



OPEN ACCESS

EDITED BY

Jim Leebens-Mack,
University of Georgia, United States

REVIEWED BY

Yu Bai,
Guiyang University, China
Guilong Lu,
Henan Institute of Science and Technology,
China
Guo Song,
Guangxi Science and Technology Normal
University, China

*CORRESPONDENCE

Min Tang
✉ mt3138@ujs.edu.cn

RECEIVED 28 June 2025

ACCEPTED 30 September 2025

PUBLISHED 15 October 2025

CITATION

Yu S, Qin M, Fleming E, Gong X and Tang M
(2025) Mitochondrial genome of *Isatis
indigotica* reveals repeat-mediated
recombination and phylogenetic
insights in *Cruciferae*.
Front. Plant Sci. 16:1655810.
doi: 10.3389/fpls.2025.1655810

COPYRIGHT

© 2025 Yu, Qin, Fleming, Gong and Tang. This
is an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Mitochondrial genome of *Isatis indigotica* reveals repeat- mediated recombination and phylogenetic insights in *Cruciferae*

Shaoshuai Yu¹, Meiling Qin², Emmanuel Fleming²,
Xun Gong³ and Min Tang^{2*}

¹Department of Pharmacy, Affiliated People's Hospital of Jiangsu University, Zhenjiang,
Jiangsu, China, ²School of Life Sciences, Jiangsu University, Zhenjiang, Jiangsu, China, ³Department
of Rheumatology and Immunology, Affiliated Hospital of Jiangsu University, Zhenjiang, Jiangsu, China

Isatis indigotica is an important medicinal plant extensively used in traditional Chinese medicine for its antiviral and anti-inflammatory properties. While previous studies have elucidated its nuclear and plastid genomes, the mitochondrial genome (mitogenome) —critical for understanding organellar evolution, intracellular DNA transfer, and stress response mechanisms—has remained uncharacterized. Here, we present a complete *de novo* assembly and comprehensive analysis of the *I. indigotica* mitogenome, generated using high-fidelity long-read sequencing technologies. The circular mitogenome spans 260,864 bp and encodes 31 protein-coding genes, 21 transfer RNAs, and 3 ribosomal RNAs. Repetitive sequences constitute 12.3% of the genome, with large repeats mediating homologous recombination and generating alternative conformations. A total of 488 RNA editing sites were identified, predominantly of the cytidine-to-uridine (C-to-U) type, indicating extensive post-transcriptional modification. We also detected 36 regions homologous to the plastid genome, reflecting active inter-organellar DNA transfer. Codon usage analysis revealed a preference for A/U-ending codons, and Ka/Ks analysis suggested strong purifying selection in most mitochondrial genes. Phylogenomic analysis based on 24 conserved mitochondrial genes placed *I. indigotica* in close proximity to *Brassica* species, supporting its taxonomic placement within the Brassicaceae family and aligning with plastid-based phylogenies. This study provides the first complete mitogenome of *I. indigotica*, offering valuable insights into mitogenome architecture, RNA editing dynamics, and plastid-mitochondrial interactions, while contributing to broader evolutionary and genomic understanding of cruciferous medicinal plants.

KEYWORDS

Isatis indigotica, mitogenome, repeat-mediated recombination, RNA-editing site, Brassicaceae

Introduction

Isatis indigotica Fortune, a biennial herb of the *Brassicaceae* (*Cruciferae*) family, has been widely cultivated in East Asia, particularly China, for both its medicinal and economic value (Chen et al., 2014; Shen et al., 2024). Known in traditional Chinese medicine as the source of *Isatidis Radix* and *Isatidis Folium*, its roots and leaves are used to treat febrile diseases, viral infections, and inflammatory conditions (Kong et al., 2008; Ding and Zhu, 2020; Sun et al., 2021). Phytochemical studies have identified a wealth of bioactive compounds including alkaloids, lignans, and flavonoids, contributing to its pharmacological efficacy. Recent functional analysis has further elucidated key glycosyltransferase genes involved in the biosynthesis of flavonoid glycosides, emphasizing the role of specialized metabolites in the plant's therapeutic profile (Tan et al., 2025). Despite its medicinal prominence, the taxonomic identity of *I. indigotica* has long been confounded with that of *Isatis tinctoria*, a European congener with historical relevance in dye production (Lu et al., 2018; Speranza et al., 2020). Chloroplast genome (cpgenome or cpDNA) comparison has provided molecular evidence distinguishing these species, with specific divergence in gene content and structure supporting their independent evolutionary lineages (Zhu et al., 2021; Wang et al., 2022; Zhang et al., 2022). Moreover, a chromosome-scale assembly has revealed a complex genomic architecture and biosynthetic gene clusters underpinning its medicinal traits, offering new resources for functional genomics (Kang et al., 2020).

Although nuclear and plastid genomes of *I. indigotica* have been increasingly characterized, its mitochondrial genome (mitogenome or mtDNA) remains unexplored. This presents a critical gap, as plant mitogenomes are not only crucial for cellular energy metabolism but are also recognized for their unique genomic properties—large size, high structural plasticity, and the prevalence of repeat-mediated homologous recombination (Liu et al., 2025; Wang et al., 2025; Zhou et al., 2025). Such recombination events can lead to genome rearrangements, multi-chromosomal configurations, and gene expression variability. In *Brassicaceae*, mitochondrial variation has been implicated in cytoplasmic male sterility, adaptive evolution, and interspecies hybridization (Chen et al., 2025; Jiang et al., 2025; Voisin et al., 2025; Zhao et al., 2025). The lack of mitochondrial data for *I. indigotica* limits phylogenetic resolution and hinders organelle-level evolutionary inferences, especially when nuclear and plastid markers show conflicting signals. Notably, chloroplast phylogenetic analysis has revealed a close sister relationship between *I. indigotica* and *Raphanus sativus* within the *Brassicaceae*, as evidenced by cpgenome data. This placement, while taxonomically consistent, warrants further validation using complementary data such as mitogenomes (Yang and Wang, 2017).

Plant mitochondrial genomes are complex and exhibit significant variability in both size and structure (Chevigny et al., 2020; Arimura and Nakazato, 2024; Forner, 2025). These genomes are essential for energy production, housing genes responsible for key components of respiration. While the gene content is relatively

minimal, the majority of the genome consists of non-coding regions that contribute to its dynamic structure. This flexibility allows for frequent rearrangements and the creation of multiple subgenomes within a single mitochondrion. Despite this structural instability, the gene sequences in plant mitochondrial genomes evolve slowly due to efficient repair mechanisms, particularly homologous recombination, which helps maintain genome integrity (Zwonitzer et al., 2024). The variability in the number of mitochondria and the genome copies within a cell further influences mitochondrial function. Recent advancements in genetic engineering tools have enabled precise modifications to mitochondrial DNA, providing new opportunities to study mitochondrial roles in plants and improve agricultural traits. In this study, we assembled and annotated the complete mitogenome of *I. indigotica* using high-fidelity (HiFi) long-read sequencing, and systematically analyzed its structural features, repeat elements, RNA editing sites, and phylogenetic position. Particular attention was paid to repeat-mediated recombination events and organellar DNA transfer, with the goal of elucidating the dynamic architecture and evolutionary trajectory of the mitogenome. The resulting data provide not only molecular tools for resolving taxonomic boundaries within *Isatis*, but also broader insights into mitogenome evolution across *Brassicaceae*.

Materials and methods

High-integrity DNA extraction and long-read sequencing

Young tender leaves of *I. indigotica* were sourced directly from Xianglian Horticulture, a verified commercial nursery located in Weitang Town, Xiangcheng District, Suzhou, China (31.2250° N, 120.6420° E) (Figure 1A). The identity of the entire plant was initially verified in-house by comparing morphological characteristics to the reference specimen image IMC0058462 from the National Plant Specimen Resource Center of China Digital Herbarium (NPSRCCDH). To reinforce the accuracy of this species determination, two independent experts from Jiangsu University—both with extensive experience in plant taxonomy—were subsequently consulted and confirmed the identification. Young leaves were freshly collected directly from potted *I. indigotica* plants and immediately rinsed with DEPC-treated water. The cleaned tissue was flash-frozen in liquid nitrogen and stored at −80 °C to preserve DNA integrity. High-molecular-weight genomic DNA was extracted using a modified CTAB protocol optimized for plant materials. DNA quality was assessed by 0.75% agarose gel electrophoresis, while purity and concentration were determined using both UV spectrophotometry (NanoDrop One, Thermo Fisher Scientific) and fluorometry (Qubit 3.0, Life Technologies). Only high-quality DNA was used for library construction with the SMRTbell Express Template Preparation Kit 2.0 (Pacific Biosciences), followed by long-read sequencing on the PacBio Sequel II system for accurate organelle genome assembly.

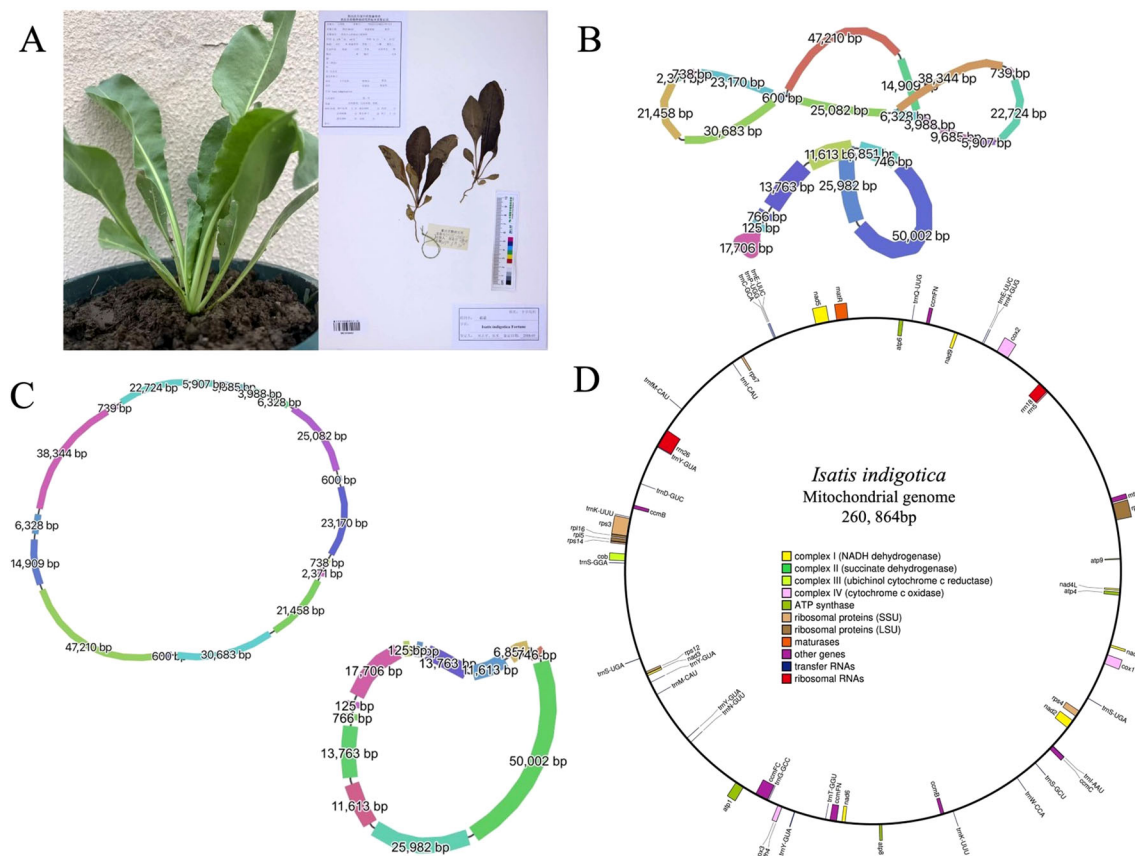


FIGURE 1

Morphological characteristics and organelle genome structures of *I. indigotica*. (A) Photograph of a living *I. indigotica* plant (left) and its herbarium specimen (right). (B) Graphical representation of the multipartite structure of the mitogenome (top) and cpgenome (bottom) of *I. indigotica*, with segment lengths (bp) labeled. Line thickness reflects relative sequencing depth, with thicker lines indicating higher coverage. (C) Circular maps of the mitogenome (top) and cpgenome (bottom), showing the lengths of each genomic segment. (D) Gene map of the complete mitogenome of *I. indigotica*, illustrating gene categories including NADH dehydrogenase, succinate dehydrogenase, cytochrome oxidase, ATP synthase, ribosomal proteins, maturases, tRNAs, and rRNAs.

Graph-based assembly and standardized annotation of mitogenome from high-fidelity reads

The mitogenome of *I. indigotica* was assembled using a multi-stage pipeline optimized for PacBio HiFi reads. To identify mitochondrial-derived sequences, HiFi reads were first aligned against three published *Brassicaceae* mitochondrial reference genomes: *Brassica rapa* (NC_049892.1) (Ren et al., 2023), *R. sativus* (NC_018551.1) (Chang et al., 2013), and *Sinapis arvensis* (KM851044.1) (Sang et al., 2020). Prior to alignment, BLAST databases were constructed for each reference using makeblastdb from BLAST+ (v2.13.0+), with default parameters and nucleotide indexing (Camacho et al., 2009). Read alignment was then conducted using blastn with the following parameters: -evalue 1e-5 -outfmt 6 -max_hsps 10 -word_size 7 -task blastn-short. The resulting alignments were filtered to retain only uniquely matched reads, and read IDs from all three comparisons were merged to generate a non-redundant mitochondrial candidate set. These filtered reads were extracted using seqkit grep and subsequently assembled *de novo* with PMAT (Plant Mitochondrial Assembler

Tool), employing the autoMito module with -st hifi mode, an estimated genome size of 300 Mb, and 24 computational threads. To improve the accuracy of the long-read data, PMAT, through its “correct_sequences.py” script, calls the Canu program to perform error correction, utilizing overlap-based assembly algorithms and error correction techniques. This procedure yielded a complete, circularized mitogenome of *I. indigotica* suitable for downstream analysis. To evaluate the assembly structure, the graphical layout in GFA format was inspected using Bandage (v0.8.1), allowing visual confirmation of the circular topology and connection paths among contigs (Wick et al., 2015).

After genome assembly, structural gene annotation was carried out using the Plant Mitochondrial Genome Annotator (PMGA, v1.5.3), a specialized tool designed for organelle genome annotation workflows (Li et al., 2025). The software was deployed within a reproducible computational environment managed by mamba, and executed inside a containerized system based on singularity (v3.7.2), ensuring compatibility and scalability across platforms. To improve annotation fidelity, the internal reference dataset (Dataset1) bundled with PMGA was utilized, offering curated models of mitochondrial genes. The initial automated annotation output

was carefully reviewed and refined using Apollo (v2.5.0) (Lee et al., 2013), with manual correction of ambiguous gene boundaries and confirmation of structural features. The resulting high-confidence GFF3 annotation file, after manual curation, was submitted along with the corresponding FASTA genome sequence to the GenBase database (<https://ngdc.cncb.ac.cn/genbase/?lang=en>) for public access and future reference (Bu et al., 2024).

Computational detection and experimental assessment of mitochondrial repeats and recombination

Repetitive sequences in eukaryotic genomes are broadly classified into simple sequence repeats (SSRs), tandem repeats, and dispersed repeats based on their structural features and genomic organization (Han et al., 2022; Ala et al., 2023; Li et al., 2023; Zhang et al., 2023; Wang et al., 2024). In this study, SSRs in the mitogenome of *I. indigotica* were identified using the *misa.pl* script (v2.1) with predefined thresholds: a minimum of 10 repeat units for mononucleotides, 5 for dinucleotides, and 4, 3, 3, and 3 for tri-, tetra-, penta-, and hexanucleotides, respectively (Thiel et al., 2003). A maximum distance of 1,000 bp was allowed between neighboring SSRs to detect potential compound repeats. Tandem repeats were identified using Tandem Repeats Finder (TRF, v4.09) under the parameter setting “2 7 7 80 10 50 500 -f -d -m”, which is optimized for the detection of tandem motifs in organelle genomes (Benson, 2023; Zhang et al., 2024). Dispersed repeats, defined as non-contiguous homologous sequences, were detected using the REPuter tool with the following parameters: minimum repeat length of 30 bp, maximum of 500 repeats per genome, and a Hamming distance threshold of 3 to accommodate minor mismatches (Kurtz et al., 2023). To gain insights into the distribution and organization of repeat elements across the mitogenome, all detected SSRs, tandem repeats, and dispersed repeats were mapped using Circos software (v0.69-8) (Krzywinski et al., 2023).

To assess whether dispersed repeat pairs mediate homologous recombination, two structural models were constructed for each candidate pair: (i) a reference configuration representing the native genomic arrangement, and (ii) a recombinant configuration simulating the outcome of recombination (Gualberto et al., 2014; Gualberto and Newton, 2017). For each model, 1000 bp flanking sequences surrounding the repeat pair were extracted. Nanopore long reads were then mapped to both models, and support for the recombinant structure was considered indicative of active recombination at that locus.

To experimentally validate selected recombination sites, PCR amplification was performed using primers designed to span the flanking regions of the repeat pair. Specifically, primers were designed in the 1000 bp regions upstream and downstream of each repeat: a forward primer (F) in the upstream region and a reverse primer (R) in the downstream region. For dispersed repeat pairs that are located far apart in the mitogenome, two sets of primers were used: F1/R1 and F2/R2 for the two respective repeat

loci. If direct (forward) recombination occurs, it can be detected by amplification of bands using the F1/R2 and F2/R1 primer combinations. Conversely, if inverted recombination occurs, it may be evidenced by amplification with F1/F1 or R1/R2 combinations. These diagnostic PCR patterns provide direct molecular evidence for the type and presence of recombination events mediated by repeat sequences [refer to Figure 2 in (Liu et al., 2025)].

Each 25 µl PCR reaction contained 1 µl genomic DNA, 1 µl each of 8 µM forward and reverse primers, 13 µl of 2× Taq PCR Master Mix, and 10 µl nuclease-free water. Thermocycling conditions were as follows: 96 °C for 3 min initial denaturation; 32 cycles of 94 °C for 30 s, 62 °C for 30 s, and 72 °C for 1 min; and a final extension at 72 °C for 10 min. PCR products were analyzed by agarose gel electrophoresis, and positive bands were sequenced using the Sanger method to confirm recombination events and evaluate repeat-associated structural variation.

Cpgenome recovery and annotation

During the structural correction phase of mitogenome assembly, the complete cpgenome of *I. indigotica* was incidentally resolved as an independent circular contig within the graphical assembly output. This structure was identified and separated using Bandage (v0.8.1) through examination of the GFA-format assembly graph generated from long-read sequencing data (Figures 1B,C). The circular chloroplast sequence was subsequently exported in FASTA format for further characterization.

Annotation of the cpgenome was conducted using the GeSeq web server, with *I. indigotica* Fortune employed as the reference genome to guide gene prediction and structural annotation (GenBank: KT939360) (Yang and Wang, 2017). The annotated cpgenome was visualized using OGDRAW (v1.3.1) (Greiner et al., 2019), resulting in a high-resolution circular genome map. The generated GFF3 file was subsequently curated manually to correct annotation discrepancies and feature boundaries. Both the finalized GFF3 and the corresponding FASTA file were submitted to the GenBase database for long-term archival and community access.

Species-level validation of the cpgenome was performed through comparative analysis with two previously published plastomes: *I. indigotica* (GenBank: OP620952) and *I. tinctoria* (GenBank: NC_028415), confirming its taxonomic identity.

Detection of mitochondrial-plastid and nuclear DNA transfer events

To investigate potential sequence transfer events between genomic compartments in *I. indigotica*, a comprehensive analysis of mitochondrial-plastid DNA transfer (MTPT) was conducted. The analysis employed BLASTn (v2.13.0) with the plastid genome used as the query and the mitogenome as the target. A stringent e-value cutoff of 1e-6 was set to ensure the detection of only high-

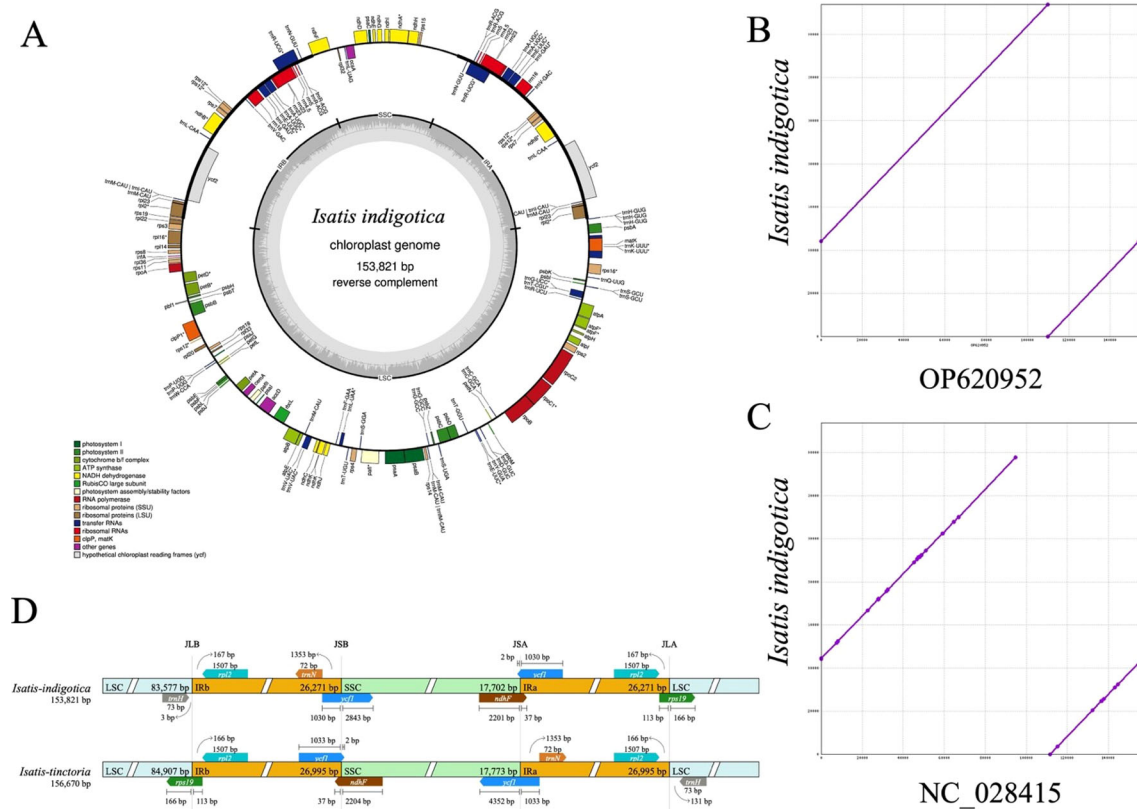


FIGURE 2

Structural and comparative analysis of the cpgenome of *I. indigotica*. (A) Circular map of the *I. indigotica* cpgenome (153,821 bp), with genes color-coded by functional category, illustrating the structural organization and annotation. (B) Sequence alignment of the *I. indigotica* cpgenome with accession OP620952 (same species) reveals near-complete sequence identity. (C) Comparative alignment with accession NC_028415 (*I. tinctoria*) shows minor sequence variations, reflecting interspecific genomic differences. (D) Structural comparison of the LSC, SSC, and IR boundary regions between *I. indigotica* and *I. tinctoria*, highlighting differences in boundary positioning and segment lengths.

confidence homologous regions (Chen et al., 2015). Regions of plastid origin identified within the mitogenome were visualized using TBtools (v2.010). The corresponding sequences were further annotated via GeSeq to assign functional roles and validate their organellar origin (Chen et al., 2020).

In parallel, the nuclear genome of *I. indigotica* (GenBank accession: VHIU00000000), which comprises seven chromosomes in its haploid complement, was also analyzed to assess possible organelle-to-nucleus DNA integration. This tri-compartmental comparative approach enabled the identification and characterization of both plastid-to-mitochondria and organelle-to-nucleus horizontal sequence transfers, offering insights into the dynamic genome interactions and evolutionary history of the species.

Dual-strategy identification of mitochondrial RNA editing sites and codon usage bias analysis

RNA editing is a crucial post-transcriptional process in plant mitochondria, predominantly characterized by cytidine-to-uridine (C-to-U) conversions. These events can restore conserved amino

acids, generate functional start or stop codons, and impact mitochondrial protein function and stability (Lukes et al., 2021). To comprehensively identify RNA editing events in *I. indigotica*, we employed two complementary strategies: (1) RNA-seq read mapping and variant calling, and (2) comparative prediction using conserved reference genomes.

For the mapping-based approach, publicly available RNA-seq data (SRA: SRR9329298, BioProject: PRJNA549758) were aligned to a custom-assembled mitogenome of *I. indigotica* (genbase: C_AA108663) (Kang et al., 2020). The processing pipeline consisted of fastp (v0.24.0) for adapter removal and quality filtering, fastQC (v0.12.1) for quality assessment, HISAT2 (v2.2.1) for splice-aware alignment, and SAMtools (v1.6) for BAM file handling (Li et al., 2009; Kim et al., 2015; Chen et al., 2018; Wingett and Andrews, 2018). Variant calling was performed using BCFtools (v1.21), and RNA editing candidates were filtered using custom scripts developed in-house (Narasimhan et al., 2016; Danecek and McCarthy, 2017; Genovese et al., 2024). In parallel, a reference-based prediction strategy was implemented using the PREPACT3 platform (<http://www.prepact.de/prepact-main.php>, accessed on 26 Jun 2025; version 3.12.0/2.2.26+) (Lenz et al., 2010). To maintain high specificity, only two mitochondrial

reference genomes were used: *Brassica napus* (NC_008285.1) and *Citrullus lanatus* (NC_014043.1). A stringent BLASTX e-value threshold of $1e-3$ was applied to identify conserved editing sites with high confidence, following established methodologies (Lenz and Knoop, 2013).

Codon usage optimization—defined as the preferential employment of synonymous codons to improve translational fidelity and effectiveness—is a recognized phenomenon in both unicellular and multicellular organisms (Presnyak et al., 2015; Sen et al., 2020). This codon usage bias arises from a combination of selective pressures, species-specific translational demands, and random genetic events (Hanson and Collier, 2018). In this study, mitochondrial protein-coding proteins (PCGs) of *Eleutherococcus senticosus* were subjected to codon bias evaluation. The analysis was conducted using TBtools software (v2.010), and relative synonymous codon usage (RSCU) values were determined via CodonW (v1.4.4) (Charif et al., 2005).

Evolutionary rate and nucleotide diversity analyses of mitochondrial PCGs in *I. indigotica* and related Brassicales species

Assessing the evolutionary patterns of PCGs is essential for understanding molecular adaptation and divergence among closely related species. In this study, the rates of nonsynonymous (K_a) and synonymous (K_s) substitutions were estimated to evaluate the selective pressures acting on mitochondrial PCGs of *I. indigotica* in comparison with 15 other representative species from the order Brassicales. The K_a/K_s ratio serves as an indicator of selection: values >1 suggest positive selection, $=1$ imply neutral evolution, and <1 indicate purifying selection.

Homologous gene pairs between *I. indigotica* and the other Brassicales species were identified using BLASTN (v2.10.1). Shared mitochondrial PCGs were aligned using MAFFT (v7.313) in auto mode to ensure accurate multiple sequence alignment (Rozewicki et al., 2019). K_a and K_s values were computed with the Maximum Likelihood (MLWL) method implemented in K_a/K_s Calculator v2.0 (Zhang et al., 2006; Wang et al., 2010; Zhang, 2022). The distribution of K_a/K_s ratios across genes was visualized using boxplots generated with the R package *ggplot2*. In parallel, to assess genetic variability at the nucleotide level, aligned homologous gene sequences were analyzed using DnaSP v5, and nucleotide diversity (π) values were calculated for each PCG (Rozas et al., 2003; Librado and Rozas, 2009; Rozas et al., 2017). This analysis provides insights into the extent of sequence polymorphism within mitochondrial genes, further contributing to our understanding of evolutionary constraints and divergence patterns across Brassicales mitogenomes.

Phylogenomic reconstruction and synteny mapping of *I. indigotica* mitogenome

To elucidate the phylogenetic position of *I. indigotica* within angiosperms, a comparative mitogenomic analysis was performed

using mitogenome sequences from 26 plant species representing four taxonomic orders. One species from the order Solanales was included as the outgroup to root the phylogenetic tree (Supplementary Table S1). Among the selected species, 16 belonged to the family Brassicaceae, including *I. indigotica* itself. These cruciferous taxa were: *Arabidopsis thaliana*, *Arabidopsis lyrata*, *Boechera stricta*, *Brassica napus*, *B. rapa*, *Brassica oleracea*, *Brassica juncea*, *Brassica carinata*, *Capsella rubella*, *Crucihimalaya lasiocarpa*, *Descurainia sophia*, *I. tinctoria*, *Lepidium apetalum*, *Lepidium sativum*, and *R. sativus*. To provide broader phylogenetic context, 10 additional non-Brassicaceae species were included: *Apium graveolens* (Apiaceae), *Panax quinquefolius* (Araliaceae), *Phaseolus vulgaris*, *Aeschynomene indica*, *Senna tora*, and *Vigna angularis* (all Fabaceae), *Saussurea costus* and *Taraxacum mongolicum* (Asteraceae), and the outgroup *Solanum aethiopicum* (Solanaceae). This diverse taxon set enabled robust resolution of both intra- and inter-family evolutionary relationships.

All genomic datasets were retrieved from GenBank and subsequently processed using PhyloSuite (v1.2.3), which facilitated the extraction and reformatting of mitochondrial coding sequences (Zhang et al., 2020; Xiang et al., 2023). Multiple sequence alignment (MSA) of conserved gene regions was performed using MAFFT (v7.313) to ensure high-accuracy alignment across all taxa involved (Rozewicki et al., 2019). To determine the most appropriate nucleotide substitution models for downstream phylogenetic inference, PartitionFinder2 was employed. Maximum likelihood phylogenetic reconstruction was carried out using IQ-TREE2 (v2.1.4) with ultrafast bootstrap support (Minh et al., 2020). The selection of the best nucleotide substitution models was based on the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), which evaluated various models to identify the one that best fit the data. The final result indicated that the GTR+F+I+G4 model was the most appropriate, as it provided the lowest log-likelihood value and the best fit according to both AIC and BIC criteria. The final tree visualization and annotation were achieved through iTOL (v6), offering customizable graphics for enhanced biological interpretation (Letunic and Bork, 2021).

To investigate mitogenome collinearity and structural conservation in *I. indigotica*, three additional mitogenomes—*Apium graveolens*, *Taraxacum mongolicum*, and *Vigna angularis*—were selected for comparative analysis. These species, representing different families within the eudicots, provided a broad phylogenetic context for assessing syntenic relationships. Pairwise comparisons were conducted using BLASTn, with an e-value cutoff of $1e-6$ to ensure specificity. Homologous regions exceeding 500 bp in length were designated as conserved collinear blocks. The collinearity relationships among *I. indigotica*, *I. tinctoria*, and the three selected eudicot species were visualized using NGenomeSyn (v1.41), enabling detailed inspection of genome rearrangements and conserved structural segments across lineages (He et al., 2023).

To assess large-scale sequence conservation and structural variation among closely related cruciferous species, comparative alignment of mitogenomes from *I. indigotica*, *I. tinctoria*, and *R. sativus* was conducted. Whole-genome alignments were generated using long-read-assembled mitogenomes and visualized with Mauve (v2.4.0) alignment software (<https://sourceforge.net/>

projects/mauve). The Mauve viewer enabled the detection of locally collinear blocks (LCBs) and facilitated inspection of genome rearrangements, inversions, and insertions/deletions across the three *Brassicaceae* species.

Results

Data generation and genome assembly

A total of 368,520 HiFi long reads were generated using the PacBio Revio platform, producing approximately 6.48 Gb of high-quality sequencing data. The mean read length was 17,596 bp, with a median length of 16,546 bp and a read length N50 of 17,479 bp, indicating a highly consistent and uniform long-read distribution across the dataset (Supplementary Figure S1A). The median read quality reached Q33, demonstrating the exceptional base-level accuracy achieved through PacBio's circular consensus sequencing (CCS) technology (Supplementary Figure S1B). The sequencing depth for the mitochondrial genome was approximately 24.9x, and for the chloroplast genome, the sequencing depth was approximately 42.2x, ensuring comprehensive coverage of both genomes. To ensure data transparency and facilitate future reuse, the raw HiFi sequencing data have been deposited in the Genome Sequence Archive (GSA) at the National Genomics Data Center (NGDC) under accession number CRA026873 (<https://ngdc.cncb.ac.cn/gsa>) (Members and Partners, 2025). Furthermore, the annotated mitogenome assembly has been submitted to GenBase, also hosted by NGDC/CNCB, and is accessible under accession number C_AA108663.1 (<https://ngdc.cncb.ac.cn/genbase>) (Bu et al., 2024). The final assembly comprises a complete circular mitogenome of 260,864 base pairs in length (Figure 1D).

Mitogenome classification and evolutionary features of mitochondrial genes

The mitogenome of *I. indigotica*, as annotated using the PMGA pipeline, comprises 65 unique genes, including 23 core PCGs, 9 variable genes, 3 ribosomal RNA (rRNA) genes, and 18 distinct transfer RNA (tRNA) species, with copy number variation observed in several tRNAs (Table 1). The core PCGs are involved in essential mitochondrial functions and include five ATP synthase subunits (*atp1*, *atp4*, *atp6*, *atp8*, *atp9*), nine NADH dehydrogenase genes (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*), one ubiquinol cytochrome c reductase gene (*cob*), and six genes related to cytochrome c biogenesis, with duplicated copies of *ccmB* and *ccmFN* (*ccmB* ×2, *ccmC*, *ccmFC*, *ccmFN* ×2). Three cytochrome c oxidase genes (*cox1*, *cox2*, *cox3*) and one maturase gene (*matR*) were also identified.

The variable gene set consists of three large ribosomal subunit genes (*rpl2*, *rpl5*, *rpl16*) and five small ribosomal subunit genes (*rps3*, *rps4*, *rps7*, *rps12*, *rps14*). Notably, *rps12*, *rps14*, and *rpl16* are

variably retained in angiosperm mitogenomes and are often considered non-canonical or plastid-derived, suggesting possible intracellular gene transfer or pseudogenization. The genome also encodes *sdh4*, a succinate dehydrogenase subunit, and *mttB*, a putative membrane transporter gene homologous to bacterial *tatB*, which may represent a case of horizontal gene transfer. The rRNA gene complement includes *rrn5*, *rrn18*, and *rrn26*, which constitute the core structural components of mitochondrial ribosomes. The tRNA set contains 18 distinct species, with multiple copies observed for *trnE-UUC* (×2), *trnK-UUU* (×2), *trnM-CAU* (×3), *trnS-UGA* (×2), and *trnY-GUA* (×4), providing redundancy typical of plant mitochondrial translation systems.

Synonymous codon preference in mitochondrial PCGs

The mitochondrial PCGs of *I. indigotica* display discernible codon usage biases, providing insights into the evolutionary dynamics and translational regulation within the mitogenome. A total of 61 sense codons encode the standard 20 amino acids, with the exception of methionine and tryptophan, both of which are encoded by single codons (Figure 3A). The distribution of relative synonymous codon usage (RSCU) values indicates varying degrees of codon preference, likely shaped by a combination of mutational biases, selective constraints, and tRNA availability.

Codons with RSCU values greater than 1 are considered preferentially used. Examples include GCT (alanine, 1.63), CGT (arginine, 1.26), AGA (arginine, 1.41), AAT (asparagine, 1.37), GAT (aspartic acid, 1.36), and TGT (cysteine, 1.24), suggesting favored usage during mitochondrial translation. In contrast, codons such as GCG (alanine, 0.49), CGG (arginine, 0.69), CGC (arginine, 0.53), AAC (asparagine, 0.63), and GAC (aspartic acid, 0.64) are underrepresented, with RSCU values below 0.7 (Supplementary Table S2). These patterns collectively suggest a moderate codon usage bias in *I. indigotica*, with a tendency toward A- and T-ending codons at the third position, consistent with observations in other angiosperm mitogenomes.

Comprehensive profiling of repetitive elements in mitogenome

A comprehensive analysis of SSRs in the mitogenome of *I. indigotica* revealed 51 distinct SSR loci, with repeat unit lengths ranging from 10 to 1,451 base pairs (Figure 3B). These SSRs were categorized into seven types—mononucleotide (p1), dinucleotide (p2), trinucleotide (p3), tetranucleotide (p4), pentanucleotide (p5), hexanucleotide (p6), and compound repeats (c)—and further grouped by repeat unit numbers (three to >7). Among them, compound SSRs were the most abundant (n = 14), followed by tetranucleotide repeats (n = 13) and dinucleotide repeats (n = 8) (Supplementary Table S3). Mononucleotide and trinucleotide motifs were observed six times each, while pentanucleotide and hexanucleotide repeats were less frequent, with only two and one

TABLE 1 Categorization of mitochondrial genes identified in *I. indigotica*.

Gene classification	Functional role	Gene list
Core genes	ATP synthase	<i>atp1, atp4, atp6, atp8, atp9</i>
	NADH dehydrogenase	<i>nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7, nad9</i>
	Ubichinol cytochrome c reductase	<i>Cob</i>
	Cytochrome <i>c</i> biogenesis	<i>ccmB</i> (x2), <i>ccmC</i> , <i>ccmFC</i> , <i>ccmFN</i> (x2)
	Cytochrome <i>c</i> oxidase	<i>cox1, cox2, cox3</i>
	Maturases	<i>matR</i>
Variable genes	Large subunit of ribosome (LSU)	<i>rbl2, rbl5, rpl16*</i>
	Small subunit of ribosome (SSU)	<i>rps3, rps4, rps7, rps12*, rps14*</i>
	Succinate dehydrogenase	<i>sdh4</i>
	Membrane transport	<i>mttB</i> (<i>tatB</i>) [#]
rRNA genes	Ribosome RNA	<i>rrn5, rrn18, rrn26</i>
tRNA genes	Transfer RNA	<i>trnC</i> -GCA, <i>trnD</i> -GUC, <i>trnE</i> -UUC (x2), <i>trnG</i> -GCC, <i>trnI</i> -AAU, <i>trnK</i> -UUU(x2), <i>trnM</i> -CAU(x3), <i>trnN</i> -GUU, <i>trnP</i> -UGG, <i>trnQ</i> -UUG, <i>trnS</i> -GCU, <i>trnS</i> -GGA, <i>trnS</i> -UGA (x2), <i>trnT</i> -GGU, <i>trnW</i> -CCA, <i>trnY</i> -GUA(x4), <i>trnH</i> -GUG

The numbers in parentheses represent gene copy numbers when applicable. Asterisks (*) denote genes with variable presence across angiosperm mitogenomes, or those showing signs of pseudogenization or dual genomic localization (Sun et al., 2018; Liu et al., 2020; Gao et al., 2022; Xiang et al., 2022). The hash symbol (#) marks genes that are rarely observed in land plant mitochondria; their annotation may be tentative and potentially of prokaryotic or horizontally transferred origin.

The mitochondrial genes of *I. indigotica* are classified into core functional groups (ATP synthase, NADH dehydrogenase, ubiquinol cytochrome *c* reductase, cytochrome *c* biogenesis, cytochrome *c* oxidase, and maturases), variable genes (including ribosomal protein genes of the large and small subunits, succinate dehydrogenase, and membrane transport proteins), plastid-derived sequences, ribosomal RNA genes, and transfer RNA genes.

instance(s), respectively. The high prevalence of compound and tetranucleotide SSRs suggests their potential role in promoting genomic plasticity and structural variability. These SSRs were visualized by repeat type and length category in a 3D bar plot.

In addition to SSRs, other repetitive elements were also analyzed, including dispersed and tandem repeats (Supplementary Table S4). A total of 396 dispersed repeats were identified, comprising 216 forward repeats and 180 palindromic repeats, with repeat lengths ranging from 30 bp to 11,396 bp. Tandem repeats were also detected, though less abundant. The spatial organization of these three major repeat types—SSRs (outermost circle), tandem repeats (middle circle), and dispersed repeats (innermost circle)—is illustrated in Figure 3C. In this circular map, green bars indicate SSR loci, light blue boxes represent tandem repeats, and arcs connect dispersed repeat pairs, with pink lines denoting forward repeats and purple lines indicating

palindromic repeats. This layered distribution provides insight into the repeat architecture and potential recombinogenic regions across the *I. indigotica* mitogenome.

Repeat-mediated structural rearrangements in mitogenome

To experimentally validate homologous recombination events facilitated by repetitive elements, five repeat pairs—two forward repeats (F96 and F146) and three palindromic repeats (P93, P134, and P326)—were selected for PCR-based analysis (Supplementary Table S5). Primers were designed to flank each repeat with ~100 bp of adjacent sequence. Two sets of primer combinations (F1/R1 and F2/R2) targeted the native configurations of each repeat region, while cross-combinations (F1/R2 and F2/R1) were designed to detect recombined conformations (Table 2).

As shown in Figure 4, agarose gel electrophoresis revealed distinct banding patterns for both native and recombinant configurations in multiple repeat pairs. For each repeat, at least two structural forms were detected: canonical conformations (F1R1 and F2R2) and recombined products (F1R2 and F2R1). The presence of clear bands in lanes corresponding to recombined primer pairs supports the hypothesis that repeat-mediated homologous recombination contributes to structural diversity in the *I. indigotica* mitogenome. Negative control lanes (NC) confirmed the specificity of the amplification reactions and excluded contamination. These findings highlight the dynamic nature of plant mitogenomes and the role of dispersed and palindromic repeats in generating structural variation.

Structural features and comparative analysis of cpgenomes

The cpgenome of *I. indigotica* was successfully assembled with a total length of 153,821 bp and a typical quadripartite structure, comprising a large single-copy (LSC) region (83,577 bp), a small single-copy (SSC) region (17,702 bp), and two identical inverted repeat (IR) regions (26,271 bp each) (Figure 2A). Genome annotation identified gene clusters associated with photosystem I and II, ATP synthase, ribosomal proteins, and tRNAs, highlighting the conserved functionality of the cpgenome in photosynthesis and plastid gene expression. Functional categories were color-coded to distinguish among coding regions such as RNA polymerase subunits, NADH dehydrogenase components, and ribosomal RNA genes.

Comparative genome alignment revealed high conservation between the newly assembled cpgenome of *I. indigotica* and a previously reported sequence from the same species (OP620952), differing by only 6 base pairs in total length (Figure 2B). In contrast, alignment with the cpgenome of *I. tinctoria* (NC_028415) revealed minor sequence variations, suggesting evolutionary divergence between these congeneric species (Figure 2C). Further structural comparison of the LSC, SSC, and IR boundary regions between

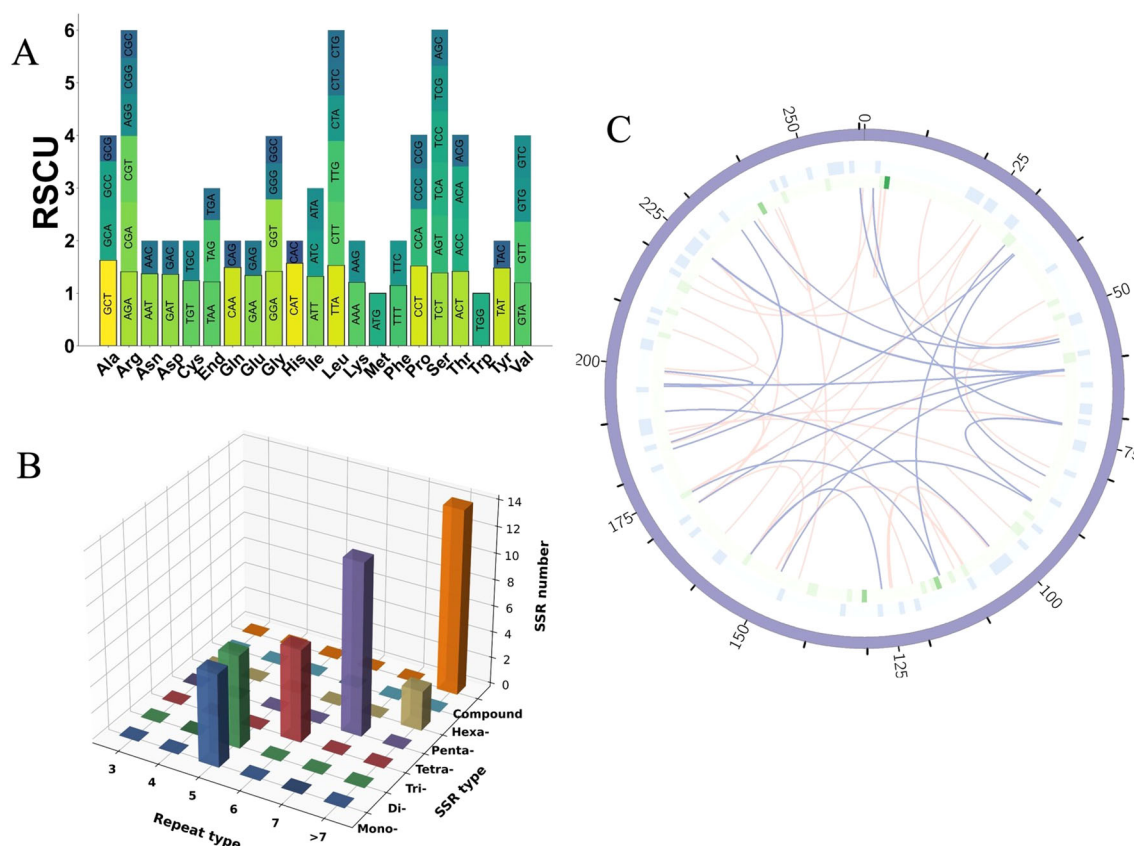


FIGURE 3

Codon usage, SSR analysis, and repeat element distribution in the mitogenome of *I. indigotica*. (A) RSCU values for PCGs in the mitogenome of *I. indigotica*. (B) Distribution of SSRs in the mitogenome of *I. indigotica*, categorized by repeat type and SSR type. (C) Distribution of repeat elements in the mitogenome of *I. indigotica*. The outermost circle represents SSRs, the middle circle shows tandem repeats, and the innermost circle illustrates dispersed repeats, with pink lines indicating forward repeats and purple lines indicating palindromic repeats.

I. indigotica and *I. tinctoria* uncovered shifts in the positions of junctions such as JLB, JSB, JSA, and JLA, along with differences in adjacent gene lengths and arrangements (Figure 2D). These structural differences at IR boundaries, including variable placement of genes such as *ycf1*, *rps19*, and *ndhF*, underscore the dynamic nature of cpgenome evolution within the *Isatis* genus.

Organellar and nuclear integration of mitochondrial sequences

Intracellular DNA transfer is a common feature of plant genomes, contributing to genomic complexity and evolutionary adaptation. In the mitogenome of *I. indigotica*, a total of 17 mitochondrial plastid DNA transfers (MTPTs) were identified, with sequence lengths ranging from 74 bp to 1,367 bp and sequence identity between 74.4% and 99.8% (Supplementary Table S6). These MTPTs collectively span several kilobases and reflect frequent plastid-to-mitochondrion DNA movement (Figure 5A). Among them, several MTPTs contained fragments of PCGs, such as *rbcl* (MTPT1), *psaB* (MTPT2, MTPT11), and *ycf1* (MTPT3, MTPT4, MTPT9, MTPT10), indicating historical

transfers of functional genetic elements. Notably, MTPT3 and MTPT4 represent bidirectional transfers of the same *ycf1* fragment, with nearly complete sequence identity (99.8%) and reverse orientation.

In addition to PCGs, various ribosomal RNA and tRNA genes were also identified among MTPTs. For example, MTPT5 and MTPT6 encode partial sequences of *rrn23*, while MTPT7 and MTPT8 harbor homologs of *rrn16*, though with lower identity (~74%) and higher mismatch counts, suggesting older or less conserved transfers. tRNA-related MTPTs include *trnP-UGG* (MTPT12), *trnW-CCA* (MTPT13), *trnD-GUC* (MTPT14), and *trnM-CAU* (MTPT15, MTPT16), as well as *trnI-CAU* (MTPT17), showing varying levels of conservation. These findings indicate that not only protein-coding but also structural RNA elements were subject to intracellular translocation. While some tRNAs may retain functional integrity, others, particularly those with high mismatch and gap counts, may be relics of ancient transfers with limited biological function.

Beyond organelle–organelle transfer, comparative analysis based on the *I. indigotica* reference nuclear genome (NCBI accession VHIU00000000) revealed extensive integration of mitochondrial DNA (NUMTs) into the nuclear genome.

TABLE 2 Primer sequences and PCR conditions for verification of homologous recombination mediated by repetitive sequences.

Repeats	Primers	Sequence 5'–3'	Product length of F1R1 and F2R2 (bp)	PCR procedure
Forward Repeat F96	F96-F1	TTCGCCTAGCTGCAAAAGGA	337	PCR Reaction Mix (10 μL): Add 1.0 μL of 10× buffer, 0.2 μL of dNTPs, 0.2 μL each of forward and reverse primers, about 0.5 μL of DNA template, and 0.1 μL of Taq polymerase. Add nuclease-free water to reach a total of 10 μL.
	F96-R1	CCAGAGGCAAGGCTATAGGC		
	F96-F2	TCGCCTAGCTGCACATGAAA	237	
	F96-R2	ACTGCCCTTCCCGTAGTAGT		
Forward Repeat F146	F146-F1	GGTTTCCCGATCGGAGATCC	398	
	F146-R1	GACTCTCCTACGGACCTGGT		
	F146-F2	CCCAGCGGAAAAGGCTAAGA	752	
	F146-R2	AGGGCCCCAAATGTCAATGT		
Palindromic Repeat P93	P93-F1	GGATCCTTCGCTGTTTGTGC	1013	
	P93-R1	AAGAGTACTTCGCGCCACAA		
	P93-F2	GCAGGGAGTGTGACAACGTA	1138	
	P93-R2	GATCCCGTCTCATTCCGACC		
Palindromic Repeat P135	P134-F1	CGCTAGATCCTGACCCCAAC	439	
	P134-R1	GCTCACTTCCCTCGTTTCGAA		
	P134-F2	TTCAAGCCATAACACGCCCT	1029	
	P134-R2	TTCCATTCTCTGTGAGCCAC		
Palindromic Repeat P326	P326-F1	CGCTAGATCCTGACCCCAAC	593	
	P326-R1	AGAGGTAGCTACAGTCGCGA		
	P326-F2	AAAGGCCGCTGTCTCTTCTC	1169	
	P326-R2	AGACCACTCGACCGATCTCA		

The designation F1&R1:57 indicates that the annealing temperature for the primer pair F1&R1 was set at 57 degrees Celsius.

As illustrated in [Figure 5B](#), mitochondrial fragments were mapped to all seven nuclear chromosomes (chr1–chr7), with chr1, chr2, and chr5 exhibiting particularly dense insertion patterns. These NUMTs contribute to nuclear genome expansion and serve as evolutionary footprints of organellar-nuclear interaction. Together, the presence of MTPTs and NUMTs highlights the dynamic and ongoing exchange of genetic material between organellar and nuclear compartments in *I. indigotica*, providing important insights into organellar genome evolution and intracellular genomic plasticity.

Consistent patterns of RNA editing in PCGs revealed by comparative prediction and transcriptome-based analyses

The prediction-based strategy identified 380 putative C-to-U RNA-editing sites across mitochondrial PCGs of *I.indigotica*, based on conserved editing patterns from related species using PREPACT3 ([Supplementary Table S7](#)). These predicted edits predominantly resulted in nonsynonymous changes, with a strong preference for specific amino acid substitutions. Notably, conversions from serine (S) to leucine (L), proline (P) to leucine (L), and serine (S) to phenylalanine (F) were the most frequent, suggesting potential functional shifts in

hydrophobicity and structural properties of mitochondrial proteins. Gene-wise analysis revealed that *nad4*, *ccmFn*, and *ccmB* harbored the greatest number of editing sites, indicating that genes involved in respiratory and cytochrome c biogenesis processes are frequent targets of RNA editing ([Supplementary Figure S2](#)). A detailed classification of RNA-editing events based on amino acid property transitions ([Supplementary Table S8](#)) further highlights these tendencies. Of the 357 identified nonsynonymous edits, nearly half (48.74%) were transitions from hydrophilic to hydrophobic residues, indicating a substantial shift toward increased protein hydrophobicity. This group was dominated by conversions such as TCA (S) $\rightarrow$ TTA (L), TCT (S) $\rightarrow$ TTT (F), and CGG (R) $\rightarrow$ TGG (W). Hydrophobic-to-hydrophobic transitions accounted for 28.29%, mainly due to P $\rightarrow$ L and A $\rightarrow$ V substitutions. Hydrophilic-to-hydrophilic and hydrophobic-to-hydrophilic edits comprised 14.00% and 8.40%, respectively, with changes like CGT (R) $\rightarrow$ TGT (C) and CCT (P) $\rightarrow$ TCT (S) being prominent. Additionally, two editing events (0.56%) led to stop codons, potentially serving regulatory or degradation-related roles. This distribution emphasizes the functional importance of RNA editing in refining the hydrophobicity and polarity of mitochondrial proteins, thereby influencing their structural and biochemical dynamics.

In parallel, transcriptome read mapping against the *I. indigotica* mitogenome (C_AA108663) identified 230 RNA-editing sites

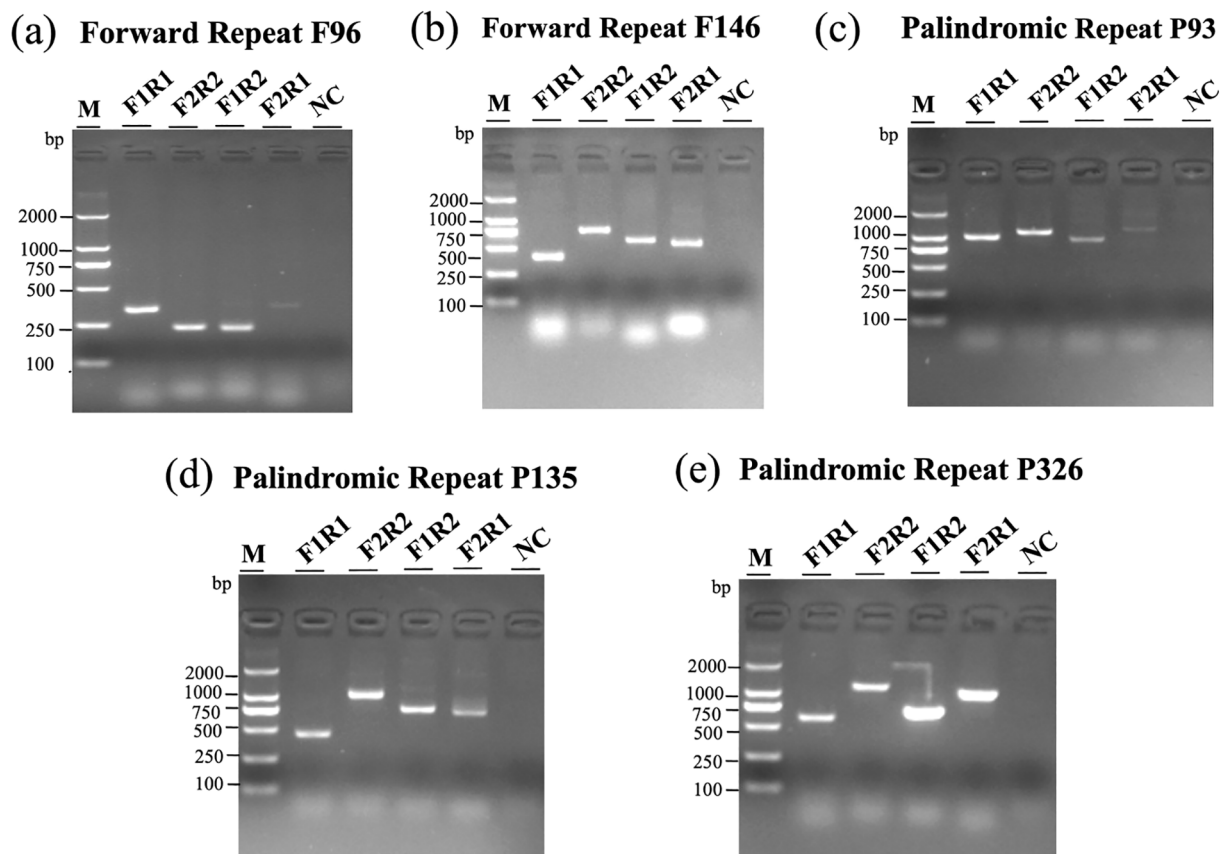


FIGURE 4

Structural variation of *I. indigotica* mtDNA mediated by repeat recombination. (a–e) PCR amplification results showing multiple conformations of mtDNA arising from recombination at two forward repeats (F96 and F146) and three palindromic repeats (P93, P134, and P326). Each gel includes: lane M, DNA marker; lanes F1R1 and F2R2, two major configurations; lanes F1R2 and F2R1, two minor configurations; NC, negative control. The observed band patterns reflect structural rearrangements driven by repeat-mediated recombination events.

(Figures 5C, D), derived from variant calling of publicly available RNA-seq data. Despite a lower total count, the editing spectrum overlapped substantially with prediction results. Key substitutions such as S→L, P→L, and S→F were consistently observed, and *nad4* remained the most edited gene. These findings indicate that high-confidence editing events are reproducible across methodologies, and that a conserved subset of edits is likely to be biologically relevant. The concordant amino acid change patterns support the hypothesis that RNA editing modulates protein properties in a directionally selective manner, favoring alterations in polarity and hydrophobicity to maintain or optimize protein function within the mitochondrial environment.

Mitogenome evolution and comparative genomics

Based on Figure 6, the evolutionary and structural characteristics of the *I. indigotica* mitogenome were investigated through phylogenetic analysis, genome feature comparison, and multi-species alignment. The phylogenetic tree (Figure 6A), constructed from conserved mitochondrial PCGs across 26 representative angiosperm species, demonstrates that *I. indigotica* clusters tightly with *I. tinctoria*,

confirming their close genetic relationship within the *Brassicaceae* family. Other genera, such as *Raphanus*, *Brassica*, and *Lepidium*, also form distinct, well-supported clades, reflecting the broader phylogenetic framework within the order Brassicales. The high bootstrap values across the tree nodes provide robust support for the inferred relationships and indicate strong conservation among mitochondrial genes across these taxa. This phylogenetic placement underscores the evolutionary consistency in mitogenomes within *Brassicaceae* and affirms the use of mitochondrial PCGs as reliable markers for resolving interspecies relationships. Furthermore, the comparison of mitogenome size and GC content among 18 *Brassicaceae* species (Figure 6B) reveals that although the total mitogenome size varies widely, ranging from approximately 220,000 to over 390,000 base pairs, the GC content remains relatively stable across species, typically clustering between 44% and 46%. *Isatis tinctoria*, marked in red, exhibits a genome size of approximately 283 kb with a GC content of ~45%, placing it near the average within its family. This trend suggests that while genome size can undergo lineage-specific expansion or reduction due to structural rearrangements or integration of foreign sequences, the GC content is more evolutionarily conserved and may be subject to functional constraints related to DNA stability or transcription efficiency.

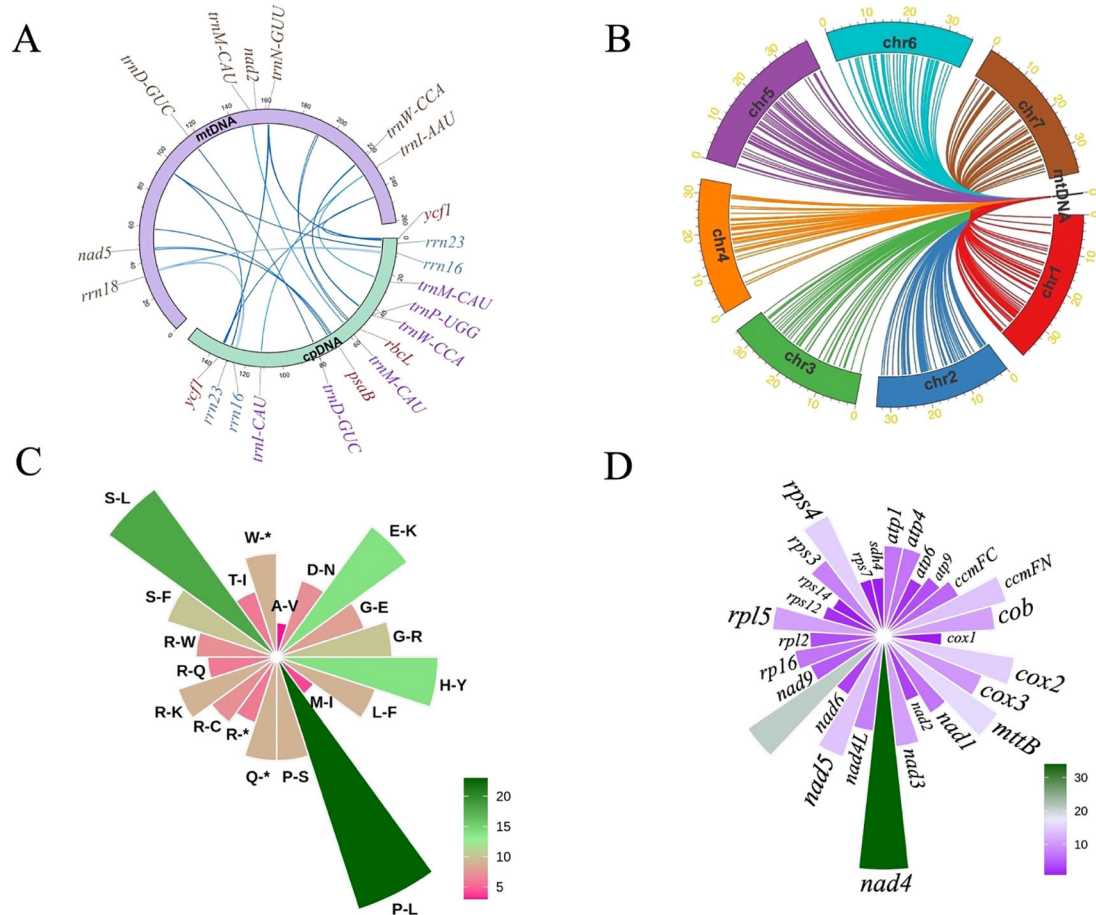


FIGURE 5

Analysis of sequence transfer, RNA editing, and gene interactions in the mitogenome of *I. indigotica*. (A) Sequence transfer between the mtDNA and cpDNA of *I. indigotica*, with transferred genes and regions indicated. Blue lines represent homologous fragments. (B) Integration of mtDNA fragments into the nuclear genome of *I. indigotica*, based on reference genome data from NCBI (accession number VHIU000000000). Colored lines link mitochondrial sequences to corresponding locations on nuclear chromosomes *chr1*–*chr7*. (C) Predicted RNA editing events in *I. indigotica* mitochondrial genes, represented as amino acid conversions. The color scale indicates the frequency of each type of editing event. (D) Distribution of RNA editing events across mitochondrial genes, with *nad4* showing the highest number of editing sites. Bar lengths indicate the editing frequency per gene.

In addition to phylogenetic and compositional features, the circular genome visualization (Figure 6C) offers a comprehensive view of the *I. indigotica* mitogenome and its structural architecture. The annotated internal rings display key genomic components, including coding sequences (CDS), transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), GC content, and GC skew, which collectively reflect the functional landscape of the mitogenome. Surrounding these are twelve concentric outer rings that depict BLAST-based alignments between *I. indigotica* and twelve other selected species within *Brassicaceae*, enabling a comparative analysis of genome conservation and divergence. The presence of numerous conserved syntenic blocks indicates that many mitochondrial regions—particularly those associated with essential genes such as *cox*, *nad*, *atp*, and *ccm*—are strongly preserved across species, consistent with their critical roles in mitochondrial respiration and metabolism. At the same time, breaks and shifts in alignment patterns highlight lineage-specific rearrangements and structural variation, which may result from recombination, gene loss, or horizontal gene transfer events. These structural differences, although

variable, often spare the core mitochondrial gene set, reinforcing the notion that while non-coding and intergenic regions are more evolutionarily plastic, the coding content remains stable to ensure essential mitochondrial functions. Taken together, this integrated analysis affirms that *I. indigotica* possesses a structurally stable mitogenome that shares extensive sequence similarity with its close relatives, while also harboring unique features that may reflect its individual evolutionary history. These findings provide valuable insight into mitogenome evolution within *Brassicaceae* and set a foundation for future investigations into mitochondrial gene function, inheritance, and adaptation.

Syntenic conservation and structural rearrangement in plant mitogenomes

The synteny and collinearity analyses of the mitogenome reveal notable patterns of both conservation and structural reconfiguration

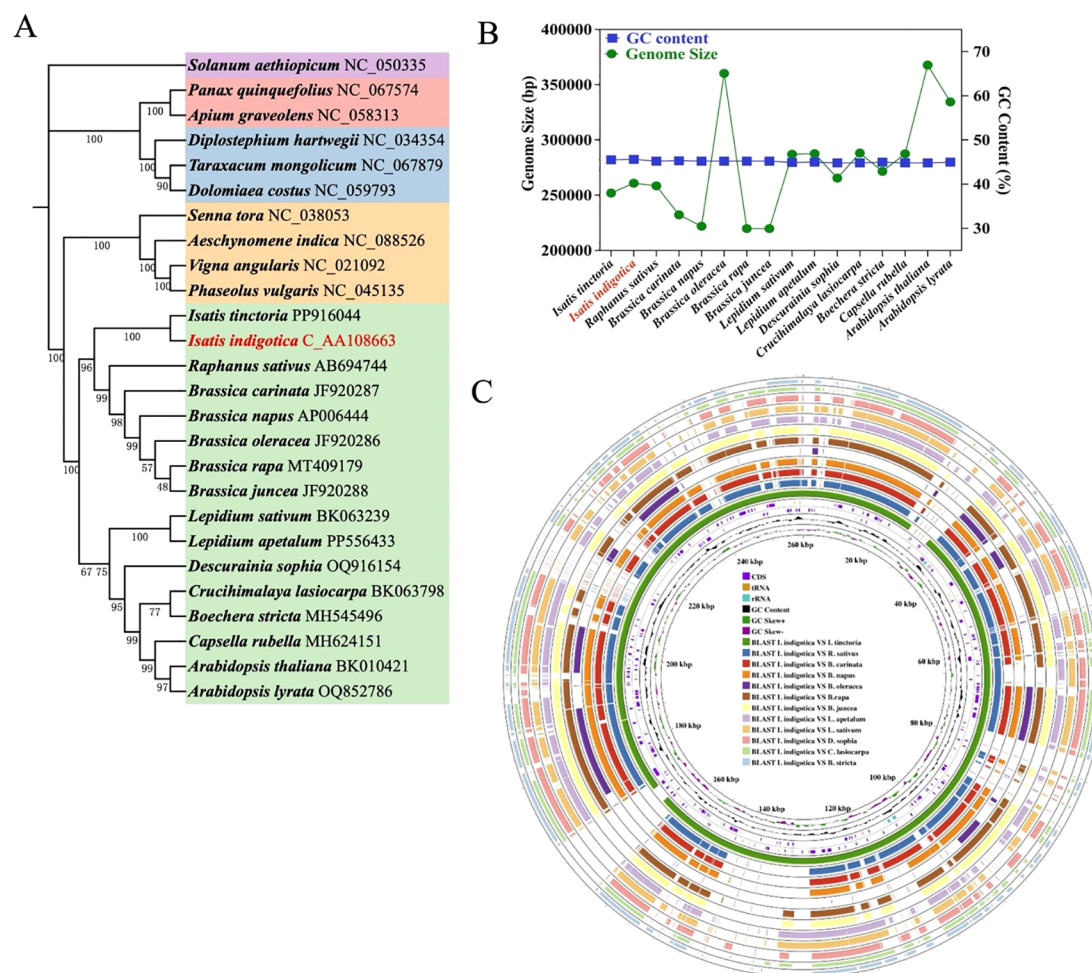


FIGURE 6

Mitogenome evolution and comparative analysis of *I. indigotica*. (A) Phylogenetic tree of 26 plant species based on conserved mitochondrial PCGs, highlighting the position of *I. indigotica* (in red) and its evolutionary relationships within the Brassicales and related taxa. (B) Comparison of mitogenome size and GC content across 18 Brassicaceae species, with *I. indigotica* marked in red. (C) Circular representation of the *I. indigotica* mitogenome, showing annotated features including CDS, tRNA, rRNA, GC content, and GC skew. The outer rings display BLAST comparisons between *I. indigotica* and 12 other species, revealing conserved regions and structural variations.

across different plant species. The syntenic comparisons, as illustrated in Figure 7A, show that extensive homologous sequence regions are retained between the focal genome and those of four other species: *I. tinctoria*, *A. graveolens*, *T. mongolicum*, and *V. angularis* (Supplementary Figure S9). These homologous segments are represented by connecting curves, varying in intensity and continuity, which suggest differing degrees of sequence similarity and genome structural stability. The greatest density and continuity of these curves are observed between the focal genome and *I. tinctoria*, indicating a high degree of conservation in sequence composition and segmental arrangement. By contrast, the connections with the other three species appear more sporadic and fragmented, suggesting a comparatively lower level of sequence conservation and a higher degree of rearrangement. This discrepancy likely reflects differing evolutionary distances, with more distantly related species exhibiting greater divergence in mitogenome architecture.

Further insights into structural organization are provided by the collinearity analysis in Figure 7B, which focuses on three species:

R. sativus, *I. tinctoria*, and the focal taxon. The visualization of LCBs, each demarcated by uniquely colored segments, facilitates a direct comparison of gene order, orientation, and overall genomic architecture. A pronounced similarity in block arrangement and orientation is evident between the focal genome and *I. tinctoria*, reinforcing the observation of strong genomic conservation inferred from the synteny analysis. In contrast, the alignment with *R. sativus* reveals a more disrupted block structure, characterized by numerous inversions, translocations, and possible insertions or deletions. These structural rearrangements result in a visually more complex and less orderly pattern of alignment, reflecting a greater divergence in genome structure. The differential collinearity among these three species suggests that mitogenome evolution in plants involves a balance between the retention of conserved functional elements and lineage-specific rearrangements. Such rearrangements may occur through recombination, duplication, or other mechanisms of genome plasticity, which contribute to the observed diversity in mitogenome organization across plant lineages.

Divergent evolutionary patterns of PCGs among angiosperms

To investigate the evolutionary dynamics of PCGs among angiosperms, three complementary analyses were conducted: gene copy number variation, Ka/Ks ratios, and Pi. Together, these results elucidate the extent of functional conservation, evolutionary constraint, and sequence variability across diverse lineages, with particular attention to the Brassicales and Fabales clades.

Gene copy number profiling (Figure 8A) revealed a heterogeneous landscape across the surveyed taxa. While the majority of PCGs were retained as single-copy genes in most species—particularly those involved in essential energy metabolism pathways such as ATP synthesis (e.g., *atp1*, *atp6*, *atp9*) and cytochrome complexes (e.g., *cob*, *cox1–3*)—a subset of genes exhibited lineage-specific duplications or absences. For instance, duplications in genes such as *rps12*, *nad4L*, and *ccmFN* were observed in several *Brassicaceae* members, suggesting possible subfunctionalization or compensatory mechanisms in mitochondrial translation and redox regulation. Meanwhile, frequent losses or underrepresentation of ribosomal protein genes (e.g., *rpl2*, *rps1*, *rps14*) in multiple lineages may indicate ongoing nuclear transfer or pseudogenization events. These patterns reflect both the functional rigidity and evolutionary plasticity of plant mitogenomes, where essential bioenergetic functions are maintained under strict constraint, while other genes tolerate structural fluidity.

The evolutionary constraints acting upon these genes were further evaluated through Ka/Ks analysis (Figure 8B), focusing on 16 representative *Brassicaceae* species. As expected for mitogenomes, purifying selection dominated across most PCGs, with Ka/Ks ratios predominantly below 1.0. However, exceptions were noted in genes such as *ccmB*, *matR*, and *rps3*, where elevated Ka/Ks values in specific taxa suggested relaxed selection or episodic positive selection. Notably, the variability observed in cytochrome *c* maturation genes (*ccmB*, *ccmC*, *ccmFN*) and ribosomal proteins (*rps3*, *rps4*) implies that these loci may be subject to adaptive pressures, potentially in response to environmental cues or cytonuclear co-evolution. Such divergence highlights regions of the mitogenome that may play a disproportionate role in lineage-specific adaptation or reproductive isolation.

Complementary to these findings, Pi analysis (Figure 8C) offered a broader population-level perspective on genetic variability. A clear disparity emerged between Brassicales and Fabales, with the latter exhibiting consistently higher Pi values across nearly all mitochondrial genes. Genes such as *rps12*, *nad6*, and *ccmFC* in Fabales showed particularly elevated polymorphism, indicating more rapid sequence turnover or relaxed constraints within this order. In contrast, Brassicales maintained low Pi, underscoring the genomic stability and strong conservation of mitogenomic content in this lineage. This contrast between lineages not only mirrors their differing evolutionary histories but also supports the notion that mitogenomes, though structurally conserved, can exhibit considerable heterogeneity in their molecular evolution.

Discussion

The complete mitochondrial and cpgenome assemblies presented in this study provide valuable genetic resources for investigating the organellar biology and evolutionary history of *I. indigotica*. By utilizing high-fidelity long-read sequencing and implementing integrative annotation strategies, this research addresses the complexity of organelle genome structure, particularly the challenges associated with extensive repeat regions and recombination events characteristic of plant mitogenomes. The circular-mapping mitogenome of *I. indigotica* spans 260,864 bp and contains 65 unique genes, including all 23 core mitochondrial PCGs, consistent with gene content observed in other *Brassicaceae* species.

A notable feature of the mitogenome is the abundance and diversity of repeat elements, including SSRs, tandem repeats, and dispersed repeats. These repetitive elements, especially the medium-to-large dispersed repeats, were shown to mediate recombination events, as evidenced by the presence of multiple structural isoforms validated through both in silico prediction and experimental PCR confirmation. This dynamic recombination landscape suggests that *I. indigotica* maintains a multipartite mitogenome architecture, in line with other angiosperms exhibiting structural plasticity.

From a taxonomic perspective, the distinction between *I. indigotica* and *I. tinctoria* has been a subject of ongoing debate. Although both species belong to the same genus and exhibit considerable morphological and genetic similarity, their practical applications and ethnobotanical roles diverge significantly. *Isatis indigotica* is predominantly used in traditional Chinese medicine for its antiviral and anti-inflammatory properties, whereas *I. tinctoria* has historically been cultivated as a source of indigo dye in Europe. This study offers the first comprehensive organellar genomic comparison between the two species based on cell line-derived mitochondrial sequences. The observed synteny and local rearrangements provide molecular evidence of their close genetic relationship, while also revealing distinct structural variations that may be linked to lineage-specific functional adaptation or domestication history.

RNA editing plays a critical role in the post-transcriptional modification of mitochondrial transcripts. A total of 587 RNA editing sites were predicted within PCGs, primarily involving C-to-U conversions. Most edits result in nonsynonymous changes, favoring the substitution of hydrophilic residues with hydrophobic ones, which may affect protein stability or membrane integration. In this study, two complementary strategies were employed to assess RNA editing patterns: computational prediction and transcriptome-based empirical validation. While prediction tools such as PREPACT revealed the full landscape of putative editing events, mapping of RNA-seq reads allowed for the confirmation of 380 high-confidence sites. The convergence between predicted and transcript-supported editing patterns—particularly in genes such as *nad4*, *ccmFn*, and *ccmB*—provides strong evidence that these modifications are functionally relevant. The dual-method approach enhances the robustness of RNA editing detection and helps to reconcile discrepancies often encountered in single-method studies.

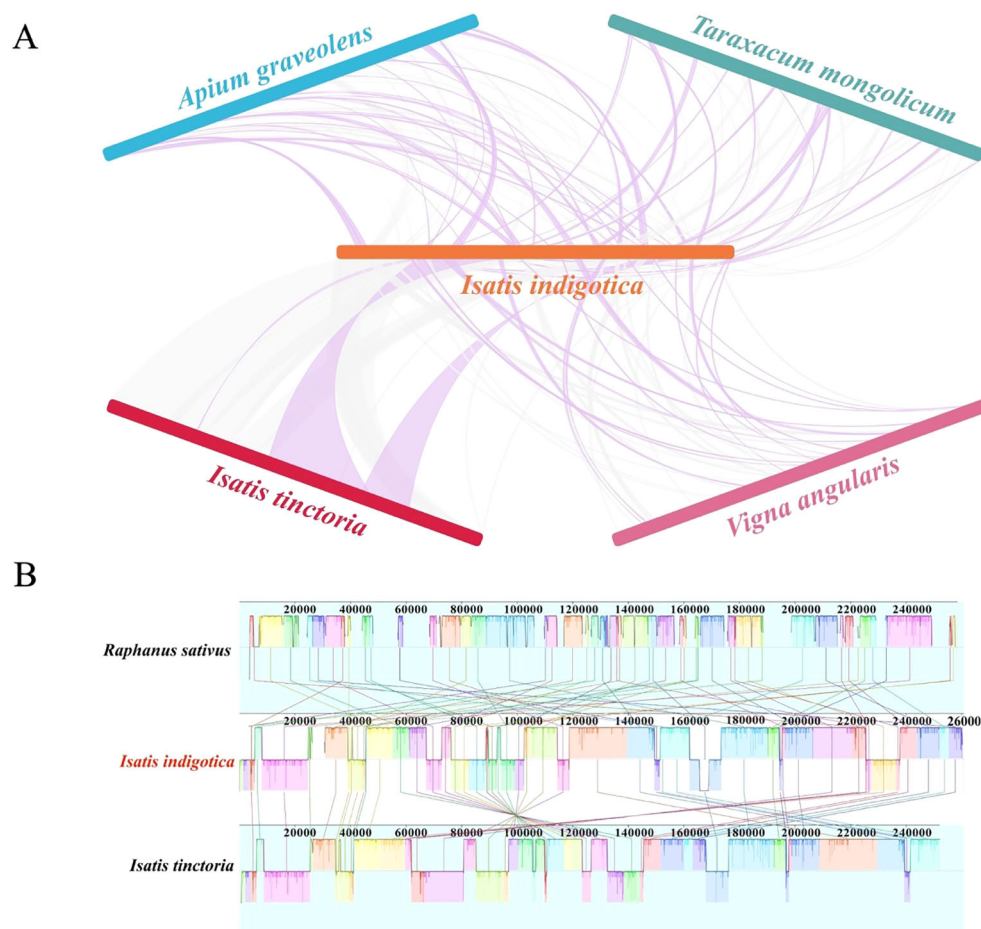


FIGURE 7

Synteny and collinearity analysis of the *I. indigotica* mitogenome. (A) Synteny relationships between *I. indigotica* and four other species (*I. tinctoria*, *A. graveolens*, *T. mongolicum*, and *V. angularis*), showing shared homologous regions and genome rearrangements. (B) Collinearity comparison among the mitogenomes of *I. indigotica*, *I. tinctoria*, and *R. sativus*, revealing conserved blocks and structural variations across the genomes.

Comparative analysis of mitogenomes further supports the close phylogenetic relationship between *I. indigotica* and *Raphanus sativus*, as initially suggested by chloroplast phylogenies (Yang and Wang, 2017). Mitogenomic comparisons reveal notable synteny and structural correspondence between these two species, characterized by conserved gene content and locally aligned genomic blocks. However, despite this broad conservation, lineage-specific rearrangements—including inversions and translocations—are evident, highlighting the dynamic nature of mitogenome evolution. Whole-genome dot plot analyses and synteny mapping underscore this pattern: while syntenic regions are often disrupted across distantly related taxa, a more continuous alignment is observed between *I. indigotica* and other *Brassicaceae* members, particularly *R. sativus*. These structural similarities lend further support to their sister relationship and suggest that mitogenomic architecture, in addition to gene content, can serve as an informative phylogenetic signal within the family.

Phylogenetic reconstruction based on 28 conserved mitochondrial PCGs placed *I. indigotica* firmly within the *Brassicaceae* clade, closely aligned with other *Cruciferae* species. Genome size and GC content across examined mitogenomes show moderate variation, with *I. indigotica* exhibiting values well within the expected range for its

taxonomic group. The presence of MTPTs and mitochondrial nuclear DNA insertions highlights ongoing intracellular DNA transfer, suggesting a dynamic genomic interplay between organelles and the nucleus. The integration of plastid and mitochondrial sequences, particularly involving tRNA and rRNA gene fragments, may contribute to genome expansion and functional redundancy.

In summary, the structural, functional, and comparative genomic analyses of the *I. indigotica* mitogenome provide significant insights into plant mitogenome evolution, with emphasis on repeat-mediated recombination, RNA editing, and inter-organellar DNA transfer. The dual approach to RNA editing analysis and the focused comparison with *I. tinctoria* also address existing gaps in taxonomic clarification and transcriptomic post-processing in *Brassicaceae*. These findings offer a robust foundation for further evolutionary, physiological, and pharmacological studies in this important medicinal lineage.

Conclusion

This study provides new insights into the mitochondrial genome of *I. indigotica*, shedding light on its structure, function,

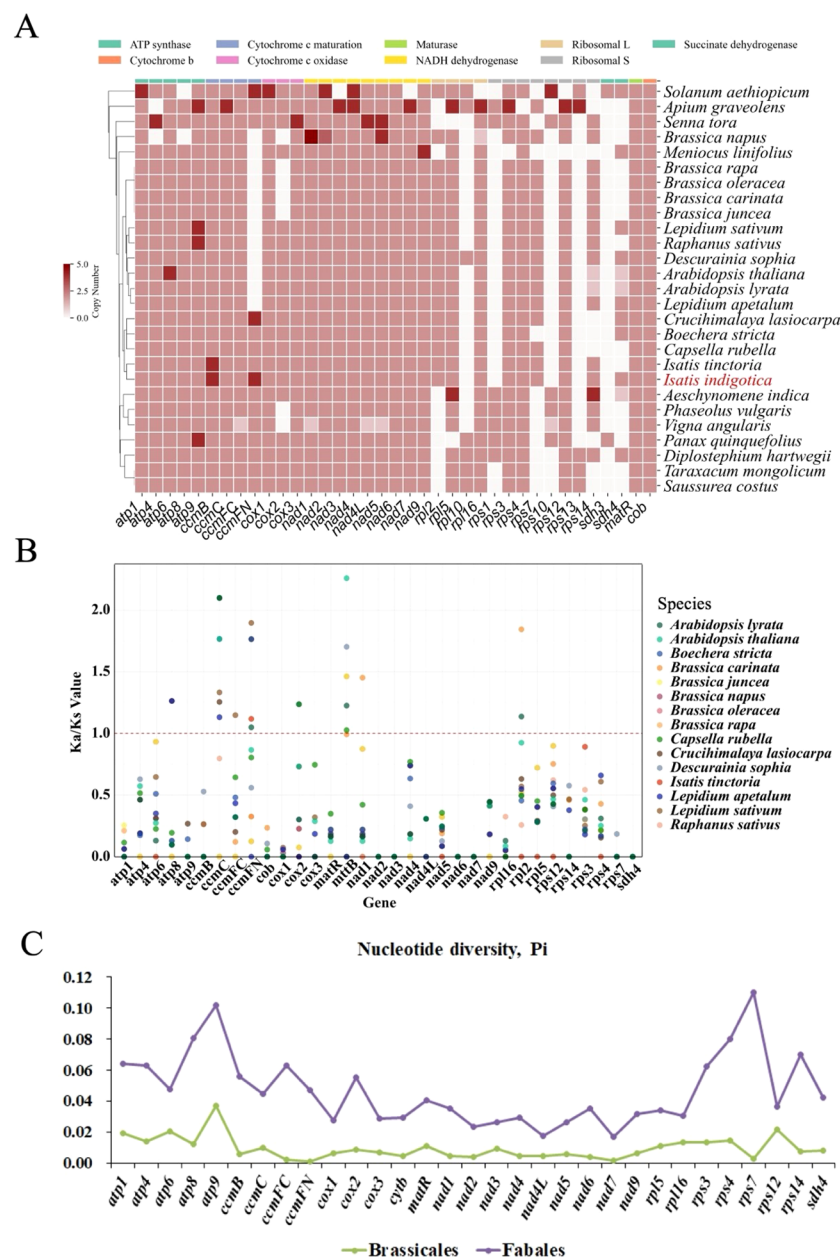


FIGURE 8

Comparative genomic and evolutionary analysis of mitochondrial genes in *I. indigotica* and related species. (A) Heatmap of mitochondrial PCGs copy numbers across 30 plant species, with functional categories indicated by color. The distribution reveals variation in gene retention and duplication among species, including *I. indigotica* (red). (B) Pairwise Ka/Ks ratios for shared mitochondrial genes among 17 Brassicaceae species, indicating patterns of selective pressure acting on individual genes. (C) π of mitochondrial genes between Brassicales and Fabales species, showing genetic variability within each group.

and evolutionary characteristics. The complete mitogenome, consisting of 260,864 base pairs, includes 31 protein-coding genes, 21 transfer RNAs, and 3 ribosomal RNAs, along with notable repeat-mediated recombination events. These features suggest active genome rearrangements and stability essential for mitochondrial function. The study also emphasizes the importance of RNA editing, particularly C-to-U conversions, in stabilizing and optimizing mitochondrial proteins critical for respiration and cytochrome c biogenesis. Additionally, phylogenetic analysis with

other Brassicaceae species revealed a close relationship between *I. indigotica* and Brassica species, highlighting the ongoing genetic exchange between mitochondria and the nucleus, particularly through mitochondrial-nuclear DNA integrations. Overall, the findings contribute significantly to our understanding of the dynamic interaction between nuclear and organellar genomes, offering valuable perspectives for future research on plant mitochondrial evolution and its relevance to medicinal plant studies.

Data availability statement

The sequencing data produced in this project have been archived in the Genome Sequence Archive (GSA) at the National Genomics Data Center (NGDC), accessible under accession number CRA026873 (<https://ngdc.cnbc.ac.cn/gsa>) (Members and Partners, 2025). In addition, corresponding annotated genome information has been submitted to GenBase at NGDC/CNCB and is available under accession number C_AA108663.1 (<https://ngdc.cnbc.ac.cn/genbase>) (Bu et al., 2024).

Ethics statement

This study was conducted in full compliance with institutional, national, and international ethical standards. Plant materials were sourced from publicly available resources, including authorized collections and databases, ensuring adherence to relevant legal and regulatory frameworks. No specific permits were required for the collection and analysis of *Eleutherococcus senticosus* samples. All experimental procedures were approved by the ethical committee of Jiangsu University, ensuring rigorous oversight of the research methodology. The study did not involve endangered or protected species, and all sampling followed sustainable and ethical practices.

Author contributions

SY: Data curation, Formal Analysis, Writing – original draft. MQ: Data curation, Investigation, Writing – review & editing. EF: Validation, Writing – review & editing. XG: Funding acquisition, Resources, Writing – review & editing. MT: Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This research was supported through funding from several prestigious sources,

including the City Social Development Project of Zhenjiang (SH2024080) and (SH2024038), the Zhenjiang Science and Technology Plan Projects (Project No. SH2023078), and the Medical Education Collaborative Innovation Fund of Jiangsu University (Grant No. JDYY2023009). These contributions were essential for the successful completion of this work.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2025.1655810/full#supplementary-material>

References

- Ala, K. G., Zhao, Z., Ni, L., and Wang, Z. (2023). Comparative analysis of mitochondrial genomes of two alpine medicinal plants of *Gentiana* (*Gentianaceae*). *PLoS One* 18, e0281134. doi: 10.1371/journal.pone.0281134
- Arimura, S. I., and Nakazato, I. (2024). Genome editing of plant mitochondrial and chloroplast genomes. *Plant Cell Physiol.* 65, 477–483. doi: 10.1093/pcp/pcad162
- Benson, G. (2023). Tandem repeats finder- a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. doi: 10.1093/nar/27.2.573
- Bu, C., Zheng, X., Zhao, X., Xu, T., Bai, X., Jia, Y., et al. (2024). GenBase: A nucleotide sequence database. *Genomics Proteomics Bioinf.* 22, qzae047. doi: 10.1093/gpbjnl/qzae047
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinf.* 10, 421. doi: 10.1186/1471-2105-10-421
- Chang, S., Chen, J., Wang, Y., Gu, B., He, J., Chu, P., et al. (2013). The mitochondrial genome of *Raphanus sativus* and gene evolution of *cruciferous* mitochondrial types. *J. Genet. Genomics* 40, 117–126. doi: 10.1016/j.jgg.2013.01.003
- Charif, D., Thioulouse, J., Lobry, J. R., and Perriere, G. (2005). Online synonymous codon usage analyses with the ade4 and seqinR packages. *Bioinformatics* 21, 545–547. doi: 10.1093/bioinformatics/bti037
- Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009
- Chen, D., Wang, C., Liu, Y., Shen, W., Cuimu, Q., Zhang, D., et al. (2025). Systematic identification of R2R3-MYB S6 subfamily genes in *Brassicaceae* and its role in anthocyanin biosynthesis in *Brassica* crops. *BMC Plant Biol.* 25, 290. doi: 10.1186/s12870-025-06296-1

- Chen, Y., Ye, W., Zhang, Y., and Xu, Y. (2015). High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic Acids Res.* 43, 7762–7768. doi: 10.1093/nar/gkv784
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, 1884–1890. doi: 10.1093/bioinformatics/bty560
- Chen, S., Zhu, Z., Ma, H., Yang, J., and Guo, Q. (2014). DNA barcodes for discriminating the medicinal plant *Isatis indigotica* Fort. (*Cruciferae*) and its adulterants. *Biochem. Syst. Ecol.* 57, 287–292. doi: 10.1016/j.bse.2014.08.007
- Chevigny, N., Schatz-Daas, D., Lotfi, F., and Gualberto, J. M. (2020). DNA repair and the stability of the plant mitochondrial genome. *Int. J. Mol. Sci.* 21, 328. doi: 10.3390/ijms21010328
- Danecek, P., and McCarthy, S. A. (2017). BCFtools/csq: haplotype-aware variant consequences. *Bioinformatics* 33, 2037–2039. doi: 10.1093/bioinformatics/btx100
- Ding, C. H., and Zhu, H. (2020). *Isatisidis* Folium alleviates acetaminophen-induced liver injury in mice by enhancing the endogenous antioxidant system. *Environ. Toxicol.* 35, 1251–1259. doi: 10.1002/tox.22990
- Forner, J. (2025). Genome modification in plant mitochondria. *Plant Physiol.* 198, kiaf197. doi: 10.1093/plphys/kiaf197
- Gao, Y., Liu, K., Li, E., Wang, Y., Xu, C., Zhao, L., et al. (2022). Dynamic evolution of the plastome in the Elm family (*Ulmaceae*). *Planta* 257, 14. doi: 10.1007/s00425-022-04045-4
- Genovese, G., Rockweiler, N. B., Gorman, B. R., Bigdeli, T. B., Pato, M. T., Pato, C. N., et al. (2024). BCFtools/liftover: an accurate and comprehensive tool to convert genetic variants across genome assemblies. *Bioinformatics* 40, btac038. doi: 10.1093/bioinformatics/btac038
- Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47, W59–W64. doi: 10.1093/nar/gkz238
- Gualberto, J. M., Mileschina, D., Wallet, C., Niazi, A. K., Weber-Lotfi, F., and Dietrich, A. (2014). The plant mitochondrial genome: dynamics and maintenance. *Biochimie* 100, 107–120. doi: 10.1016/j.biochi.2013.09.016
- Gualberto, J. M., and Newton, K. J. (2017). Plant mitochondrial genomes: dynamics and mechanisms of mutation. *Annu. Rev. Plant Biol.* 68, 225–252. doi: 10.1146/annurev-arplant-043015-112232
- Han, F., Qu, Y., Chen, Y., Xu, L., and Bi, C. (2022). Assembly and comparative analysis of the complete mitochondrial genome of *Salix wilsonii* using PacBio HiFi sequencing. *Front. Plant Sci.* 13, 1031769. doi: 10.3389/fpls.2022.1031769
- Hanson, G., and Collier, J. (2018). Codon optimality, bias and usage in translation and mRNA decay. *Nat. Rev. Mol. Cell Biol.* 19, 20–30. doi: 10.1038/nrm.2017.91
- He, W., Yang, J., Jing, Y., Xu, L., Yu, K., and Fang, X. (2023). NGenomeSyn: an easy-to-use and flexible tool for publication-ready visualization of syntenic relationships across multiple genomes. *Bioinformatics* 39, btad121. doi: 10.1093/bioinformatics/btad121
- Jiang, X., Hu, Q., Mei, D., Li, X., Xiang, L., Al-Shehbaz, I. A., et al. (2025). Chromosome fusions shaped karyotype evolution and evolutionary relationships in the model family *Brassicaceae*. *Nat. Commun.* 16, 4631. doi: 10.1038/s41467-025-59640-2
- Kang, M., Wu, H., Yang, Q., Huang, L., Hu, Q., Ma, T., et al. (2020). A chromosome-scale genome assembly of *Isatis indigotica*, an important medicinal plant used in traditional Chinese medicine: An *Isatis* genome. *Hortic. Res.* 7, 18. doi: 10.1038/s41438-020-0240-5
- Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317
- Kong, W. J., Zhao, Y. L., Shan, L. M., Xiao, X. H., and Guo, W. Y. (2008). Investigation on the spectrum-effect relationships of EtOAc extract from *Radix Isatidis* based on HPLC fingerprints and microcalorimetry. *J. Chromatogr. B. Analyt. Technol. BioMed. Life Sci.* 871, 109–114. doi: 10.1016/j.jchromb.2008.06.053
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascogne, R., Horsman, D., et al. (2023). CircoS: An information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2023). REPuter- the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 29, 4633–4642. doi: 10.1093/nar/29.22.4633
- Lee, E., Helt, G. A., Reese, J. T., Munoz-Torres, M. C., Childers, C. P., Buels, R. M., et al. (2013). Web Apollo- a web-based genomic annotation editing platform. *Genome Biol.* 14, R93. doi: 10.1186/gb-2013-14-8-r93
- Lenz, H., and Knoop, V. (2013). PREPACT 2.0: predicting C-to-U and U-to-C RNA editing in organelle genome sequences with multiple references and curated RNA editing annotation. *Bioinform. Biol. Insights* 7, 1–19. doi: 10.4137/BBI.S11059
- Lenz, H., Rudinger, M., Volkmar, U., Fischer, S., Herres, S., Grewe, F., et al. (2010). Introducing the plant RNA editing prediction and analysis computer tool PREPACT and an update on RNA editing site nomenclature. *Curr. Genet.* 56, 189–201. doi: 10.1007/s00294-009-0283-5
- Leticia, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, X., Li, M., Li, W., Zhou, J., Han, Q., Lu, W., et al. (2023). Comparative Analysis of the Complete Mitochondrial Genomes of *Apium graveolens* and *Apium leptophyllum* Provide Insights into Evolution and Phylogeny Relationships. *Int. J. Mol. Sci.* 24, 14615. doi: 10.3390/ijms241914615
- Li, J., Ni, Y., Lu, Q., Chen, H., and Liu, C. (2025). PMGA: A plant mitochondrial genome annotator. *Plant Commun.* 6, 101191. doi: 10.1016/j.xplc.2024.101191
- Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187
- Liu, S., Wang, Z., Wang, H., Su, Y., and Wang, T. (2020). Patterns and Rates of Plastid rps12 Gene Evolution Inferred in a Phylogenetic Context using Plastomic Data of Ferns. *Sci. Rep.* 10, 9394. doi: 10.1038/s41598-020-66219-y
- Liu, J., Yu, S., Lu, P., Gong, X., Sun, M., and Tang, M. (2025). De novo assembly and characterization of the complete mitochondrial genome of *Phellodendron amurense* reveals three repeat-mediated recombination. *Gene* 935, 149031. doi: 10.1016/j.gene.2024.149031
- Lu, L. M., Mao, L. F., Yang, T., Ye, J. F., Liu, B., Li, H. L., et al. (2018). Evolutionary history of the angiosperm flora of China. *Nature* 554, 234–238. doi: 10.1038/nature25485
- Lukes, J., Tesarova, M., Yurchenko, V., and Votpyka, J. (2021). Characterization of a new cosmopolitan genus of trypanosomatid parasites, *Obscuromonas* gen. nov. (*Blastocrithidiinae* subfam. nov.). *Eur. J. Protistol.* 79, 125778. doi: 10.1016/j.jeop.2021.125778
- Members, C.-N., and Partners, (2025). Database resources of the national genomics data center, China national center for bioinformation in 2025. *Nucleic Acids Res.* 53, D30–D44. doi: 10.1093/nar/gkac978
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015
- Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., and Durbin, R. (2016). BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* 32, 1749–1751. doi: 10.1093/bioinformatics/btw044
- Presnyak, V., Alhusaini, N., Chen, Y. H., Martin, S., Morris, N., Kline, N., et al. (2015). Codon optimality is a major determinant of mRNA stability. *Cell* 160, 1111–1124. doi: 10.1016/j.cell.2015.02.029
- Ren, H., Xu, D., Xiao, W., Zhou, X., Li, G., Zou, J., et al. (2023). Chromosome-level genome assembly and annotation of *Zicaitai* (*Brassica rapa* var. *purpuraria*). *Sci. Data* 10, 759. doi: 10.1038/s41597-023-02668-0
- Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248
- Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X., and Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19, 2496–2497. doi: 10.1093/bioinformatics/btg359
- Rozewicki, J., Li, S., Amada, K. M., Standley, D. M., and Katoh, K. (2019). MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res.* 47, W5–W10. doi: 10.1093/nar/gkz342
- Sang, S., Cheng, H., Mei, D., Fu, L., Wang, H., Liu, J., et al. (2020). Complete organelle genomes of *Sinapis arvensis* and their evolutionary implications. *Crop J.* 8, 505–514. doi: 10.1016/j.cj.2019.12.001
- Sen, A., Kargar, K., Akgun, E., and Pinar, M. C. (2020). Codon optimization: a mathematical programming approach. *Bioinformatics* 36, 4012–4020. doi: 10.1093/bioinformatics/btaa248
- Shen, Y., Wu, S., Song, M., Zhang, H., Zhao, H., Wu, L., et al. (2024). The isolation, structural characterization and anti-inflammatory potentials of neutral polysaccharides from the roots of *isatis indigotica* fort. *Molecules* 29, 2683. doi: 10.3390/molecules29112683
- Speranza, J., Miceli, N., Taviano, M. F., Ragusa, S., Kwiecien, I., Szopa, A., et al. (2020). *Isatis tinctoria* L. (Woad): A Review of its Botany, Ethnobotanical Uses, Phytochemistry, Biological Activities, and Biotechnological Studies. *Plants (Basel)* 9, 298. doi: 10.3390/plants9030298
- Sun, Y. K., Gutmann, B., Yap, A., Kindgren, P., and Small, I. (2018). Editing of chloroplast rps14 by PPR editing factor EMB2261 is essential for arabidopsis development. *Front. Plant Sci.* 9, 841. doi: 10.3389/fpls.2018.00841
- Sun, H.-J., Wu, Y.-X., Wu, Z.-F., Han, F., Yang, M., and Wang, Y.-Q. (2021). Optimization of vacuum assisted heat reflux extraction process of *radix isatidis* using least squares-support vector machine algorithm. *Phytochem. Lett.* 43, 108–113. doi: 10.1016/j.phytol.2021.03.009
- Tan, Y., Yang, J., Sun, S., Zhu, H., Jiang, Y., Wei, X., et al. (2025). Functional analysis of tandem glycosyltransferases catalyzing the O-glycosylation of flavonoid 4'-O-glycosyl-6-C-glycosides biosynthesis from *Isatis indigotica*. *Plant Physiol. Biochem.* 224, 109920. doi: 10.1016/j.plaphy.2025.109920
- Thiel, T., Michalek, W., Varshney, R. K., and Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 106, 411–422. doi: 10.1007/s00122-002-1031-0

- Voisin, C., Choler, P., Vacher, C., Perrier, C., Renaud, J., Rioux, D., et al. (2025). Deciphering the drivers of speciation in *Alpine Noccaea* (Brassicaceae) reveal an evolution predominantly shaped by geography. *Mol. Phylogenet. Evol.* 211, 108381. doi: 10.1016/j.ympev.2025.108381
- Wang, R., Deng, Z., and Luo, Y. (2022). The complete chloroplast genome and phylogenetic analysis of *Cardamine circaeoides* Hook. f. et Thoms. 1861 (Brassicaceae). *Mitochondrial. DNA B. Resour.* 7, 1964–1967. doi: 10.1080/23802359.2022.2141081
- Wang, L., Liu, X., Xu, Y., Zhang, Z., Wei, Y., Hu, Y., et al. (2024). Assembly and comparative analysis of the first complete mitochondrial genome of a traditional Chinese medicine *Angelica biserrata* (Shan et Yuan) Yuan et Shan. *Int. J. Biol. Macromol.* 257, 128571. doi: 10.1016/j.ijbiomac.2023.128571
- Wang, Z., Wang, H., Gong, X., Ou, X., Guo, Y., and Tang, M. (2025). *De Novo* assembly and phylogenetic analysis of the complete mitochondrial genome of *Eleutherococcus senticosus* and related *araliaceous* species. *BMC Plant Biol.* 25, 760. doi: 10.1186/s12870-025-06809-y
- Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinf.* 8, 77–80. doi: 10.1016/S1672-0229(10)60008-3
- Wick, R. R., Schultz, M. B., Zobel, J., and Holt, K. E. (2015). Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics* 31, 3350–3352. doi: 10.1093/bioinformatics/btv383
- Wingett, S. W., and Andrews, S. (2018). FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res* 7, 1338.
- Xiang, C. Y., Gao, F., Jakovlic, I., Lei, H. P., Hu, Y., Zhang, H., et al. (2023). Using PhyloSuite for molecular phylogeny and tree-based analyses. *Imeta* 2, e87. doi: 10.1002/imt2.87
- Xiang, K. L., Mao, W., Peng, H. W., Erst, A. S., Yang, Y. X., He, W. C., et al. (2022). Organization, phylogenetic marker exploitation, and gene evolution in the plastome of *thalictrum* (Ranunculaceae). *Front. Plant Sci.* 13, 897843. doi: 10.3389/fpls.2022.897843
- Yang, S., and Wang, Z. (2017). The complete chloroplast genome sequence of the medicinal and economic plant woad *Isatis indigotica* (Brassicaceae). *Mitochondrial. DNA B. Resour.* 2, 514–515. doi: 10.1080/23802359.2017.1361356
- Zhang, Z. (2022). KaKs_Calculator 3.0: calculating selective pressure on coding and non-coding sequences. *Genomics Proteomics Bioinf.* 20, 536–540. doi: 10.1016/j.gpb.2021.12.002
- Zhang, X., Chen, H., Ni, Y., Wu, B., Li, J., Burzynski, A., et al. (2024). 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.* 24, e13952. doi: 10.1111/1755-0998.13952
- Zhang, D., Gao, F., Jakovlic, I., Zou, H., Zhang, J., Li, W. X., et al. (2020). PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.* 20, 348–355. doi: 10.1111/1755-0998.13096
- Zhang, Z., Li, J., Zhao, X. Q., Wang, J., Wong, G. K., and Yu, J. (2006). KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinf.* 4, 259–263. doi: 10.1016/S1672-0229(07)60007-2
- Zhang, X., Shan, Y., Li, J., Qin, Q., Yu, J., and Deng, H. (2023). Assembly of the complete mitochondrial genome of *pereskia aculeata* revealed that two pairs of repetitive elements mediated the recombination of the genome. *Int. J. Mol. Sci.* 24, 8366. doi: 10.3390/ijms24098366
- Zhang, Z., Tao, M., Shan, X., Pan, Y., Sun, C., Song, L., et al. (2022). Characterization of the complete chloroplast genome of *Brassica oleracea* var. *italica* and phylogenetic relationships in *Brassicaceae*. *PloS One* 17, e0263310.
- Zhao, J., Liu, S., Ren, H., Afriyie, O. E., Zhang, M., Xu, D., et al. (2025). Genome-wide identification and comparative evolution of 14-3-3 gene family members in five *Brassicaceae* species. *BMC Genomics* 26, 309. doi: 10.1186/s12864-025-11513-0
- Zhou, G., Qin, M., Liu, X., Qi, Y., Ou, X., and Tang, M. (2025). *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* 26, 13. doi: 10.1186/s12864-024-11190-5
- Zhu, B., Qian, F., Hou, Y., Yang, W., Cai, M., and Wu, X. (2021). Complete chloroplast genome features and phylogenetic analysis of *Eruca sativa* (Brassicaceae). *PloS One* 16, e0248556. doi: 10.1371/journal.pone.0248556
- Zwonitzer, K. D., Tressel, L. G., Wu, Z., Kan, S., Broz, A. K., Mower, J. P., et al. (2024). Genome copy number predicts extreme evolutionary rate variation in plant mitochondrial DNA. *Proc. Natl. Acad. Sci. U.S.A.* 121, e2317240121. doi: 10.1073/pnas.2317240121

Glossary

MTPTs	mitochondrial plastid DNAs	cob	cytochrome b gene
NUMTs	nuclear mitochondrial DNAs	matR	maturase-related gene
cpDNA	chloroplast DNA	LSC	large single-copy
PCGs	protein-coding genes	SSC	small single-copy
SSRs	simple sequence repeats	IR	inverted repeat
tRNAs	transfer RNAs	N50	minimum contig length covering 50% of the genome
rRNAs	ribosomal RNAs	mt	mitochondrial contig
RNA	ribonucleic acid	APG	Angiosperm Phylogeny Group
RSCU	relative synonymous codon usage	BLAST	Basic Local Alignment Search Tool
Ka/Ks	nonsynonymous/synonymous substitution ratio	PCR	polymerase chain reaction
Pi	nucleotide diversity	CTAB	cetyltrimethylammonium bromide
MSA	multiple sequence alignment	DEPC	diethyl pyrocarbonate
CDS	coding sequence	Qubit	fluorometric DNA quantification system
GFF3	General Feature Format version 3	NanoDrop	UV-based spectrophotometer
GC	guanine–cytosine content	Leu	leucine
ATP	adenosine triphosphate	Val	valine
NADH	nicotinamide adenine dinucleotide hydride	Ser	serine
ccm	cytochrome c biogenesis genes	Arg	arginine
cox	cytochrome c oxidase genes	Gly	glycine.