC_082539.1; OQ101613.1–OQ101615.1). Then tRNAscan-SE v2.0 [56] was used to check the tRNA genes and BLASTn [34] was utilized to identify rRNA genes. PMGmap (http://www.1kmpg.cn/pmgview) [57] was used to detect introns of PCGs in the mitogenome of *E. ulmoides*. All genes and introns were manually verified and corrected by MacVector v18.5. MC1 and MC2 were submitted to GenBank with the accession numbers OR296708.1 and OR296709.1. PMGmap [57] was utilized to visualize the circular maps of the *E. ulmoides* mitogenome.

Analysis of repeats

Three sorts of repeats were identified in the *E. ulmoides* mitogenome. Firstly, the simple sequence repeats (SSRs) of the *E. ulmoides* mitogenome were detected using the MISA web service (https://webblast.ipk-gatersleben.de/misa/) [58]. The minimum values of SSRs for mono-, di-, tri-, tetra-, penta-, and hexanucleotides were established followed as 10, 5, 4, 3, 3, and 3, respectively. The maximum length of sequences between two SSRs was set to 100 bp. Secondly, the tandem repeats in the mitogenome of *E. ulmoides* were identified using Tandem Repeats Finder v4.09 (https://tandem.bu.edu/trf/trf.html) [59] with the following alignment parameters: 2, 7, 7 for matches, mismatches, and indels, respectively; The minimum alignment score was set to 60, while the maximum period size was set to 500. Thirdly, the dispersed repeats were detected using the online program REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer/) [60] with set parameters: hamming distances: 3, minimal repeat size: 50, and maximum computed repeats: 5000. The module of 'Advanced Circos' in the TBtools v2.019 [61] was utilized to visualize the distribution of repeats in the *E. ulmoides* mitogenome.

Analysis of codon usage bias

The codon usage of bias differs in various plant mitogenomes owing to the uneven distribution of mutation patterns in a long-time selection of evolution [62, 63]. CodonW v1.4.2 (http://codonw.sourceforge.net) [64] was utilized to analyze the codon preference of 38 PCGs within the *E. ulmoides* mitogenome, enabling the calculation of RSCU values. We used the 'ggplot' module of R v4.3.2 to visualize RSCU values and frequency of codon usage.

Analysis of collinearity

To investigate the collinearity of mitogenomes between *E. ulmoides* and relative species, two mitogenomes from the NCBI database were downloaded including *Ipomoea nil* (*I. nil*; NC_031158.1) and *Scyphiphora hydrophyllacea* (*S. hydrophyllacea*; NC_057654.1). These three mitogenomes were aligned with each other using local BLASTn [34] with the following parameters '-evalue 1e-5 -word_size 9 -gapextend 2 -reward 2- penalty −3 -gapopen 5'. Homologous sequences with lengths exceeding 500 bp were selected as local collinear blocks (LCBs). Then we used the 'RIdeogram' module of R v4.3.2 to visualize the LCBs in these mitogenomes.

Phylogenetic analysis

To determine the phylogenetic position of *E. ulmoides* in asterids, the phylogenetic tree was constructed based on 21 conserved PCGs (*atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFc*, *cox2*, *matR*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*, *rps3*, and *rps4*) of 20 plant mitogenomes. MAFTT v7.45 was used to align these shared PCGs with the codon alignment model [65]. Then they were connected into a combined dataset using PhyloSuite [66]. ModelFinder v2.2.0 was utilized to select the best-fit partition model (edge-linked) using the Bayesian Information Criterion (BIC) [67]. The best-fit model among mafft results of shared PCGs: TIM + F + I + I + R2: *atp1* + *atp4* + *ccmC* + *ccmFc* + *matR*; TPM3u + F + I + G4: *atp6*; GTR + F + I + G4: *atp8* + *cox2* + *nad3* + *nad6*; HKY + F + G4: *atp9*; GTR + F + R2: *ccmB* + *nad1* + *nad2* + *nad4L* + *nad4* + *nad5* + *nad7* + *nad9*; K3Pu + F + R2: *rps3*; TVM + F + I + I + R2: *rps4*. IQ-TREE was used to construct the maximum likelihood (ML) tree based on the best-fit model with the ultrafast bootstrap pattern [68]. The mitogenomes of *Arabidopsis thaliana* and *Sapindus mukorossi* were set as outgroups. Additionally, a plastid ML tree based on 48 PCGs of 20 plant plastomes was constructed using the same pipeline. We used the online program iTOL v6 (https://itol.embl.de/) [69] to visualize the two ML trees.

Comparison of PCG content among land plants

Mitogenomes of liverworts and mosses possess nearly a complete set of unique functional genes of land plants including 24 core PCGs and 19 variable PCGs, which may descend from their common ancestors [21]. However, PCG loss is frequent in the mitogenomes of angiosperms due to convergent evolution. The mitogenomes of 22 land plants including liverworts, ferns, gymnosperms, and angiosperms lineages, were downloaded from the

NCBI database. IPMGA was utilized to verify and correct the annotations of these mitogenomes.

Abbreviations

mitogenome	mitochondrial genome
plastome	plastid genome
MC	Master circle
PCG	Protein-coding gene
APG IV	Angiosperm Phylogeny Group IV
SSR	Simple sequence repeat
LCB	Local collinear block
RSCU	Relative synonymous codon usage

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06771-9.

Supplementary Material 1: Fig. S1. The assembly graph of the E. ulmoides mitogenome. MC1 and MC2 refer to the two chromosomes in the E. ulmoides mitogenome. Different colors represent different DNA fragments sorted and labeled with LR1/contig1-3 according to length. Fig. S2. Comparison of read coverage depth between nuclear genome(A), plastome(B), and mitogenome(C) of E. ulmoides. Fig. S3. Cis-splicing PCGs and trans-splicing PCGs in the E. ulmoides mitogenome. Fig. S4. Dispersed repeats in the E. ulmoides mitogenome. Fig. S5. Phylogenetic relationships of 7 orders of asterids based on mitogenomes (A) and plastomes (B).

Supplementary Material 2: Table S1. SSRs identified in the E. ulmoides mitogenome. Table S2. Tandem repeats identified in the E. ulmoides mitogenome. Table S3. Dispersed repeats identified in the E. ulmoides mitogenome. Table S4. Dispersed repeats identified in the mitogenome of E. ulmoides with a length of more than 150 bp. Table S5. LCBs identified in the three mitogenomes including E. ulmoides, Ipomoea nil, and Scyphiphora hydrophyllacea mitogenomes. Table S6. Species included in the phylogenetic analysis. Table S7. Species abbreviations included in the PCG content analysis.

Acknowledgements
We appreciate that Assoc. Prof. Kewang Xu from Nanjing Forestry University collected the sample of E. ulmoides for us.

Author contributions
CB and ML planned and designed the research. CB provided the sequencing materials and conducted experiments. JQ analyzed the data and WZ wrote the initial version of the manuscript. CB, JQ, and ML revised this and provided comments. All authors read and approved the manuscript.

Funding
The work is supported by the Natural Science Foundation of Jiangsu Province (BK20220414), and the Natural Science Foundation of the Higher Education Institutions of Jiangsu Province (22KJB220003).

Data availability
In this study, the PacBio HiFi sequencing data of ***E. ulmoides*** have been deposited in NCBI with accession numbers PRJNA1129334 and SRR29636512. The mitogenome of ***E. ulmoides*** has been represented in GenBank with accession numbers OR296708.1 and OR296709.1.

Declarations

Ethics approval and consent to participate
This study has rigorously adhered to relevant institutional, national, and international guidelines and regulations. Moreover, the study did not involve the use of any endangered or protected species. The *E. ulmoides* plant leaves utilized in this study were collected in Nanjing Forestry University.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 1 July 2024 / Accepted: 23 May 2025
Published online: 29 May 2025

References

1. Gray MW, Burger G, Lang BF. Mitochondrial evolution. Science. 1999;283:1476–81.
2. Allen J. Why chloroplasts and mitochondria retain their own genomes and genetic systems: colocation for redox regulation of gene expression. Proc Natl Acad Sci. 2015;112:10231–8.
3. Alverson AJ, Rice DW, Dickinson S, Barry K, Palmer JD. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. Plant Cell. 2011;23:2499–513.
4. Bi C, Lu N, Xu Y, He C, Lu Z. Characterization and analysis of the mitochondrial genome of common bean (Phaseolus vulgaris) by comparative genomic approaches. Int J Mol Sci. 2020;21:3778.
5. Backert S, Lynn Nielsen B, Börner T. The mystery of the rings: structure and replication of mitochondrial genomes from higher plants. Trends Plant Sci. 1997;2:477–83.
6. Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, et al. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol. 2012;10:e1001241.
7. 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;200:978–85.
8. Bi C, Sun N, Han F, Xu K, Yang Y, Ferguson DK. The first mitogenome of Lauraceae (*Cinnamomum chekiangense*). Plant Divers. 2024;46:144–8.
9. Bi C, Shen F, Han F, Qu Y, Hou J, Xu K, et al. PMAT: an efficient plant mitogenome assembly toolkit using low-coverage HiFi sequencing data. Hortic Res. 2024;11:uhae023.
10. Xie L, Gong X, Yang K, Huang Y, Zhang S, Shen L, et al. Technology-enabled great leap in Deciphering plant genomes. Nat Plants. 2024;10:551–66.
11. Gualberto JM, Newton KJ. Plant mitochondrial genomes: dynamics and mechanisms of mutation. Annu Rev Plant Biol. 2017;68:225–52.
12. Kozik A, Rowan BA, Lavelle D, Berke L, Schranz ME, Michelmore RW, et al. The alternative reality of plant mitochondrial DNA: one ring does not rule them all. PLoS Genet. 2019;15:e1008373.
13. Varré J-S, D'Agostino N, Touzet P, Gallina S, Tamburino R, Cantarella C, et al. Complete sequence, multichromosomal architecture and transcriptome analysis of the *Solanum tuberosum* mitochondrial genome. Int J Mol Sci. 2019;20:4788.
14. Liu F, Fan W, Yang J-B, Xiang C-L, Mower JP, Li D-Z, et al. Episodic and GC-biased bursts of intragenomic and interspecific synonymous divergence in Ajugoideae (Lamiaceae) mitogenomes. New Phytol. 2020;228:1107–14.
15. Yu R, Sun C, Zhong Y, Liu Y, Sanchez-Puerta MV, Mower JP, et al. The minicircular and extremely heteroplasmic mitogenome of the holoparasitic plant *Rhopalocnemis phalloides*. Curr Biol. 2022;32:470–e4795.
16. 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. 2015;112.
17. Shan Y, Li J, 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.
18. Bi C, Paterson AH, Wang X, Xu Y, Wu D, Qu Y, 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.
19. Yang H, Chen H, Ni Y, Li J, Cai Y, Ma B, et al. De Novo hybrid assembly of the *Salvia miltiorrhiza* mitochondrial genome provides the first evidence of the multi-chromosomal mitochondrial Dna structure of *Salvia* species. Int J Mol Sci. 2022;23:14267.
20. Chang S, Wang Y, Lu J, Gai J, Li J, Chu P, et al. The mitochondrial genome of soybean reveals complex genome structures and gene evolution at intercellular and phylogenetic levels. PLoS ONE. 2013;8:e56502.
21. Mower JP. Variation in protein gene and intron content among land plant mitogenomes. Mitochondrion. 2020;53:203–13.
22. Dong S, Zhao C, Zhang S, Zhang L, Wu H, Liu H, et al. Mitochondrial genomes of the early land plant lineage liverworts (Marchantiophyta): conserved

genome structure, and ongoing low frequency recombination. BMC Genomics. 2019;20:953.
23. Wilms S. The divine Farmer's classic of materia medica. Shen Nong Bencao Jing. 3rd ed. Happy Goat Productions; 2017.
24. Suzuki N, Uefuji H, Nishikawa T, Mukai Y, Yamashita A, Hattori M, et al. Construction and analysis of EST libraries of the trans-polyisoprene producing plant, *Eucommia ulmoides* Oliver. Planta. 2012;236:1405–17.
25. Zhu M-Q, Sun R-C. *Eucommia ulmoides* Oliver: A potential feedstock for bioactive products. J Agric Food Chem. 2018;66:5433–8.
26. Wang C-Y, Tang L, He J-W, Li J, Wang Y-Z. Ethnobotany, phytochemistry and Pharmacological properties of Eucommia ulmoides: a review. Am J Chin Med. 2019;47:259–300.
27. Huang L, Lyu Q, Zheng W, Yang Q, Cao G. Traditional application and modern Pharmacological research of *Eucommia ulmoides* Oliv. Chin Med. 2021;16:73.
28. Yao X, Pan Y, Ma X, Yin S, Zhu M. Efficient separation and production of high-quality rubber, lignin nanoparticles and fermentable sugars from *Eucommia ulmoides* pericarp via deep eutectic solvent pretreatment. Int J Biol Macromol. 2023;253:127221.
29. Wang L, Wuyun T, Du H, Wang D, Cao D. Complete Chloroplast genome sequences of *Eucommia ulmoides*: genome structure and evolution. Tree Genet Genomes. 2016;12:12.
30. Du Q, Wu Z, Liu P, Qing J, He F, Du L, et al. The chromosome-level genome of *Eucommia ulmoides* provides insights into sex differentiation and α-linolenic acid biosynthesis. Front Plant Sci. 2023;14:1118363.
31. Handa H. The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica Napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. Nucleic Acids Res. 2003;31:5907–16.
32. Lee HJ, Lee Y, Lee S-C, Kim C-K, Kang J-N, Kwon S-J, et al. Comparative analysis of mitochondrial genomes of *Schisandra Repanda* and *Kadsura Japonica*. Front Plant Sci. 2023;14:1183406.
33. Maréchal A, Brisson N. Recombination and the maintenance of plant organelle genome stability[J]. New Phytol. 2010;186(2):299–317.
34. Chen Y, Ye W, Zhang Y, Xu Y. High speed BLASTN: an accelerated megablast search tool. Nucleic Acids Res. 2015;43:7762–8.
35. Stull GW, Duno de Stefano R, Soltis DE, Soltis PS. Resolving basal lamiid phylogeny and the circumscription of Icacinaceae with a plastome-scale data set. Am J Bot. 2015;102:1794–813.
36. Wuyun T-N, Wang L, Liu H, Wang X, Zhang L, Bennetzen JL, et al. The hardy rubber tree genome provides insights into the evolution of polyisoprene biosynthesis. Mol Plant. 2018;11:429–42.
37. Li Y, Wei H, Yang J, Du K, Li J, Zhang Y, et al. High-quality de Novo assembly of the *Eucommia ulmoides* haploid genome provides new insights into evolution and rubber biosynthesis. Hortic Res. 2020;7:183.
38. 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;181:1–20.
39. Bi C, Qu Y, Hou J, Wu K, Ye N, Yin T. Deciphering the multi-chromosomal mitochondrial genome of *Populus simonii*. Front Plant Sci. 2022;13:914635.
40. Feng L, Wang Z, Wang C, Yang X, An M, Yin Y. Multichromosomal mitochondrial genome of *Punica granatum*: comparative evolutionary analysis and gene transformation from Chloroplast genomes. BMC Plant Biol. 2023;23:512.
41. Yang H, Chen H, Ni Y, Li J, Cai Y, Wang J, et al. Mitochondrial genome sequence of *Salvia officinalis* (lamiales: lamiaceae) suggests diverse genome structures in cogeneric species and finds the stop gain of genes through RNA editing events. Int J Mol Sci. 2023;24:5372.
42. Wu Z-Q, Liao X-Z, Zhang X-N, Tembrock LR, Broz A. Genomic architectural variation of plant mitochondria—A review of multichromosomal structuring. J Syst Evol. 2022;60:160–8.
43. Wang W, Chen S, Zhang X. Whole-genome comparison reveals heterogeneous divergence and mutation hotspots in Chloroplast genome of *Eucommia ulmoides* Oliver. Int J Mol Sci. 2018;19:1037.
44. Hu H, Sun P, Yang Y, Ma J, Liu J. Genome-scale angiosperm phylogenies based on nuclear, plastome, and mitochondrial datasets. JIPB. 2023;65:1479–89.
45. Funk DJ, Omland KE. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu Rev Ecol Evol Syst. 2003;34:397–423.
46. Folk RA, Mandel JR, Freudenstein JV. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. Syst Biol. 2016. syw083.
47. Lee-Yaw JA, Grassa CJ, Joly S, Andrew RL, Rieseberg LH. An evaluation of alternative explanations for widespread cytonuclear discordance in annual sunflowers (*Helianthus*). New Phytol. 2019;221:515–26.
48. Wang H, Morales-Briones DF, Moore MJ, Wen J, Wang H. A phylogenomic perspective on gene tree conflict and character evolution in Caprifoliaceae using target enrichment data, with Zabelioideae recognized as a new subfamily. J Sytematics Evol. 2021;59:897–914.
49. Johnston IG, Williams BP. Evolutionary inference across eukaryotes identifies specific pressures favoring mitochondrial gene retention. Cell Syst. 2016;2(2):101–11.
50. Adams KL, Palmer JD. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus[J]. Mol Phylogenet Evol. 2003;29(3):380–95.
51. Sloan DB, Warren JM, Williams AM, et al. Cytonuclear integration and co-evolution[J]. Nat Rev Genet. 2018;19(10):635–48.
52. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of *de Novo* genome assemblies. Bioinformatics. 2015;31:3350–2.
53. Zhou C, Brown M, Blaxter M, The Darwin Tree of Life Project Consortium, McCarthy SA, Durbin R. Oatk: A de novo assembly tool for complex plant organelle genomes. 2024.
54. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–100.
55. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of samtools and BCFtools. Gigascience. 2021;10:giab008.
56. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research. 1997;25:955–64.
57. Zhang X, Chen H, Ni Y, Wu B, Li J, Burzyński A, et al. Plant mitochondrial genome map (PMGmap): A software tool for the comprehensive visualization of coding, noncoding and genome features of plant mitochondrial genomes. Mol Ecol Resour. 2024;24:e13952.
58. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33:2583–5.
59. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27:573–80.
60. Kurtz S. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29:4633–42.
61. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 2020;13:1194–202.
62. Wu P, Xiao W, Luo Y, Xiong Z, Chen X, He J et al. Comprehensive analysis of codon bias in 13 *Ganoderma* mitochondrial genomes. Front Microbiol. 2023;14.
63. Ma QP, Li C, Wang J, Wang Y, Ding ZT. Analysis of synonymous codon usage in *FAD7* genes from different plant species. Genet Mol Res. 2015;14:1414–22.
64. Peden JF. Analysis of codon usage[D]. Nottingham: University of Nottingham; 1999.
65. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
66. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, et al. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 2020;20:348–55.
67. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9.
68. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74.
69. Letunic I, Bork P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024;:gkae268.

Publisher's note