BMC Genomics

RESEARCH

Open Access

Mitochondrial genome of *Iodes seguinii* reveals repeat-mediated recombination and phylogenetic insights in Icacinaceae

Shaoshuai Yu[1], Man Wang[1], Xun Gong[2], Meiling Qin[3] and Min Tang[3*]

Abstract

The Icacinales order includes numerous genera and species, many of which have medicinal properties. *Iodes seguinii*, a species from this order, is traditionally used for treating rheumatic and inflammatory diseases. This study presents the first complete mitochondrial genome (mitogenome) of *I. seguinii*, the first characterized in the Icacinales order. A hybrid sequencing approach combining Oxford Nanopore long reads and Illumina short reads was used to assemble a 423,925 bp circular mitogenome, which includes 32 protein-coding genes, 3 rRNAs, and 23 tRNAs, as well as repeat elements that mediate homologous recombination. Additionally, RNA editing analysis revealed extensive post-transcriptional modifications, with C-to-U conversions primarily occurring in genes related to mitochondrial energy metabolism. The chloroplast genome, with a length of 150,599 bp, was also reconstructed. It contains typical features such as large and small single-copy regions and inverted repeats, with genes for photosynthesis, carbon fixation, and energy production. Mitochondrial plastid sequences suggested active inter-organelle gene transfer. Phylogenetic analysis based on mitochondrial genes positioned *I. seguinii* as a distinct lineage within basal eudicots, closely related to *Eucommia ulmoides*. Comparative mitogenome analyses revealed lineage-specific rearrangements and structural evolution. These findings provide insights into the evolution and functional dynamics of mitogenomes in Icacinales, offering a valuable genomic resource for future studies on plant mitochondrial biology and medicinal applications.

Keywords *Iodes seguinii*, RNA-editing site, Mitochondrial plastid sequences, Plant genome, Repeat elements

Introduction

Globally, the Icacinaceae family consists of approximately 58 genera and 400 species, distributed mainly across tropical and subtropical regions, with the Southern Hemisphere being their primary habitat [1]. In China, this family includes around 13 genera and 25 species, primarily located in the southern and southwestern regions. Among these, 10 genera have been identified in Yunnan Province, including *Nothapodytes*, *Pittosporopsis*, *Platea*, *Mappianthus*, *Apodytes*, *Gomphandra*, *Gonocaryum*, *Natsiatum* and *Iodes*. Many genera contain plants with notable medicinal value [2–4]. For instance, *Mappianthus iodoides* is traditionally used by the Dai

*Correspondence:
Min Tang
mt3138@ujs.edu.cn
[1]Department of Pharmacy, Affiliated People's Hospital of Jiangsu University, Zhenjiang, Jiangsu Province 212001, China
[2]Department of Rheumatology & Immunology, Affiliated Hospital of Jiangsu University, Zhenjiang, Jiangsu Province 212001, China
[3]School of Life Sciences, Jiangsu University, Zhenjiang, Jiangsu Province 212013, China

community for conditions such as irritability, excessive urination, and injuries, while the Yao people employ its roots and stems to treat jaundice, menstrual irregularities, and snake bites [5]. Within China, the genus includes species such as *Iodes cirrhosa*, *Iodes balansae*, *Iodes vitiginea*, and *Iodes seguinii (Lévl.) Rehd*, primarily distributed in Yunnan Province. Studies on *I. cirrhosa* have identified various phenolic compounds, including lignans and phenylpropanoids, which show promise in managing inflammation, rheumatic diseases, kidney disorders, and even cancer [6–8]. Genomic research on this genus is scarce, with *I. cirrhosa* being the only species investigated at the genomic level. Its chloroplast genome, comprising 151,994 base pairs and encoding 80 protein-coding genes (PCGs), 28 tRNA genes, and 4 rRNA genes, has been reported and serves as a reference plastome for the genus [9]. However, the overall lack of genomic studies has impeded comprehensive research on the genus.

Distinguished by its nodular lenticels and aromatic, sweet fruit, *I. seguinii* is regarded for its medicinal applications (Fig. 1a of [10]). In Guangxi Province, it is traditionally used to treat nephritis and is often employed as an alternative to *Stephania tetrandra*, a remedy for alleviating rheumatic pain and edema [4, 11]. Recent research has made significant advancements in understanding this species, including the first chromosome-level nuclear genome assembly of *I. seguinii*. This was achieved using PacBio HiFi and Hi-C sequencing technologies, producing a genome with a total length of 273.58 Mb anchored into 13 pseudochromosomes. The genome annotation revealed 25,270 PCGs, including those involved in the biosynthesis of secondary metabolites. Metabolomic profiling identified 84 active compounds, with luteolin highlighted for its therapeutic potential in treating rheumatoid arthritis (RA) [12]. Network pharmacology and molecular docking studies confirmed luteolin targets key RA-related proteins, such as AKT1 and TNF, through pathways like NF-κB and Toll-like receptor signaling [13–18]. These findings underscore the value of *I. seguinii* in managing immune and inflammatory conditions, providing a robust foundation for further genomic, metabolomic, and pharmacological exploration.

This study marks the first-ever reconstruction and annotation of the mitochondrial genome (referred to as mitogenome) of a plant from the family Icacinaceae, specifically *I. seguinii*. It provides a detailed analysis of the mitogenome, alongside the chloroplast genome (referred to as cpgenome), emphasizing the investigation of repeat sequences and codon usage preferences. A systematic prediction of RNA-editing sites within mitochondrial PCGs was performed, shedding light on potential post-transcriptional modifications. Furthermore, the mitochondrial plastid sequences (MTPTs) were analyzed to elucidate inter-organellar gene transfer events. Phylogenetic analysis incorporating mitochondrial PCGs from 25 plant species was conducted, offering insights into evolutionary relationships. Additionally, homologous collinear blocks were identified and compared across four species, enhancing our understanding of genomic synteny and structural evolution. As the first mitogenome characterized in the Icacinaceae family, this study provides a foundation for exploring the biological characteristics of *I. seguinii* and serves as a valuable resource for studying genetic evolution and diversity in plant mitogenomes.

Materials and methods

Plant material collection, dual-platform library preparation, and sequencing using second- and third-generation technologies

Whole plants of *I. seguinii* were gathered from the wild from Xichou County, located in the Wenshan Zhuang and Miao Autonomous Prefecture (23.368075° N, 104.246194° E), Yunnan Province. The species was identified by two qualified taxonomists, and the voucher specimen has been documented and made publicly accessible via the iNaturalist platform (https://www.inaturalist.org/observations/346360782), ensuring traceability and c ompliance with data availability requirements. To preserve their integrity, the samples were rapidly frozen in liquid nitrogen and stored at − 80 °C. Detailed observations and photographs were taken to capture the plants' dimensions, morphology, coloration, surface texture, and other distinguishing features. The collected samples were identified by Prof. Xiaobin Ou from the Gansu Key Laboratory of Protection and Utilization for Biological Resources and Ecological Restoration in Longdong, Longdong University.

Based on the experimental design described in our previously published studies [19–23], leaf samples were treated with DEPC water and stored at − 80 °C prior to nucleic acid extraction. For Oxford Nanopore Technologies (ONT) sequencing, high-quality genomic DNA was extracted by the sequencing service provider using an optimized CTAB-based protocol. The quality and quantity of this sequencing-grade DNA were assessed by 0.75% agarose gel electrophoresis, a NanoDrop One spectrophotometer (Thermo Fisher Scientific), and a Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA, USA), with OD260/OD280 ratios maintained between 1.7 and 1.9. ONT libraries were prepared using the SQK-LSK110 ligation sequencing kit and sequenced on the PromethION platform (Oxford Nanopore Technologies, Oxford, UK). Raw ONT reads were processed using Porechop_ABI (v0.5.0) to remove adapter sequences and NanoFilt (v2.8.0) to filter low-quality reads [24–26]. For Illumina sequencing, genomic DNA was extracted by the sequencing service provider using a commercial kit-based

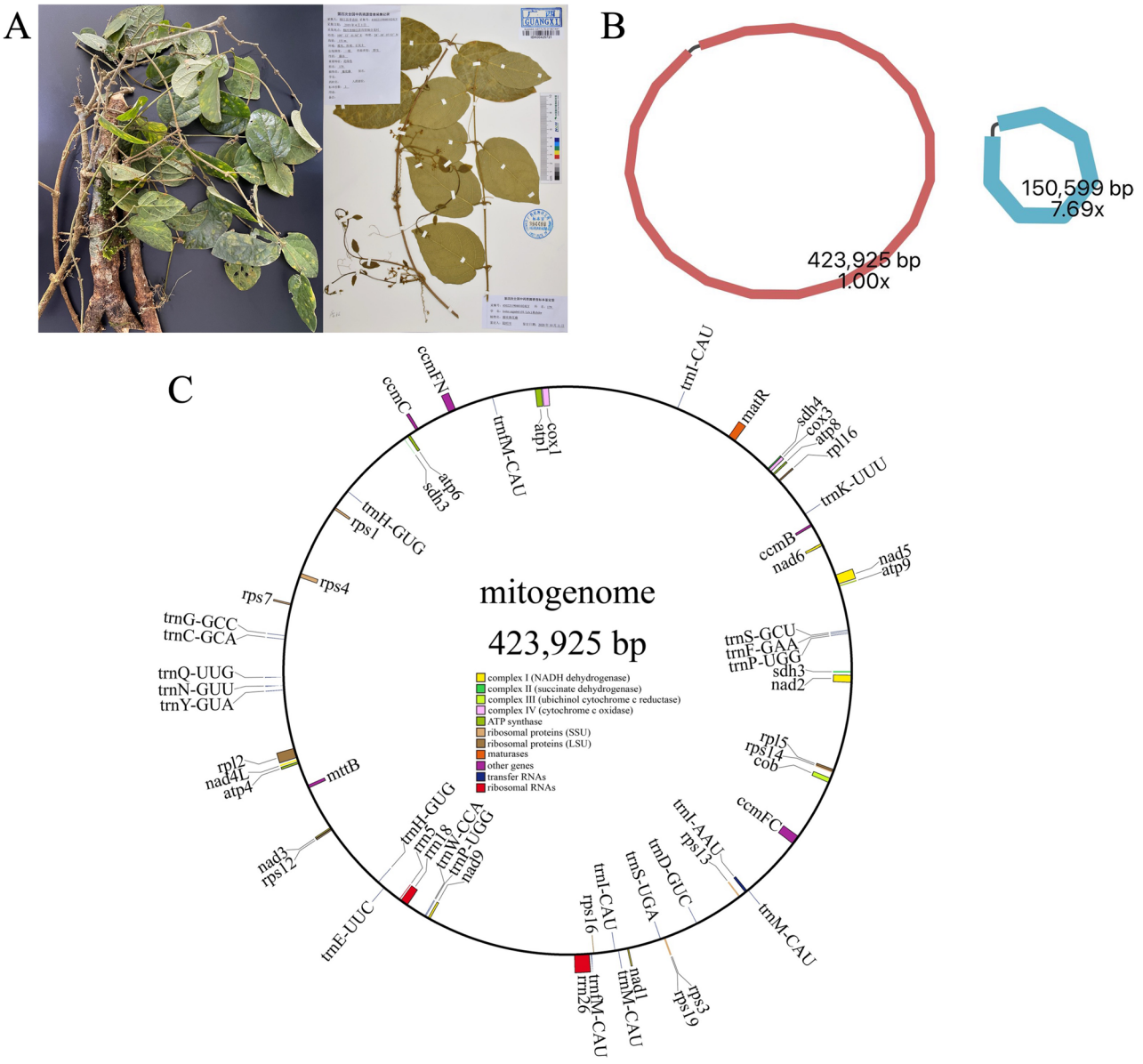

Fig. 1 Morphological characteristics and organellar genome structure of *I. seguinii*. **A** Morphological features of *I. seguinii*: freshly collected plant materials (left) and the corresponding herbarium specimen deposited at the Chinese National Herbarium (right). **B** Circular genome structures obtained from long-read sequencing and hybrid assembly. The larger circle represents the complete mitogenome (423,925 bp; red), while the smaller circle corresponds to the cpgenome (150,599 bp; blue). The mean sequencing coverages for the mitogenome and cpgenome are approximately 100× and 7.69×, respectively. **C** Gen e map of the *I. seguinii* mitochondrial genome. Genes are color-coded according to their functional categories: respiratory complexes (I–V), ribosomal proteins (SSU and LSU), transfer RNAs (tRNAs), and ribosomal RNAs (rRNAs)

method for short-read library construction. Paired-end libraries with an insert size of approximately 300 bp were prepared using the Nextera DNA Flex Library Prep Kit (Illumina, San Diego, CA, USA) and sequenced on the Illumina NovaSeq platform. Raw Illumina reads were quality-filtered and adapter-trimmed using fastp (v0.22.0) with default parameters [27]. In addition, total genomic DNA was independently extracted in our laboratory using the EZ-10 Spin Column Plant Genomic DNA Purification Kit (B518261-0100, Sangon Biotech, China) for subsequent Polymerase Chain Reaction (PCR) validation. This DNA was used exclusively for downstream validation experiments and was not subjected to sequencing-related quality control.

Species identification methodology using DNA barcoding

A detailed comparison of the samples was conducted using the National Plant Specimen Resource Center of China Digital Herbarium to ensure correct identification. Specimens of this species are available in the

Guangxi Institute of Botany Herbarium under specimen number IBK00425721 (Fig. 1A), which was primarily used for comparison with the collected samples. After confirming the accuracy of the identification, the *rbcL*, *matK*, and *psbA-trnH* gene regions were amplified from genomic DNA for precise species identification. The primer sequences and corresponding PCR protocols for amplifying these DNA barcodes are provided in Table S1 [10]. Sequencing results for *rbcL* and *matK* were compared with reference sequences in the BOLD Systems database (https://www.boldsystems.org), while *psbA–trnH* sequences were aligned using the online BLASTn tool provided by NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi) against the Core Nucleotide Database (core_nt). Species identification was determined based on comparison scores, sequence similarity, and E-values from these databases.

Mitogenome assembly and annotation

The mitogenome of *I seguinii* was assembled using a hybrid approach that integrated Oxford Nanopore (ONT) long reads and Illumina short reads [21]. First, the clean ONT reads were subjected to comprehensive quality control, including evaluation of read length distribution, read quality distribution, and other standard sequencing quality metrics, and reads shorter than 1000 bp were removed prior to downstream analyses. The remaining ONT reads were then aligned with a reference mitogenome from *Arabidopsis thaliana* (GenBank: BK010421) using a customized BASH script (Supplemental_code_1). The initial alignment was performed with a minimum matching length of 50 bp and quality control criteria, such as a mapping quality threshold of 60 and a minimum ratio of 0.7 between the match length and total alignment length. Seed sequences were iteratively extracted across five cycles, with the length threshold increasing to 1000 bp in subsequent iterations, and the requirement of at least 2 matching genes for each sequence. The iterative process continued until no new candidate mitochondrial fragments were identified, at which point the cycle for extracting candidate mitochondrial fragments was concluded. Next, Canu (v2.2) [9] was used to correct and trim the filtered mitochondrial-related ONT reads, with parameters such as corOutCoverage set to 120, corMinCoverage at 2, and minOverlapLength set to 1000 to optimize read overlap. The correction and trimming steps were performed on a high-performance computing server equipped with dual Intel® Xeon® E5-2698 v4 processors (80 CPU cores in total) and 755 GB of system memory, utilizing multi-threaded parallel computation to accelerate processing. Illumina short reads were then mapped to the trimmed mitochondrial long reads with Bowtie 2 (v2.5.1) with default parameters [10] to select the relevant mitochondrial short reads. A hybrid assembly of long and short mitochondrial reads was conducted using Unicycler (v0.4.8) [11] with default settings. The generated GFA format assembly was visualized using Bandage (v0.8.1) [12].

After the genome assembly, gene annotation was conducted using the Plant Mitochondrial Genome Annotator (PMGA, v1.5.3), which is tailored for organelle genome annotation (Li et al., 2025). The PMGA tool was run within a reproducible computational setup managed by mamba and utilizing a singularity-based containerized system (v3.7.2) to ensure cross-platform compatibility. For accurate gene annotation, the internal reference dataset (Dataset1) bundled with PMGA, which includes carefully curated mitochondrial gene models, was employed. The initial annotation, generated automatically, was then thoroughly examined and refined using Apollo (v2.5.0) (Lee et al., 2013), making necessary adjustments to gene boundaries and structural features. The finalized high-confidence GFF3 annotation file, together with the corresponding FASTA genome sequence, was submitted to the GenBase database (https://ngdc.cncb.ac.cn/genbase/?lang=en) for public availability and future reference (Bu et al., 2024).

Assembly of the cpgenome and detection of mitochondrial plastid-derived sequences

In the mitogenome assembly process described in Sect. 2.3, a product resembling the chloroplast genome was also obtained, as indicated by its circular structure and length. To verify this observation, the sequence was subsequently annotated using the CPGAVAS2 web server (http://47.96.249.172:16019/analyzer/annotate) for functional prediction of the genes [28]. The plastome maps for *I. seguinii* were visualized using OGDRAW (v1.3.1) (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html), which allowed for a detailed graphical representation of the genome's structure [29].

Mitochondrial plastid DNAs transfer (MTPTs) contributes to the variability in mitogenome structures across plant species, with dispersed repetitive sequences facilitating homologous recombination. This recombination results in the formation of multiple genome conformations and increased complexity [30–33]. To detect homologous sequences transferred between the plastome and mitogenome of *I. seguinii*, BLASTn (v2.13.0) was applied [34]. This analysis identified the presence of MTPTs—fragments of plastid DNA integrated within the mitogenome, a common occurrence in plants [33, 35]. In this case, the mitogenome was used as the target sequence for constructing the database, while the plastome served as the query sequence, with BLASTn parameters set to 'evalue 1e-6'. The resulting MTPTs were visualized using TBtools (v2.010) [36] and further

annotated through the GESeq platform (https://chlorobox.mpimp-golm.mpg.de/geseq.html) [37].

Analysis of repeats and repeat-mediated homologous recombination in mitogenome

Eukaryotic genomes are known to contain a considerable proportion of repetitive elements, which can be categorized into simple repetitive sequences, tandem repeats, and dispersed repeats, depending on their structure and distribution within the genome. Simple Sequence Repeats (SSRs), which are tandem repeats of 1–6 nucleotides, are widely spread across eukaryotic genomes [38–41]. SSRs were identified using the misa.pl (v2.1) tool [42], with thresholds set for mono-, di-, tri-, tetra-, penta-, and hexanucleotides as 10, 5, 4, 3, 3, and 3, respectively. A maximum difference of 1000 bp between two SSRs was allowed. Tandem repeats consist of multiple adjacent copies of repeat units of at least 7 nucleotides in length, and are found extensively in both eukaryotic and certain prokaryotic genomes. To detect these, Tandem Repeats Finder (TRF v4.09) [43] was employed, using the parameters '2 7 7 80 10 50 500 -f -d -m'. Dispersed repeats were identified with REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer/) [44], using the following settings: Ham ming Distance of 3, Maximum Computed Repeats of 500, and a Minimal Repeat Size of 30. The identified repeat elements were visualized using Circos (v0.69-8) [45].

To explore the role of repeat elements in the mitogenome as potential mediators of homologous recombination, pairs of repeats and their 100 bp flanking regions were extracted. These sequences were used to construct two genomic conformations: a reference and a recombinant conformation, based on recombination. Nanopore sequencing reads were aligned to both conformations to assess the potential of repeat pairs to facilitate recombination. For verification, PCR amplification was performed with a reaction mixture containing 1 µl (60–200 ng) of genomic DNA template, 1 µl of 8 µM forward and reverse primers, 13 µl of 2× Taq plus Mater Mix with Dye (BL553A, biosharp, China; containing Taq DNA polymerase, dNTPs, Mg^{2+}, and reaction buffer), and 10 µl of ddH2O. The PCR program included an initial denaturation at 96 °C for 3 min, followed by 32 cycles of denaturation at 94 °C for 30 s, annealing at 62 °C for 30 s, extension at 72 °C for 1 min, and a final extension at 72 °C for 10 min. The PCR products were then sequenced using the Sanger method.

RNA-editing sites prediction and codon usage analysis

RNA editing is a widespread post-transcriptional modification found in the mitogenomes of plants [46, 47]. The PREPACT3 tool (available at http://www.prepact.de/prepact-main.php) was employed to predict RNA-editing sites, with the data accessed on June 6, 2024. For the analysis, a set of 25 species from the mitochondrion database was selected as protein references, and the BLASTX threshold was set to 0.001 [48].

Codon usage refers to the non-random usage of synonymous codons in protein-coding sequences and has been widely described across both prokaryotic and eukaryotic organisms [49, 50]. This preference is believed to be influenced by factors such as natural selection, species-specific variations, and genetic drift [51, 52]. To analyze codon preference in *I seguinii*, PCGs were extracted from the mitogenome using TBtools (v2.010), and the relative synonymous codon usage (RSCU) values were calculated with CodonW (v1.4.4) [36].

Phylogenetic and collinearity analyses

For the phylogenetic analysis, a total of 29 species with high-quality mitogenome assemblies were selected, representing a broad taxonomic range across 10 plant families and six orders, along with *I. seguinii* (Icacinaceae) (Table S2). These taxa included representatives from Apocynaceae, Solanales, Gentianales, Lamiales, Fabales, Garryales, Sapindales, and other related groups, thereby providing a comprehensive framework for resolving the phylogenetic placement of *I. seguinii* within angiosperms. *Arabidopsis thaliana* (Brassicaceae) was designated as the outgroup.

Mitogenome sequences in GenBank format were imported into PhyloSuite (v1.2.3) for dataset standardization and feature extraction [53]. Using the "EXTRACT" module, all mitochondrial protein-coding genes (PCGs), tRNAs, and rRNAs were retrieved from each species. Multiple sequence alignment was then performed using MAFFT (v7.313) [54, 55], and the alignment accuracy was further optimized with MACSE (v2.0) to correct potential frameshifts and stop codon errors [56–58]. Conserved sequence blocks were subsequently identified using Gblocks (v0.91b) to remove poorly aligned or ambiguous regions. The best-fit nucleotide substitution model for each partition was determined using the ModelFinder module integrated within IQ-TREE2 (v2.1.4) [59]. Phylogenetic inference was then conducted under the maximum likelihood (ML) framework, with bootstrap support values and SH-aLRT branch test with 1000 replicates calculated to assess node reliability. The resulting tree topology was visualized and annotated in iTOL (v6) (https://itol.embl.de/) [60].

Following the phylogenetic reconstruction, the types and numbers of mitochondrial PCGs from all analyzed species were statistically summarized using an in-house Python script to ensure consistency across taxa (Supplemental_code_2). Subsequently, the PCGs used for phylogenetic analysis were employed to calculate the nonsynonymous (Ka) and synonymous (Ks) substitution rates using KaKs_Calculator (v3.0) [61, 62], with *I.*

seguinii serving as the reference species. The aligned PCG sequences generated by MAFFT were finally visualized and manually inspected using Geneious Prime (v2024.0.7) to evaluate codon alignment accuracy and confirm the distribution of conserved and variable regions.

For the collinearity analysis, five representative species from distinct evolutionary lineages were selected based on the phylogenetic relationships inferred from the mitochondrial genome dataset. These included *I. seguinii* and four other angiosperms representing diverse taxonomic clades, ensuring a broad comparison across lineages. Conserved collinear blocks among these mitogenomes were identified using BLASTn (v2.13.0) with an *E*-value threshold of 1e − 6, retaining homologous sequences longer than 500 bp [34]. The structural synteny and conserved genomic regions were visualized using both NGenomeSyn (v1.41) [63] and Mauve (v2.3.1) [64], which enabled the detection of locally collinear blocks (LCBs) and lineage-specific genome rearrangements. This combined approach provided a comprehensive view of mitochondrial genome conservation and structural evolution among distantly related species.

Results

Genomic features and structure of the *I. seguiniiis* mitogenome

All of the raw sequencing data of *Iodes seguinii* are available in the Genome Sequence Archive (GSA) of the National Genomics Data Center (NGDC, https://ngdc.cncb.ac.cn) under project accession number PRJCA034565 and GSA accession number CRA021844, with the Illumina dataset registered as CRX1326754 and the ONT dataset as CRX1326755 [65].The sequencing process generated approximately 52 Gb of ONT long-read data, achieving an N50 value of 1,531 bp, and 26 Gb of paired-end Illumina reads, each 150 bp in length. Second-generation sequencing generated 95,907,578 high-quality reads, while third-generation sequencing produced 18,793,638 clean reads. The final assembly yielded a complete circular mitochondrial genome of 423,925 bp, as shown in Fig. 1B. This genome has been deposited in GenBase at the NGDC under the registration number C_AA107961.1.

The mitogenome of *I. seguinii* encodes a total of 32 PCGs, including both core and variable genes, along with three ribosomal RNA (rRNA) genes and 23 transfer RNA (tRNA) genes (Table 1). Among the core PCGs, five are associated with ATP synthase (*atp1*, *atp4*, *atp6*, *atp8*, *atp9*), nine with NADH dehydrogenase (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*), one with ubiquinol cytochrome c reductase (*cob*), four with cytochrome c biogenesis (*ccmB*, *ccmC*, *ccmFC*, *ccmFN*), and three with cytochrome *c* oxidase (*cox1*, *cox2*, *cox3*). The mitogenome also contains a single maturase gene (*matR*) and a transport membrane protein gene (*mttB*). The variable PCGs include four large subunit ribosomal protein genes (*rpl2*, *rpl5*, *rpl16*), five small subunit ribosomal

Table 1 Gene composition and functional categorization of the *I. seguinii* mitogenome. The mitochondrial genes of *I. seguinii* can be categorized into several functional groups. These include core genes involved in essential processes such as ATP synthesis (ATP synthase), NADH dehydrogenation, ubiquinol cytochrome c reductase activity, cytochrome c biosynthesis, cytochrome c oxidation, and maturase functions. Additionally, there are variable genes, which encompass ribosomal proteins from both the large and small subunits, succinate dehydrogenase, and membrane transporters. The genome also contains sequences derived from plastids, as well as genes encoding ribosomal RNA and transfer RNA

Gene classification	Functional role	Gene list
Core genes	ATP synthase	*atp1*, *atp4*, *atp6*, atp8, *atp9*
	NADH dehydrogenase	*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*
	Ubichinol cytochrome c reductase	*cob*
	Cytochrome c biogenesis	*ccmB*, *ccmC*, *ccmFC*, *ccmFN*
	Cytochrome c oxidase	*cox1*, *cox2*, *cox3*
	Maturases	*matR*
	Transport membrane protein	*mttB* (*tatB*)[#]
Variable genes	Large subunit of ribosome	*rpl2*, *rpl5*, *rpl16*
	Small subunit of ribosome	*rps1*, *rps3*, *rps4*, *rps7*, *rps12*[*], *rps13*, *rps14*[*], *rps19*
	Succinate dehydrogenase	*sdh3* (x2), *sdh4*
rRNA genes	Ribosome RNA	*rrn5*, *rrn18*, *rrn26*
tRNA genes	Transfer RNA	*trnI-AAU*, *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnG-GCC*, *trnH-GUG* (x2), *trnK-UUU*, *trnI-CAU* (x2), *trnM-CAU* (x4), *trnN-GUU*, *trnP-UGG* (x2), *trnQ-UUG*, *trnS-GCU*, *trnS-UGA*, *trnW-CCA*, *trnY-GUA*

Gene copy numbers are indicated in parentheses where applicable. Asterisks (*) are used to highlight genes that exhibit variable presence in angiosperm mitogenomes, or those that may have undergone pseudogenization or are located in dual genomic regions [66–69]. Genes marked with a hash symbol (#) are infrequently found in land plant mitochondria, and their annotations should be considered provisional, potentially representing prokaryotic or horizontally acquired genes

protein genes (*rps1*, *rps3*, *rps4*, *rps7*, *rps19*), and two succinate dehydrogenase genes (*sdh3*, *sdh4*). The three rRNA genes include *rrn5*, *rrn18*, and *rrn26*, while the tRNA genes span 21 different codons, including *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnH-GUG*, and *trnY-GUA*. The presence of both core and variable genes provides a basis for exploring mitochondrial biology and evolution in *I. seguinii*.

Structural organization, genomic interactions, and evolutionary dynamics in the organellar genomes of *I. seguinii*

The cpgenome of *I. seguinii* was successfully assembled and annotated, and has been deposited in GenBase at the NGDC under the registration number C_AA070695.1. The annotated cpgenome reveals a typical circular structure containing functional gene regions related to photosynthesis, NADH dehydrogenase activity, ribosomal RNA, and transfer RNA, as well as large and small single-copy regions separated by a pair of inverted repeats (Fig. 2A). The gene arrangement and functional annotations highlight its role in photosynthetic and metabolic processes. A comparative analysis of the cpgenomes among *I. seguinii*, *I. cirrhosa* (GenBank: MF574684), *E. ulmoides*, and *Ipomoea x leucantha* (GenBank: NC_041208; Convolvulaceae) revealed significant conservation across species, including the presence of LSC, SSC, and IR regions. However, *I. seguinii* exhibited distinct differences in genome length and the size of the SSC and IR regions compared to its relatives (Fig. 2B). Notably, the SSC region of *I. seguinii* was shorter than that of *I. cirrhosa*, while its LSC region was elongated.

The interaction between the *I. seguinii* mitogenome and cpgenome was analyzed through the identification of shared sequences, termed MTPTs. A total of 33 MTPTs were identified in the mitogenome of *I. seguinii*, including several that contained partial plastidial PCGs. For instance, MTPT1 encoded fragments of *psbE*, MTPT9 contained parts of *atpA*, MTPT13 harbored sections of *ycf3*, MTPT15 also included *atpA*, while MTPT18 and MTPT19 contained fragments of *ycf2* (Fig. 2C).

Lastly, syntenic relationships between the *I. seguinii* mitogenome, cpgenome, and nuclear chromosomes were examined to identify NUMTs and NUPTs (Table S3). Collinear blocks and sequence alignments revealed homologous regions shared between the mitogenome and nuclear chromosomes (Fig. 2D) [70–73]. These observations describe the presence and distribution of organellar-derived sequences within the nuclear genome of *I. seguinii*. Together, these analyses provide a descriptive overview of the organellar genome features examined in this study.

Codon usage patterns and evolutionary insights in mitogenome of *I. seguinii*

The codon usage analysis revealed that certain codons with high relative synonymous codon usage (RSCU) values are preferentially used in the mitochondrial PCGs of *I. seguinii*, indicating their higher frequency in mitochondrial gene expression. Among the codons with RSCU values greater than 1.3, the most frequently used codons include AGA (Arg), UCU (Ser), GGA (Gly), GUU (Val), and CUU (Leu) (Fig. 3A, Table S4). These codons are likely optimized for more efficient protein translation in the mitochondrial environment [74–77]. In contrast, codons like CUG (Leu), CGC (Arg), and GGC (Gly) exhibited lower RSCU values, which suggests a reduced frequency of usage and potentially reflects translational or regulatory constraints within the *I. seguinii* mitogenome. Furthermore, the analysis showed a clear A/T bias at the third codon position, a common feature in plant mitogenomes, which may be linked to the replication and repair mechanisms specific to plant mitochondrial DNA [78]. Additionally, stop codons did not show a strong preference for specific codons, indicating functional neutrality in their selection across mitochondrial genes. These results provide a better understanding of the codon usage patterns within *I. seguinii* mitochondria.

Repeat element composition and structural insights into the single circular mitogenome of *I. seguinii*

As shown in Fig. 3B, the distribution of repeat types in the *I. seguinii* mitogenome reveals a clear dominance of forward repeats, which greatly outnumber the other categories (Table S5). The forward repeats, defined as direct repeat pairs oriented in the same direction (5′→3′ / 5′→3′), range in length from 75 to 339 bp, while the palindromic repeats, which exhibit inverted symmetry within a single sequence, vary between 75 and 343 bp, indicating substantial diversity in repeat structure [79]. Tandem repeats are also abundant, though slightly less frequent than forward repeats (Table S6), whereas reverse repeats, characterized by inverted orientation between paired sequences (5′→3′ / 3′←5′), occur in relatively smaller numbers. The types, lengths, and distribution of these repeat elements are summarized in Fig. 3C.

SSRs, comprising 86 units (Table S7), represent a significant feature of the *I. seguinii* mitogenome. As depicted in Fig. 3D, the most abundant SSR type is compound SSRs, consisting of combinations of multiple repeat units, particularly those involving hexanucleotide and longer repeats. Mono- and di-nucleotide SSRs are also present but less frequent compared to compound repeats. Tri-, tetra-, and penta-nucleotide SSRs are distributed moderately throughout the mitogenome. The overall distribution indicates that compound SSRs, especially those with

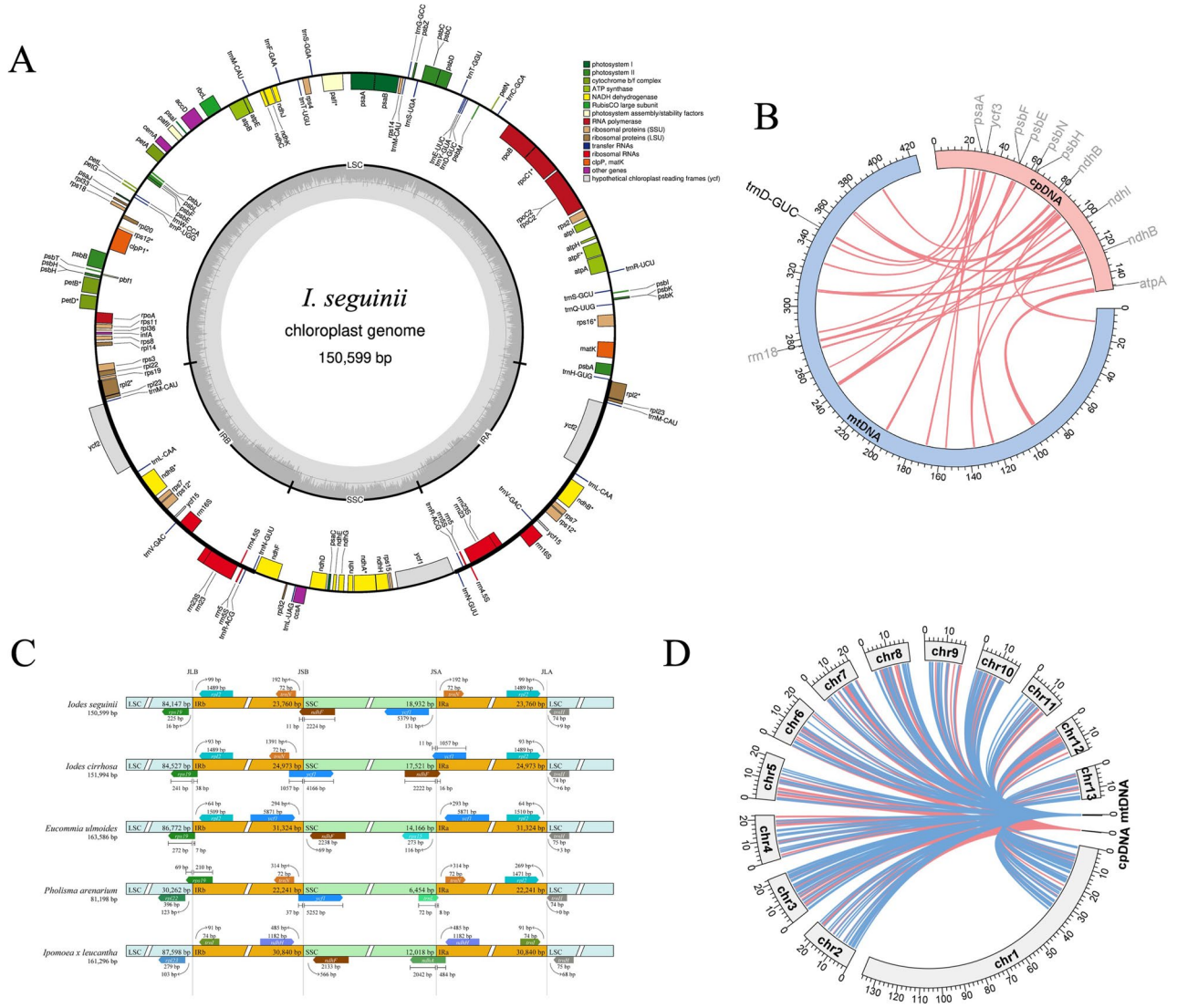

Fig. 2 Comprehensive analysis and interactions between the cpgenome and mitogenome of *I. seguinii*.**A** Circular map of the *I. seguinii* cpgenome. Functional gene categories are color-coded, including photosystem I and II, cytochrome b6/f complex, NADH dehydrogenase, and ribosomal RNA genes. The inner circle represents GC content variation, while the central region illustrates repeat-mediated connections across the genome. **B** Structural comparison of cpgenomes between *I. seguinii* and other species (*I. cirrhosa*, *E. alatum*, *P. amurense*, and *I. lacunosa*). Conserved regions, including LSC, SSC, and IR, are shown along with genome lengths. **C** Circos plot depicting homologous sequences shared between the *I. seguinii* cpgenome (cpDNA) and mitogenome (mtDNA). Red lines indicate regions of sequence homology, highlighting inter-organellar gene transfer events. **D** Circos plot illustrating syntenic relationships between the *I. seguinii* mitogenome, cpDNA and its nuclear chromosomes (chr1–chr12). Blue and red lines represent homologous connections, showcasing genomic interactions between the chloroplast, mitochondria, and nucleus

more than six repeat units, are a characteristic of the *I. seguinii* mitogenome.

Validation of forward and palindromic repeats in the *I. seguinii* Mitogenome

The structural analysis of the *I. seguinii* mitogenome revealed the presence of forward and palindromic repeats, which were experimentally validated to confirm their configuration and potential role in genomic organization. Forward repeats were identified as regions with high sequence similarity, represented schematically in [80, 81]. These repeats were hypothesized to facilitate homologous recombination and contribute to structural rearrangements within the mitogenome. An example of forward repeat validation through PCR amplification is shown in Fig. 4A, demonstrating the precise mapping and connectivity of the repeat region.

In addition to forward repeats, palindromic repeats were analyzed and visualized. These palindromic repeats, characterized by their sequence identity when read in both directions, were experimentally validated using specifically designed primer pairs to target junction regions

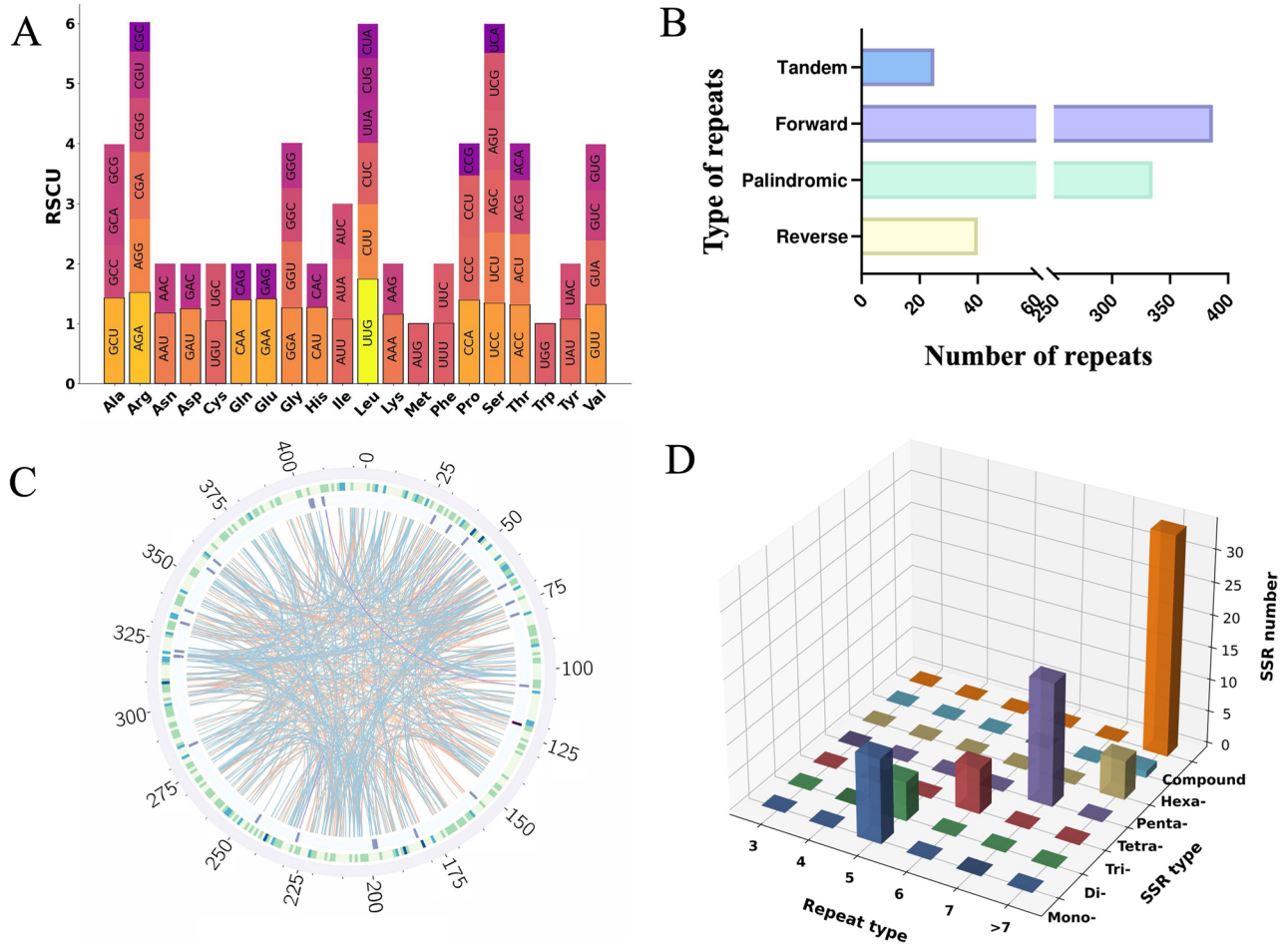

Fig. 3 Analysis of repeats and SSRs in the *I. seguinii* mitogenome. **A** RSCU distribution in the *I. seguinii* mitogenome. Each block corresponds to codons encoding a specific amino acid, with the height of the column representing the cumulative RSCU values for all codons associated with that amino acid. **B** Circos plot illustrating the distribution of dispersed repeats within the *I. seguinii* mitogenome. The colored lines represent different types of repeats, highlighting the genomic regions involved in repeat-mediated homologous recombination. **C** Classification and frequency of repeats in the mitogenome. The repeats are categorized into tandem, forward, palindromic, and reverse types, with forward repeats being the most prevalent. **D** Distribution of SSRs by type and repeat number. The SSR types include mononucleotide, dinucleotide, trinucleotide, and tetranucleotide repeats, with trinucleotide repeats showing the highest abundance across all categories

[82, 83]. Three such major repeats, P94, P103, and P127, were investigated in detail (Table 2). PCR amplification using forward-forward, reverse-reverse, and forward-reverse primer combinations produced distinct bands corresponding to expected fragment sizes, confirming the structural integrity of the inverted repeats (Fig. 4B-D). These results underscore the role of repetitive sequences in the structure of the *I. seguinii* mitogenome, a concept well-documented in the existing literature [80, 82, 84, 85]. Forward and palindromic repeats are likely involved in homologous recombination, contributing to genomic stability and facilitating potential structural rearrangements [86].

RNA editing patterns and their functional implications in *I. seguinii* mitogenome

RNA editing sites in protein-coding genes (PCGs) of the *I. seguinii* mitogenome were analyzed, describing patterns of post-transcriptional modification observed in this study (Table S8). A total of 501 RNA-editing sites were identified across 32 PCGs, predominantly involving C-to-U transitions. These edits resulted in amino acid changes that are critical for the functionality of the mitochondrial proteins (Fig. 5A). Among the PCGs, *nad6* and *ccmFN* displayed the highest number of RNA-editing sites, with 35 and 33 sites, respectively, followed by *ccmFC* and *nad1*, each with more than 25 editing sites. These findings suggest that RNA editing is highly concentrated in genes essential for mitochondrial energy metabolism and biogenesis.

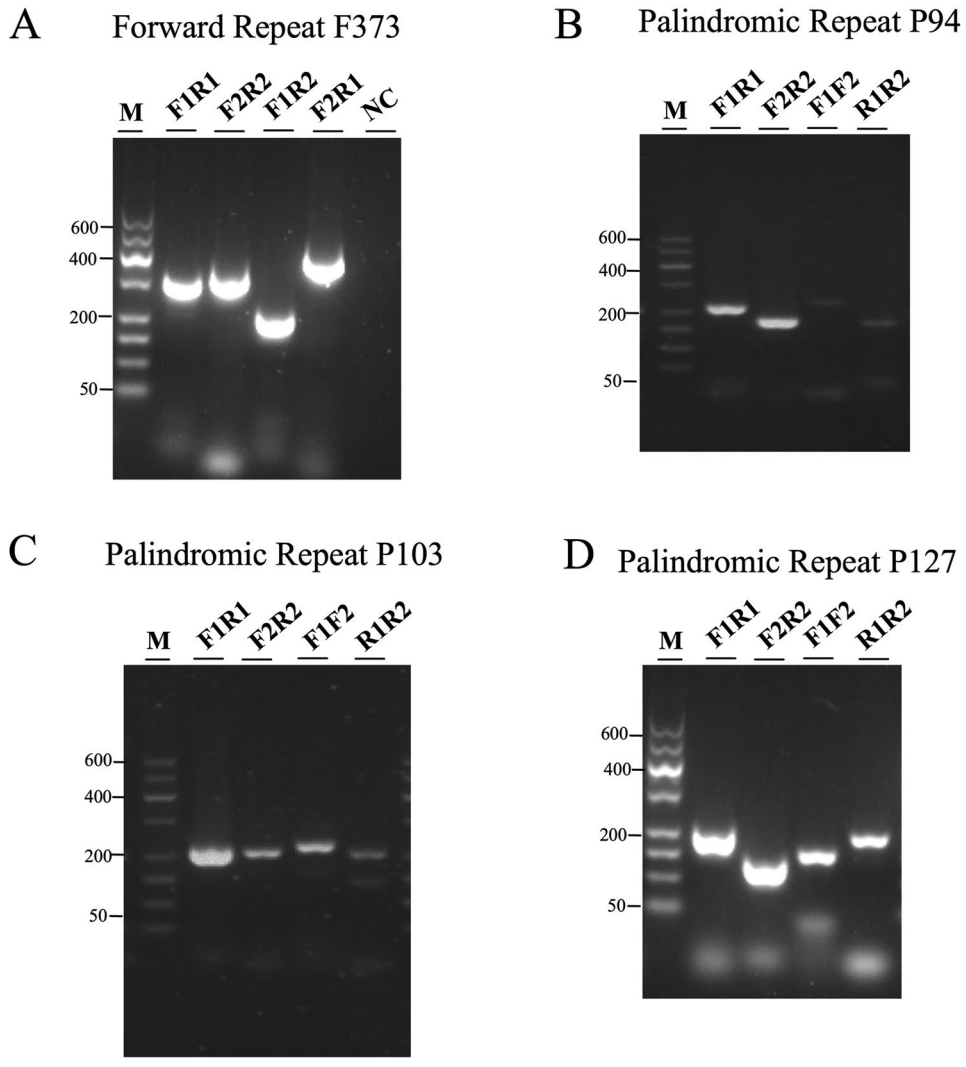

Fig. 4 PCR validation of forward and palindromic repeats in the *I. seguinii* mitogenome. **A** Gel electrophoresis results confirm the presence of Forward Repeat F373 in the *I. seguinii* mitogenome, with amplification products obtained using various primer combinations. **B** PCR amplification of Palindromic Repeat P94 shows the expected product size, while (**C**) PCR results for Palindromic Repeat P103 reveal the corresponding amplification bands. Finally, (**D**) PCR validation of Palindromic Repeat P127 demonstrates the presence of the repeat through the amplification of the expected product size

The analysis of amino acid substitutions resulting from RNA editing revealed clear and biologically meaningful patterns in the *I. seguinii* mitogenome (Fig. 5B). The most frequent conversions were Ser-to-Leu, Pro-to-Leu, and Ser-to-Phe, which together accounted for the majority of editing events. These transitions typically involve a shift from polar to hydrophobic residues, suggesting that RNA editing contributes to enhancing the structural stability and hydrophobicity of mitochondrial proteins, thereby promoting proper folding and membrane integration (Table S9). Beyond these dominant substitutions, additional changes such as Thr-to-Met, Ala-to-Val, and His-to-Tyr were also identified, indicating a broader spectrum of amino acid remodeling. Notably, several editing events led to the formation of novel start or stop codons, which could play a regulatory role in maintaining correct translation and protein maturation.

Comparative genomic features and evolutionary dynamics of mitochondrial PCGs across plant species

The analysis of the mitogenome size and GC content across 29 plant species, as shown in Fig. 5C, reveals significant variation in genome size, with species like *C. maxima* and *S. melongena* having larger mitogenomes compared to smaller ones such as *A. reptans* and *V. radiata*. The GC content also fluctuates across species, with a notable dip in *C. maxima* and *S. melongena*, suggesting potential correlations between genome size and nucleotide composition. Additionally, the large differences in mitogenome length across species imply corresponding variation in the length and organization of

Table 2 Primer design and PCR conditions for repeated sequences in *I. seguinii* mitogenome

Repeats	Primers	Sequence 5′-3′	Product Length of F1R1 and F2R2 (bp)	PCR Program
Forward Repeat F373	F373-F1	GGTTCGCAGCTTCTTAGAT	325	95℃ 3 min; 94℃ 15 s, 53℃ 30s, 72℃ 1 min (32 cycles); 72℃ 5 min
	F373-R1	TAGTCTTAGTAGTGGTTCTTCC		
	F373-F2	CAGAACTGGCTGACTAGAC	375	
	F373-R2	GAGAGTACGATAGGTACGATT		
Palindromic Repeat P94	P94-F1	TGCTGACTCTCCTAGTTAGA	219	
	P94-R1	AGTGTGCTGTTGACTTCAT		
	P94-F2	CTTAACCAGGAACAAGATAACG	179	
	P94-R2	ATGGAGTCTCGCAGAAGAA		
Palindromic Repeat P103	P103-F1	ATTCCGCTCGTTCCCTAT	225	
	P103-R1	GCAGATGGACTTACTTTACTAC		
	P103-F2	TTATAGTGCGGATGGACTTA	216	
	P103-R2	AACTCACTCTTCCAGATCC		
Palindromic Repeat P127	P127-F1	CGGATGCAGCTTTCTTTG	209	
	P127-R1	TGTACGAGTCAAGGATATGG		
	P127-F2	CAAAGAATCTCCGTCCCTA	150	
	P127-R2	GGCTAACTGAACTTACTTGA		

PCGs, reflecting species-specific genomic evolution. This is further illustrated in Fig. 5D, which presents a heatmap visualizing the copy number distribution of mitochondrial PCGs across 29 species. It shows varying copy numbers for mitochondrial genes such as ATP synthase, cytochrome c oxidase, and NADH dehydrogenase, with *I. seguinii* exhibiting higher copy numbers for genes like *atpA* and *atp9*. Notably, *S. miltiorrhiza* in the heatmap contain the *rbcL* gene, which is typically associated with the chloroplast genome, not the mitogenome. This is unusual, as it suggests possible horizontal gene transfer events between the mitochondria and plastid genomes, or potential contamination during sequencing. Such occurrences emphasize the need for careful genome annotation and validation to ensure accurate classification of genes within their respective organelles.

After processing the mitogenomes of 29 species through the PhyloSuite pipeline, a multiple sequence alignment of 13 mitochondrial PCGs was generated for Ka/Ks value analysis (Fig. 6A). This analysis provides insights into the selective pressures acting on these genes across species. The Ka/Ks ratio, which reflects the degree of selection, revealed distinct patterns for different genes. For example, the *nad9* gene (Fig. 6B) showed a relatively high Ka/Ks value, suggesting a degree of positive selection, while *ccmFC* (Fig. 6C) exhibited a ratio close to 1, indicating a neutral evolutionary pattern. In contrast, *nad3* (Fig. 6D) displayed a lower Ka/Ks value. Moreover, specific amino acid substitutions were observed at unique positions, with a distinctive mutation at position 11 in *nad9* and another at position 104 in *ccmFC*, both of which appear exclusive to *I. seguinii*. These unique substitutions, along with several other rare mutations found across the dataset, may reflect lineage-specific adaptive evolution in mitochondrial function.These variations in Ka/Ks ratios highlight the different evolutionary pressures and functional constraints that shape the mitochondrial genome across diverse plant species.

Synteny and repeat analysis reveal structural dynamics of the *I. seguinii* mitogenome

Comparative mitogenomic analyses revealed substantial genomic divergence and rearrangements in *I. seguinii*, emphasizing its distinct evolutionary history within angiosperms (Table S10). In the maximum likelihood phylogenetic tree constructed from concatenated mitochondrial PCGs (Fig. 7A), *I. seguinii* was positioned as a unique representative of the order Icacinales, clearly separated from other analyzed taxa. It formed a well-supported sister relationship with *E. ulmoides* (order Garryales), indicating a relatively close evolutionary connection between these two basal lineages within the malvid clade. Bootstrap support values at major nodes were consistently high, confirming the robustness of this phylogenetic placement. The color-coded branches, representing different plant families, further highlight the distinct evolutionary separation of *I. seguinii* from Solanales, Lamiales, and Brassicales species.

Synteny analysis between *I. seguinii* and four representative species—*C. maxima*, *N. attenuata*, *R. serrata*, and *V. radiata*—revealed extensive genomic rearrangements and variable degrees of collinearity (Fig. 7B). *I. seguinii* exhibited the highest syntenic coverage with *N. attenuata* (29.7%), followed by *C. maxima* (25.8%), *R. serrata* (21.7%), and *V. radiata* (19.9%). Despite moderate sequence identity across these comparisons, the rearranged homologous blocks indicate a history of structural reshuffling and repeat-mediated recombination events

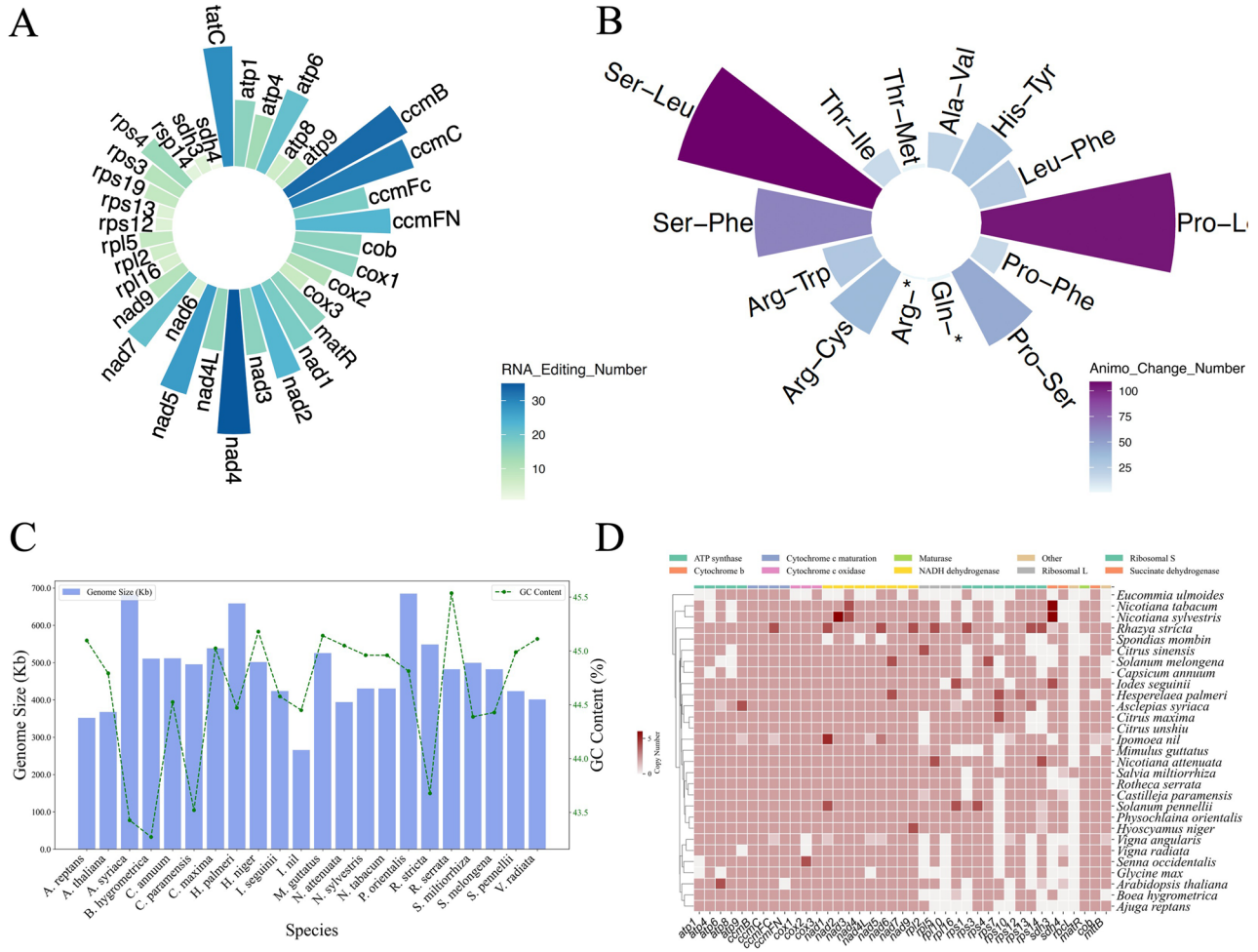

Fig. 5 RNA editing, amino acid substitutions, and genomic features in the I. seguinii mitogenome. **A** Distribution of RNA editing events across PCGs in the *I. seguinii* mitogenome, shown as a circular bar chart, with darker shades indicating genes with a higher number of editing events. **B** Circular bar chart illustrating the frequency of amino acid substitutions caused by RNA editing in the *I. seguinii* mitogenome, with prominent changes such as Ser→Leu and Pro→Leu. **C** Comparison of mitogenome size and GC content across 28 plant species. **D** Heatmap displaying the copy number of mitochondrial PCGs across 28 plant species, with functional categories indicated by color, revealing variation in gene retention and duplication among species

that have contributed to the mitogenome's structural diversification.

Further comparative alignment of *I. seguinii* with *A. thaliana* and *E. ulmoides* (Fig. 7C) revealed numerous inversions, translocations, and segmental duplications. While *E. ulmoides* displayed a more conserved mitochondrial architecture, *I. seguinii* showed extensive fragmentation and reorganization of collinear blocks. This pattern supports the notion that *I. seguinii* has undergone lineage-specific mitochondrial evolution, likely shaped by repeat-mediated recombination and inter-organelle DNA transfer.

Discussion

The *Iodes* genus has long been recognized for its medicinal importance, with species regionally used in traditional remedies due to their rich bioactive compounds and therapeutic potential. Among these, *I. seguinii*, an endemic species native to China, stands out for its potential in treating rheumatic immune diseases. However, genomic studies on *I. seguinii*, particularly its taxonomic classification and phylogenetic relationships, have been limited. Investigating its genetic architecture through mitogenomic analysis provides an excellent opportunity to uncover its evolutionary significance and further its medicinal applications. This study presents the first assemble and annotation report of the mitogenome of *I. seguinii*, providing describe features into its structural organization, evolutionary dynamics, and inter-organellar interactions within Icacinaceae. The 423,925 bp circular mitogenome of *I. seguinii* exhibits a conserved gene repertoire typical of angiosperms but also displays multiple repeats and rearrangements. This complexity is largely shaped by repeat-mediated recombination, extensive RNA editing, and MTPTs, all of which contribute to its structural divergence in this dataset.

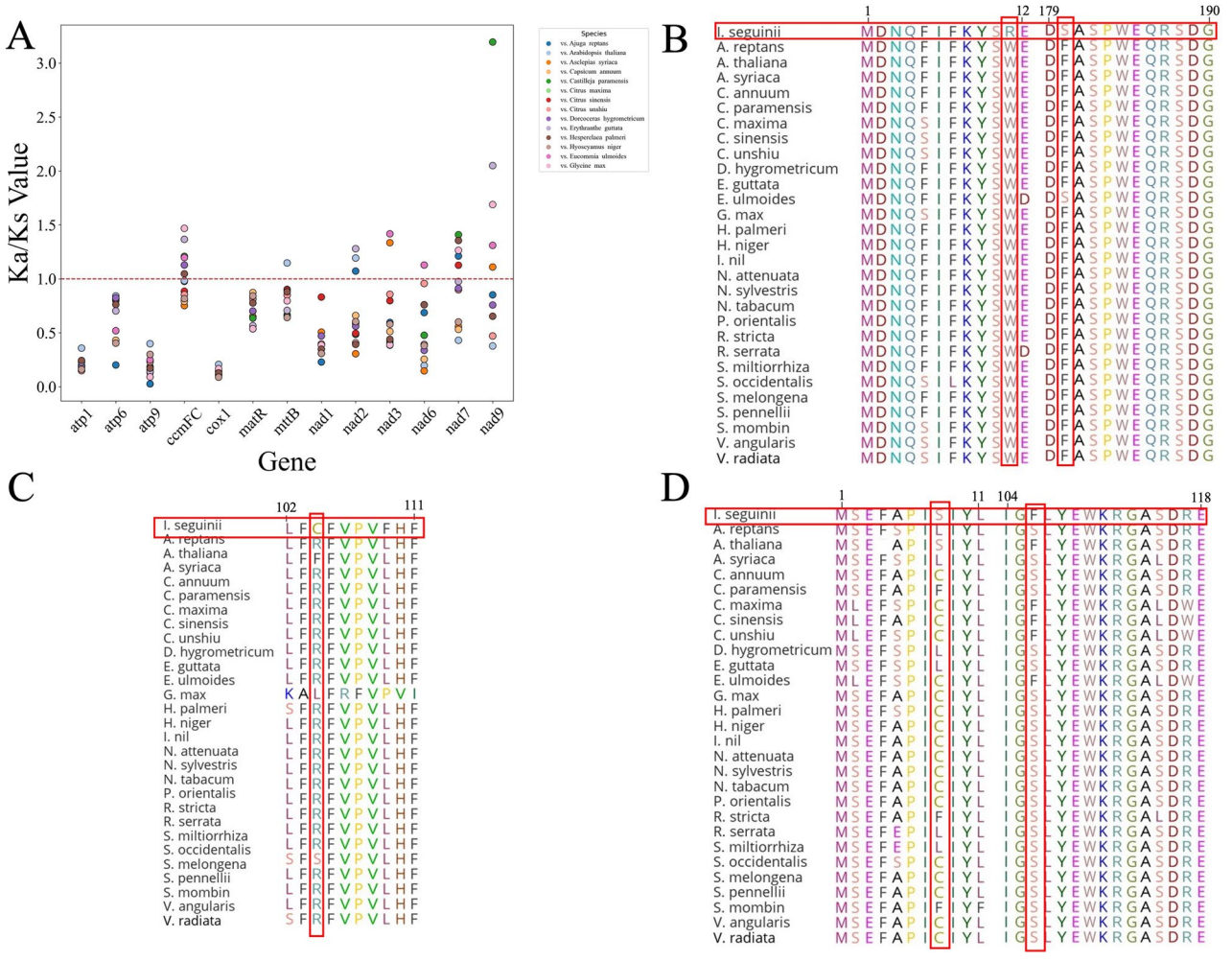

Fig. 6 Comparative Ka/Ks analysis and sequence alignment of mitochondrial genes across 14 species. **A** Pairwise Ka/Ks ratios for shared mitochondrial genes across 14 species, indicating selective pressures acting on each gene. Genes with a Ka/Ks value greater than 1 suggest positive selection, while values below 1 indicate purifying selection. **B** Multiple sequence alignment of the *nad9* gene, showing conserved regions and variation among species. **C** Alignment of the *ccmFC* gene, with corresponding sequence conservation and amino acid substitutions. **D** Sequence alignment of the *nad3* gene, highlighting the variation in amino acid residues and sequence identity across species

The repeat composition analysis demonstrated that forward and tandem repeats constitute the predominant types within the *I. seguinii* mitogenome, indicating their involvement in maintaining and reorganizing mitochondrial structure. Experimental validation through PCR further confirmed that both forward and palindromic repeats are capable of mediating homologous recombination, resulting in alternative genomic configurations. Such recombination events, frequently reported in other plant mitochondrial genomes, are recognized as major forces driving structural rearrangements and promoting genomic diversity during evolution [87–90]. Moreover, given the multicopy nature of plant mitochondria, even low-frequency recombination between dispersed repeats may give rise to coexisting structural isoforms, thereby enhancing the overall plasticity and evolutionary adaptability of the *I. seguinii* mitogenome.

Beyond structural variation, repeat-mediated recombination has been discussed in previous studies in the context of mitochondrial genome organization and diversity in plants [80, 91, 92]. In some cases, excessive or aberrant recombination gives rise to cytoplasmic male sterility (CMS), a phenomenon of considerable agricultural importance resulting from chimeric or rearranged mitochondrial genes that impair pollen development [83, 93]. Collectively, these findings highlight the dual role of repeat-mediated recombination in maintaining genomic stability while driving evolutionary innovation in the *I. seguinii* mitogenome and other angiosperms.

Analysis of nonsynonymous and synonymous substitution rates (Ka/Ks) among 13 conserved mitochondrial protein-coding genes revealed distinct evolutionary pressures. Most genes were under strong purifying selection, consistent with their essential roles in energy metabolism

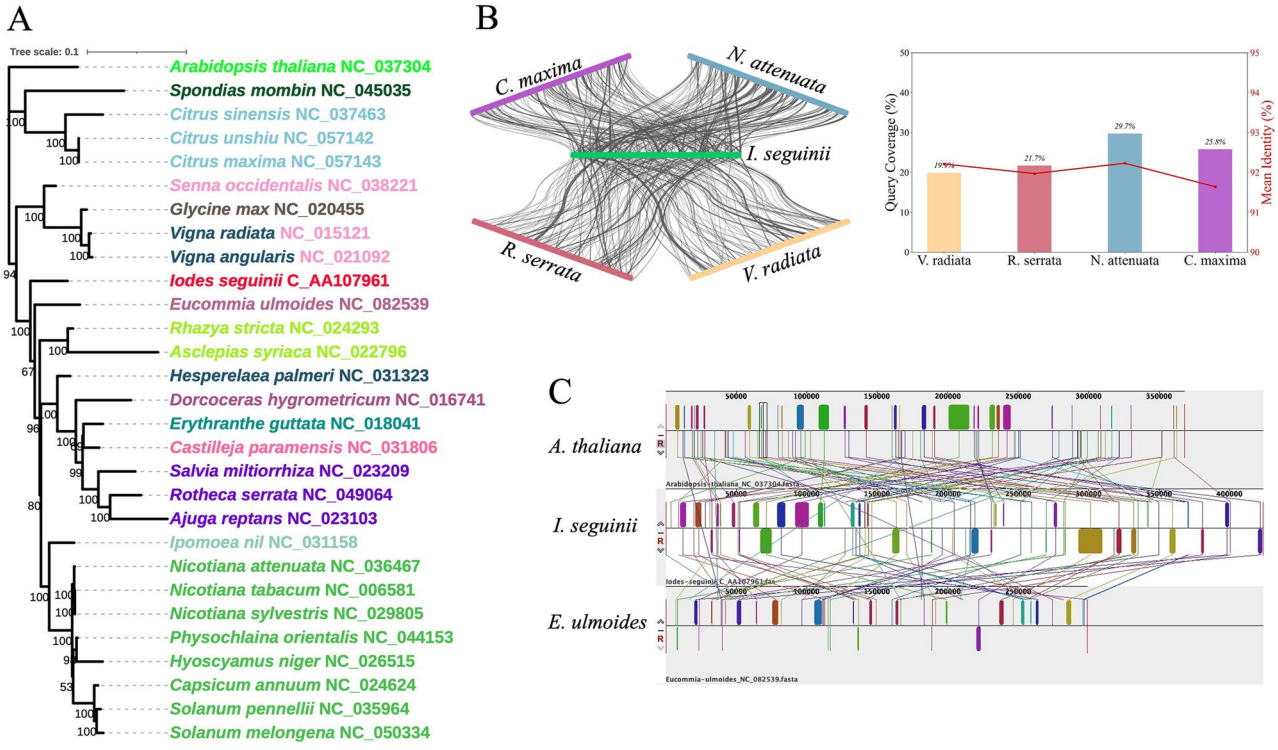

Fig. 7 Phylogenetic relationships and comparative mitogenomic synteny of *I. seguinii* and related species. **A** Maximum likelihood phylogenetic tree constructed from concatenated mitochondrial PCGs of *I. seguinii* and 28 representative angiosperm species, with *A. thaliana* used as an outgroup. Bootstrap support values are indicated at each node, and different colors represent distinct plant families, illustrating the evolutionary placement of I. seguinii within the Icacinaceae clade. **B** Synteny analysis of *I. seguinii* and four representative species (*C. maxima*, *N. attenuata*, *R. serrata*, and *V. radiata*), highlighting homologous genomic regions and sequence collinearity. The bar chart (right) summarizes overall coverage and sequence identity among the aligned mitogenomes. **C** Collinearity comparison among the mitogenomes of *A. thaliana* (top), *I. seguinii* (middle), and *Eucommia ulmoides* (bottom), showing conserved locally collinear blocks and major rearrangements, indicating lineage-specific structural evolution

and cellular respiration. However, *nad9* exhibited a relatively high Ka/Ks ratio, suggesting episodes of adaptive evolution possibly related to mitochondrial respiratory adaptation. In contrast, *ccmFC* showed near-neutral evolution, whereas *nad3* was highly conserved, reflecting strong functional constraints. Notably, unique amino acid substitutions—such as those at positions 11 in *nad9* and 104 in *ccmFC*—were exclusive to *I. seguinii*, suggesting lineage-specific adaptive signatures.

Comparative phylogenomic analyses across major angiosperm lineages showed that *I. seguinii* represents the only sampled representative of the order Icacinales in the current dataset. The maximum-likelihood phylogenetic reconstruction placed *I. seguinii* as a separate branch within the phylogeny, most closely related to *E. ulmoides* (Garryales), while *A. thaliana*—a well-established model species—was used as the outgroup for tree rooting. This phylogenetic placement indicates an early divergence of the Icacinales lineage relative to other sampled malvid orders, based on the current taxon sampling. The relatively long branch length associated with *I. seguinii* reflects substantial sequence divergence in the mitochondrial genome.

Synteny and collinearity analyses revealed pronounced genomic rearrangements between *I. seguinii* and other representative angiosperms, such as *N. attenuata* (Solanales) and *C. maxima* (Cucurbitales), consistent with the phylogenetic distances inferred from the tree topology. In contrast, *I. seguinii* shared relatively fewer collinear blocks with *A. thaliana*, indicating a higher degree of mitochondrial genome differentiation. Collectively, these observations describe notable structural divergence in the mitochondrial genome of *I. seguinii* relative to other sampled malvid taxa, based on comparative phylogenetic and synteny analyses.

This comprehensive mitogenomic assembly provides a foundational resource for studying the genetic and evolutionary dynamics of *I. seguinii*. The identification of recombination events, nucleotide composition biases in SSRs, and repeat-mediated genomic variations enhances our understanding of plant mitogenome plasticity. These findings also contribute to comparative genomics and phylogenetics, supporting future investigations into the taxonomy, evolutionary relationships, and genetic diversity within the *Iodes* genus.

Conclusion

This study reports the first complete mitochondrial genome of *I. seguinii* and represents the first mitogenome reported for the order Icacinales. Using a hybrid assembly strategy combining Oxford Nanopore long reads and Illumina short reads, a circular mitochondrial genome was assembled. Analyses of repeat elements, codon usage, RNA editing sites, and MTPTs were conducted. Repeat-mediated recombination was experimentally examined, and multiple genomic configurations associated with repeat regions were identified. Phylogenetic and synteny analyses described the placement of *I. seguinii* within the sampled angiosperm taxa and documented patterns of structural variation relative to other species. Overall, this study provides a descriptive mitochondrial genome dataset for *I. seguinii* that may serve as a reference for future comparative and evolutionary studies.

Abbreviations

PCGs	Protein-coding genes
NADH	Nicotinamide adenine dinucleotide hydride
ATP	Adenosine triphosphate
IR	Inverted repeat
LSC	Large single-copy
SSC	Small single-copy
SSRs	Simple sequence repeats
RSCU	Relative synonymous codon usage
N50	Median read length of the longest contigs
GC	Guanine-cytosine
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
mt	Mitochondrial contig
tRNAs	Transfer RNAs
rRNA	Ribosomal RNA
MSA	Multiple sequence alignment
Leu	Leucine
Val	Valine
Ser	Serine
Arg	Arginine
Gly	Glycine
APG	Angiosperm Phylogeny Group
DEPC	Diethyl pyrocarbonate
CTAB	Cetyltrimethylammonium bromide
ccm	Cytochrome c biogenesis genes
BP	Base pairs

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-026-12871-z.

Supplementary Material 1.

Authors' contributions

S.S. Yu and M. Wang contributed to the experimental design, data analysis, and manuscript preparation. X. Gong was responsible for sample collection and DNA extraction. M. Qin participated in genome assembly, annotation, and data visualization. M. Tang supervised the project, coordinated the research efforts, and finalized the manuscript. All authors reviewed and approved the final version of the manuscript.

Funding

This research was funded by the National Natural Science Foundation of China (Grant No. 32002235), the Science and Technology Plan Projects of Zhenjiang (Grant No. SH2023078), and the Medical Education Collaborative Innovation Fund of Jiangsu University (Grant No. JDYY2023009).

Data availability

All raw sequencing data generated in this study have been deposited in the GSA hosted by the NGDC, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences, and are publicly accessible at (https://ngdc.cncb.ac.cn/gsa) [65]. The data are available under project accession number PRJCA034565 and GSA accession number CRA021844, comprising sequencing results from two platforms: Oxford Nanopore Technology (ONT) and Illumina NovaSeq 6000. ONT sequencing (Experiment ID: CRX1326755, Sample ID: SAMC4514995) produced data stored under Run ID CRR1444242 (file name: CRR1444242.fq.gz), while Illumina sequencing (Experiment ID: CRX1326754, Sample ID: SAMC4514994) generated paired-end reads under Run ID CRR1444241 (files: CRR1444241_r1.fq.gz and CRR1444241_r2.fq.gz). In addition, the assembled mitochondrial and chloroplast genomes have been deposited in GenBase at NGDC under registration numbers C_AA107961.1 and C_AA070695.1, respectively.

Declarations

Ethics approval and consent to participate

This research utilized diploid *I. seguinii* samples collected from natural populations where the species is abundant. All sample collection and experimental procedures were conducted in accordance with relevant institutional, national, and international guidelines and regulations. No specific permissions or licenses were required for collecting samples from the wild.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 11 January 2025 / Accepted: 15 April 2026
Published online: 24 April 2026

References

1. Allen SE, Stull GW, Manchester SR. Icacinaceae from the eocene of Western North America. Am J Bot. 2015;102(5):725–44.
2. Su Y, Bai Q, Tao H, Xu B. Prospects for the application of traditional Chinese medicine network pharmacology in food science research. J Sci Food Agric. 2023;103(11):5183–200.
3. Chen M, Xiao J, El-Seedi HR, Wozniak KS, Daglia M, Little PJ, Weng J, Xu S. Kaempferol and atherosclerosis: From mechanism to medicine. Crit Rev Food Sci Nutr. 2024;64(8):2157–75.
4. Cassotta M, Forbes-Hernandez TY, Cianciosi D, Elexpuru Zabaleta M, Sumalla Cano S, Dominguez I, Bullon B, Regolo L, Alvarez-Suarez JM, Giampieri F et al. Nutrition and Rheumatoid Arthritis in the 'Omics' Era. Nutrients. 2021;13(3).
5. Hazarika S, Borah P, Deb PK, Venugopala KN, Hemalatha S. Icacinaceae Plant Family: A Recapitulation of the Ethnobotanical, Phytochemical, Pharmacological, and Biotechnological Aspects. Curr Pharm Des. 2023;29(15):1193–217.
6. Gan M, Zhang Y, Lin S, Liu M, Song W, Zi J, Yang Y, Fan X, Shi J, Hu J, et al. Glycosides from the Root of Iodes cirrhosa. J Nat Prod. 2008;71(4):647–54.
7. Ramesha BT, Suma HK, Senthilkumar U, Priti V, Ravikanth G, Vasudeva R, Kumar TR, Ganeshaiah KN, Shaanker RU. New plant sources of the anti-cancer alkaloid, camptothecine from the Icacinaceae taxa, India. Phytomedicine. 2013;20(6):521–7.
8. Thi Ngoc N, Hong Quang T, Thi Hong Hanh T, Xuan Cuong N, The Cuong N, Hoang Ha C, Hoai Nam N, Van Minh C. Phenolic Glycosides from the Leaves of Iodes cirrhosa Turcz. with Cytotoxic and Antimicrobial Effects. Chem Biodivers. 2022;19(9):e202200182.
9. Wang L, Zhang H, Jiang M, Chen H, Huang L, Liu C. Complete plastome sequence of Iodes cirrhosa Turcz., the first in the Icacinaceae, comparative genomic analyses and possible split of Idoes species in response to climate changes. PeerJ. 2019;7:e6663.
10. Gong X, Zhang H, Guo Y, Yu S, Tang M. Chromosome-level genome assembly of Iodes seguinii and its metabonomic implications for rheumatoid arthritis treatment. The Plant Genome; 2024.

11. Jiang Y, Liu M, Liu H, Liu S. A critical review: traditional uses, phytochemistry, pharmacology and toxicology of Stephania tetrandra S. Moore (Fen Fang Ji). Phytochem Rev. 2020;19(2):449–89.
12. Miao L, Zhang H, Cheong MS, Zhong R, Garcia-Oliveira P, Prieto MA, Cheng K-W, Wang M, Cao H, Nie S, et al. Anti-diabetic potential of apigenin, luteolin, and baicalein via partially activating PI3K/Akt/Glut-4 signaling pathways in insulin-resistant HepG2 cells. Food Sci Hum Wellness. 2023;12(6):1991–2000.
13. Weng J, Wu XF, Shao P, Liu XP, Wang CX. Medicine for chronic atrophic gastritis: a systematic review, meta- and network pharmacology analysis. Ann Med. 2023;55(2):2299352.
14. Conti P, Caraffa A, Gallenga CE, Ross R, Kritas SK, Frydas I, Younes A, Di Emidio P, Ronconi G, Pandolfi F. Powerful anti-inflammatory action of luteolin: Potential increase with IL-38. BioFactors. 2021;47(2):165–9.
15. Li J, Chang JY, Jiang ZL, Yin YK, Chen JY, Jin W, Li H, Feng L. Network Pharmacology and in vitro Experimental Verification on Intervention of Quercetin, Present in Chinese Medicine Yishen Qutong Granules, on Esophageal Cancer. Chin J Integr Med. 2023;29(3):233–43.
16. Liu J, Liu J, Tong X, Peng W, Wei S, Sun T, Wang Y, Zhang B, Li W. Network Pharmacology Prediction and Molecular Docking-Based Strategy to Discover the Potential Pharmacological Mechanism of Huai Hua San Against Ulcerative Colitis. Drug Des Devel Ther. 2021;15:3255–76.
17. Sun T, Quan W, Peng S, Yang D, Liu J, He C, Chen Y, Hu B, Tuo Q. Network Pharmacology-Based Strategy Combined with Molecular Docking and in vitro Validation Study to Explore the Underlying Mechanism of Huo Luo Xiao Ling Dan in Treating Atherosclerosis. Drug Des Devel Ther. 2022;16:1621–45.
18. Taban K, İlhan M, Süntar I. Luteolin: Advances on Resources, Biosynthesis Pathway, Bioavailability, Bioactivity, and Pharmacology. In: Handb Diet Flavonoids. 2023:1–37.
19. Liu J, Yu S, Lu P, Gong X, Sun M, Tang M. De novo assembly and characterization of the complete mitochondrial genome of Phellodendron amurense reveals three repeat-mediated recombination. Gene. 2025;935:149031.
20. Zhou G, Qin M, Liu X, Qi Y, Ou X, Tang M. De novo assembly of the mitochondrial genome of Glycyrrhiza glabra and identification of two types of homologous recombination configurations caused by repeat sequences. BMC Genomics. 2025;26(1).
21. Wang Z, Wang H, Gong X, Ou X, Guo Y, Tang M. De Novo assembly and phylogenetic analysis of the complete mitochondrial genome of Eleutherococcus senticosus and related araliaceous species. BMC Plant Biol. 2025;25(1).
22. Yu S, Qin M, Fleming E, Gong X, Tang M. Mitochondrial genome of Isatis indigotica reveals repeat-mediated recombination and phylogenetic insights in Cruciferae. Front Plant Sci. 2025;16.
23. Qin M, Lü P, Tang M, Yu S, Gong X. Characterization and evolutionary insights into complete mitochondrial genome of Sedum sarmentosum within the family Crassulaceae. Front Plant Sci. 2026;17.
24. Bonenfant Q, Noe L, Touzet H. Porechop_ABI: discovering unknown adapters in Oxford Nanopore Technology sequencing reads for downstream trimming. Bioinform Adv. 2023;3(1):vbac085.
25. De Coster W, Rademakers R. NanoPack2: population-scale evaluation of long-read sequencing data. Bioinf (Oxford England). 2023;39(5).
26. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinf (Oxford England). 2018;34(15):2666–9.
27. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinf (Oxford England). 2018;34(17):i884–90.
28. Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C. CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res. 2019;47(W1):W65–73.
29. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47(W1):W59–64.
30. Gualberto JM, Mileshina D, Wallet C, Niazi AK, Weber-Lotfi F, Dietrich A. The plant mitochondrial genome: dynamics and maintenance. Biochimie. 2014;100:107–20.
31. Su X, Dong Z, Yu H, Yi S, Gao S, Liu J, Kan S, Zhang W. Pan-mitogenome in Dipterocarpoideae: mitochondrial plastid DNAs and repeats shape the dynamic evolution of mitogenomes. BMC Plant Biol. 2025;25(1):1440.
32. Nhat Nam N, Pham Anh Thi N, Do HDK. New Insights into the Diversity of Mitochondrial Plastid DNA. Genome Biol Evol. 2024;16(9).
33. Park HS, Jayakodi M, Lee SH, Jeon JH, Lee HO, Park JY, Moon BC, Kim CK, Wing RA, Newmaster SG, et al. Mitochondrial plastid DNA can cause DNA barcoding paradox in plants. Sci Rep. 2020;10(1):6112.
34. Chen Y, Ye W, Zhang Y, Xu Y. High speed BLASTN: an accelerated MegaBLAST search tool. Nucleic Acids Res. 2015;43(16):7762–8.
35. Wang XC, Chen H, Yang D, Liu C. Diversity of mitochondrial plastid DNAs (MTPTs) in seed plants. Mitochondrial DNA DNA Mapp Seq Anal. 2018;29(4):635–42.
36. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol Plant. 2020;13(8):1194–202.
37. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45(W1):W6–11.
38. Tautz D, Renz M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res. 1984;12(10):4127–38.
39. Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res. 1989;17(16):6463–71.
40. Subramanian S, Mishra RK, Singh L. Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. Genome Biol. 2003;4(2):R13.
41. Srivastava S, Avvaru AK, Sowpati DT, Mishra RK. Patterns of microsatellite distribution across eukaryotic genomes. BMC Genomics. 2019;20(1):153.
42. Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L). TAG Theoretical Appl Genet Theoretische und angewandte Genetik. 2003;106(3):411–22.
43. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80.
44. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29(22):4633–42.
45. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639–45.
46. Lukeš J, Kaur B, Speijer D. RNA Editing in Mitochondria and Plastids: Weird and Widespread. Trends Genet. 2021;37(2):99–102.
47. Ye Z, Hu F, Zhang W, Fang D, Dong K, Cao J. Amino acid transporters of Brassica napus: Identification, evolution, expression and response to various stresses. Ind Crops Prod. 2023;194.
48. Lenz H, Hein A, Knoop V. Plant organelle RNA editing and its specificity factors: enhancements of analyses and new database features in PREPACT 3.0. BMC Bioinformatics. 2018;19(1):255.
49. Gouy M, Gautier C. Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. 1982;10(22):7055–74.
50. dos Reis M, Wernisch L. Estimating translational selection in eukaryotic genomes. Mol Biol Evol. 2009;26(2):451–61.
51. Hershberg R, Petrov DA. Selection on codon bias. Annu Rev Genet. 2008;42:287–99.
52. Bulmer M. The selection-mutation-drift theory of synonymous codon usage. Genetics. 1991;129(3):897–907.
53. Zhang D, Gao F, Jakovlic I, Zou H, Zhang J, Li WX, Wang GT. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 2020;20(1):348–55.
54. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
55. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30(14):3059–66.
56. Ranwez V, Chantret N, Delsuc F. Aligning Protein-Coding Nucleotide Sequences with MACSE. Methods Mol Biol. 2021;2231(1):51–70.
57. Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. Mol Biol Evol. 2018;35(10):2582–4.
58. Ranwez V, Harispe S, Delsuc F, Douzery EJ. MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. PLoS ONE. 2011;6(9):e22594.
59. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37(5):1530–4.
60. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics. 2007;23(1):127–8.

61. Zhang Z. KaKs_Calculator 3.0: Calculating Selective Pressure on Coding and Non-coding Sequences. Genomics Proteom Bioinf. 2022;20(3):536–40.
62. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteom Bioinf. 2010;8(1):77–80.
63. He W, Yang J, Jing Y, Xu L, Yu K, Fang X. NGenomeSyn: an easy-to-use and flexible tool for publication-ready visualization of syntenic relationships across multiple genomes. Bioinf (Oxford England). 2023;39(3).
64. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE. 2010;5(6):e11147.
65. Members C-N, Partners. Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2025. Nucleic Acids Res. 2025;53(D1):D30–44.
66. Sun YK, Gutmann B, Yap A, Kindgren P, Small I. Editing of Chloroplast rps14 by PPR Editing Factor EMB2261 Is Essential for Arabidopsis Development. Front Plant Sci. 2018;9:841.
67. Xiang KL, Mao W, Peng HW, Erst AS, Yang YX, He WC, Wu ZQ. Organization, Phylogenetic Marker Exploitation, and Gene Evolution in the Plastome of Thalictrum (Ranunculaceae). Front Plant Sci. 2022;13:897843.
68. Gao Y, Liu K, Li E, Wang Y, Xu C, Zhao L, Dong W. Dynamic evolution of the plastome in the Elm family (Ulmaceae). Planta. 2022;257(1):14.
69. Liu S, Wang Z, Wang H, Su Y, Wang T. Patterns and Rates of Plastid rps12 Gene Evolution Inferred in a Phylogenetic Context using Plastomic Data of Ferns. Sci Rep. 2020;10(1):9394.
70. Yoshida T, Furihata HY, Kawabe A. Analysis of nuclear mitochondrial DNAs and factors affecting patterns of integration in plant species. Genes Genet Syst. 2017;92(1):27–33.
71. Michalovova M, Vyskot B, Kejnovsky E. Analysis of plastid and mitochondrial DNA insertions in the nucleus (NUPTs and NUMTs) of six plant species: size, relative age and chromosomal localization. Heredity (Edinb). 2013;111(4):314–20.
72. Robles P, Quesada V. Organelle Genetics in Plants. Int J Mol Sci. 2021;22(4).
73. Richly E, Leister D. NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs. Mol Biol Evol. 2004;21(10):1972–80.
74. Mazumdar P, Binti Othman R, Mebus K, Ramakrishnan N, Ann Harikrishna J. Codon usage and codon pair patterns in non-grass monocot genomes. Ann Bot. 2017;120(6):893–909.
75. Li G, Zhang L, Xue P. Codon usage pattern and genetic diversity in chloroplast genomes of Panicum species. Gene. 2021;802:145866.
76. Wang ZK, Liu Y, Zheng HY, Tang MQ, Xie SQ. Comparative Analysis of Codon Usage Patterns in Nuclear and Chloroplast Genome of Dalbergia (Fabaceae). Genes (Basel). 2023;14(5).
77. Xu P, Zhang L, Lu L, Zhu Y, Gao D, Liu S. Patterns in Genome-Wide Codon Usage Bias in Representative Species of Lycophytes and Ferns. Genes (Basel). 2024;15(7).
78. Zhang Y, Shen Z, Meng X, Zhang L, Liu Z, Liu M, Zhang F, Zhao J. Codon usage patterns across seven Rosales species. BMC Plant Biol. 2022;22(1):65.
79. Yu J, Li J, Zuo Y, Qin Q, Zeng S, Rennenberg H, Deng H. Plastome variations reveal the distinct evolutionary scenarios of plastomes in the subfamily Cereoideae (Cactaceae). BMC Plant Biol. 2023;23(1):132.
80. Yang Z, Ni Y, Lin Z, Yang L, Chen G, Nijiati N, Hu Y, Chen X. De novo assembly of the complete mitochondrial genome of sweet potato (Ipomoea batatas [L.] Lam) revealed the existence of homologous conformations generated by the repeat-mediated recombination. BMC Plant Biol. 2022;22(1):285.
81. Jiang M, Ni Y, Zhang J, Li J, Liu C. Complete mitochondrial genome of Mentha spicata L. reveals multiple chromosomal configurations and RNA editing events. Int J Biol Macromol. 2023;251:126257.
82. Jiang M, Ni Y, Li J, Liu C. Characterisation of the complete mitochondrial genome of Taraxacum mongolicum revealed five repeat-mediated recombinations. Plant Cell Rep. 2023;42(4):775–89.
83. You C, Cui T, Zhang C, Zang S, Su Y, Que Y. Assembly of the Complete Mitochondrial Genome of Gelsemium elegans Revealed the Existence of Homologous Conformations Generated by a Repeat Mediated Recombination. Int J Mol Sci. 2022;24(1).
84. Cole LW, Guo W, Mower JP, Palmer JD. High and Variable Rates of Repeat-Mediated Mitochondrial Genome Rearrangement in a Genus of Plants. Mol Biol Evol. 2018;35(11):2773–85.
85. Li J, Xu Y, Shan Y, Pei X, Yong S, Liu C, Yu J. Assembly of the complete mitochondrial genome of an endemic plant, Scutellaria tsinyunensis, revealed the existence of two conformations generated by a repeat-mediated recombination. Planta. 2021;254(2):36.
86. Wang J, Kan S, Liao X, Zhou J, Tembrock LR, Daniell H, Jin S, Wu Z. Plant organellar genomes: much done, much more to do. Trends Plant Sci. 2024;29(7):754–69.
87. Rottmann WH, Brears T, Hodge TP, Lonsdale DM. A mitochondrial gene is lost via homologous recombination during reversion of CMS T maize to fertility. Embo j. 1987;6(6):1541–6.
88. Tanaka Y, Tsuda M, Yasumoto K, Yamagishi H, Terachi T. A complete mitochondrial genome sequence of Ogura-type male-sterile cytoplasm and its comparative analysis with that of normal cytoplasm in radish (Raphanus sativus L). BMC Genomics. 2012;13:352.
89. Wang R, Cai X, Hu S, Li Y, Fan Y, Tan S, Liu Q, Zhou W. Comparative Analysis of the Mitochondrial Genomes of Nicotiana tabacum: Hints Toward the Key Factors Closely Related to the Cytoplasmic Male Sterility Mechanism. Front Genet. 2020;11:257.
90. He T, Ding X, Zhang H, Li Y, Chen L, Wang T, Yang L, Nie Z, Song Q, Gai J, et al. Comparative analysis of mitochondrial genomes of soybean cytoplasmic male-sterile lines and their maintainer lines. Funct Integr Genomics. 2021;21(1):43–57.
91. Cao P, Huang Y, Zong M, Xu Z. De Novo Assembly and Comparative Analysis of the Complete Mitochondrial Genome of Chaenomeles speciosa (Sweet) Nakai Revealed the Existence of Two Structural Isomers. Genes (Basel). 2023;14(2).
92. Yang H, Chen H, Ni Y, Li J, Cai Y, Ma B, Yu J, Wang J, Liu C. 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(22).
93. Liberatore KL, Dukowic-Schulze S, Miller ME, Chen C, Kianian SF. The role of mitochondria in plant development and stress tolerance. Free Radic Biol Med. 2016;100:238–56.

Publisher's Note