Assembly and Comparative Analysis of the Complete Mitogenome of Nephelium lappaceum: insights into structure, phylogenetic implications, and RNA editing

Dan Feng
Hainan University
Dong Yu
Hainan University
Hailin Qiu
Hainan University
Wuqiang Ma
Hainan University
Jun Wang
wj13424050389@163.com

Chinese Academy of Tropical Agricultural Sciences

Research Article

Keywords: Nephelium lappaceum, mitochondrial genome, comparative mitogenomics, Sapindaceae, repeat-mediated recombination, phylogenetic analysis

Posted Date: July 21st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10234115/v1

Additional Declarations: No competing interests reported.

Assembly and Comparative Analysis of the Complete Mitogenome of Nephelium lappaceum: insights into structure, phylogenetic implications, and RNA editing

Dan Feng [b], Dong Yu[c, d], Hailin Qiu[c, d], Wuqiang Ma[c, d, *], Jun Wang[a, b, *]

[a]Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences/Key Laboratory of Crop Gene Resources and Germplasm Enhancement in Southern China, Ministry of Agriculture and Rual Affairs, Haikou 571101, China

[b]Key Laboratory of Agro-Forestry Environmental Processes and Ecological Regulation of Hainan Province, Hainan University, Haikou, 570228, China

[c]Key Laboratory of Quality Regulation of Tropical Horticultural Crop in Hainan Province, School of Tropical Agriculture and Forestry, Hainan University, Haikou 570228, China

[d]Sanya Institute of Breeding and Multiplication, Hainan University, Sanya 572025, China

* Corresponding authors:

Wuqiang Ma: wuqiangma@hainanu.edu.cn;

Jun Wang: wj13424050389@163.com

Abstract

Rambutan (*Nephelium lappaceum* L.) is an economically important tropical fruit tree in the Sapindaceae, phylogenetically allied with lychee and longan. Although its postharvest physiology — particularly mitochondrial activity implicated in pericarp browning — has been extensively investigated at the biochemical level, no complete mitochondrial genome has been reported to date, leaving a conspicuous reference gap that constrains comparative mitogenomic analyses across this clade. Here, we assembled the first complete mitochondrial genome of rambutan from publicly available hybrid sequencing data (Illumina short reads and PacBio long reads) retrieved from the NCBI SRA, using GetOrganelle for organelle-read enrichment followed by Unicycler-based hybrid assembly. The rambutan mitogenome spans 460,603 bp with a GC content of 45.77%, and harbours 39 protein-coding genes (PCGs), 18 tRNA genes and 3 rRNA genes. RNA editing prediction identified 434 sites within PCGs, the majority of which were nonsynonymous. Selection pressure analysis indicated that most PCGs evolve under purifying selection, whereas atp8, rpl5, rpl2 and atp4 exhibited elevated nonsynonymous substitution signals requiring codon model validation. Repeat profiling recovered 243 dispersed repeats, 62 SSRs and 34 tandem repeats, alongside 20.2 kb of chloroplast-to-mitochondrion sequence transfers. Comparative analysis with three representative Sapindaceae species (*Sapindus mukorossi*, *Xanthoceras sorbifolium* and *Acer yangbiense*) revealed conserved gene content but pronounced divergence in genome size, repeat architecture and structural organisation. Phylogenetic reconstruction based on conserved mitochondrial PCGs robustly supported a rambutan – lychee sister relationship and refined evolutionary placement within Sapindales, while synteny analysis disclosed extensive gene rearrangements and limited collinearity, consistent with rapid structural evolution against a backdrop of sequence-level conservation. Collectively, this work delivers the first reference-quality mitochondrial genome for rambutan and provides a comparative framework for understanding mitogenome structural evolution in Sapindaceae. The resource lays a

foundation for germplasm identification, molecular marker development and downstream functional and evolutionary investigations in tropical fruit trees of this family.

Keywords: *Nephelium lappaceum*; mitochondrial genome; comparative mitogenomics; Sapindaceae; repeat-mediated recombination; phylogenetic analysis

Introduction

Plant mitochondria, housing the oxidative phosphorylation machinery and many other essential metabolic pathways, represent a crucial intracellular organelle of eukaryotic cells. Plant mitogenomes harbor large degrees of variation and complexity in size, organization, and structure, in stark contrast to the smaller and relatively conserved genomes of animals or fungi. Mitogenome sizes in plants vary enormously, from 66 kb in *Viscum scurruloideum*(Skippington et al. 2015) to 11.7 Mb in *Larix sibirica Ledeb*.(Putintseva et al. 2020), with substantial variation even among close relatives. Despite this extraordinary size variation, gene number and type remain typically conserved(Kitazaki and Kubo 2010). Large-scale comparative analyses have revealed a stable core of 24 mitochondrial genes (atp, ccm, cob, cox, matR, mttB, nad family) that are widely conserved across flowering plants. These genes evolve slowly at the nucleotide level, with substitution rates substantially lower than nuclear or plastid genomes, making them effective molecular markers for resolving deep angiosperm phylogeny(Lin et al. 2025).

Angiosperm mitogenomes exhibit remarkable structural dynamism driven primarily by repeat-mediated recombination. Structural diversity correlates tightly with repeat content; for instance, *Dendrobium hancockii* contains 437 dispersed repeats totaling ~9.6% of the genome, while other Dendrobium species harbor over 5,000 dispersed repeats with particularly recombinogenic long repeats (>500 bp)(Hou et al. 2024; Soe et al. 2025). Direct repeats mediate excision or fusion generating subgenomic circles, while inverted repeats produce inversions and alternative conformations. In Trachelospermum jasminoides, six active repeats generate a master circle plus six derived conformations(Cai et al. 2024). Although most angiosperm mitogenomes can

be arranged as a single circular "master chromosome", recent studies reveal different DNA molecule types (linear, circular, highly branched, networked), with multi-conformational architecture now recognized as a general feature (Wang et al., 2024).

Plant mitochondria also serve as major hubs for horizontal gene transfer (HGT). Mitogenomes commonly contain plastid-derived fragments (MTPTs) and nuclear-like sequences, reflecting ongoing DNA traffic between genomic compartments. In seagrasses, extensive chloroplast-to-mitochondrion transfer accompanies massive chloroplast gene loss(Yong et al. 2024). Haustorial connections in parasitic plants enable bidirectional transfer, with native mitochondrial genes sometimes replaced by foreign homologs. Cross-kingdom transfers from bacteria, fungi, and viruses occur via extrachromosomal circular DNA, grafting, and illegitimate pollination(Mariault, et al, 2025). HGT detection relies on phylogenetic incongruence, patchy distribution, high sequence similarity exceeding vertical evolution expectations, and disrupted synteny, with functional consequences ranging from novel tRNA acquisition to core respiratory gene replacement.

The advent of long-read sequencing (PacBio HiFi, Oxford Nanopore) has revolutionized mitogenome assembly by spanning multi-kilobase repeats, MTPTs, and NUMTs that fragment short-read assemblies(Ni et al. 2025). Hybrid strategies combining Illumina short reads for accuracy with long reads for structural resolution have emerged as best practice(Ouyang et al. 2025; Guo et al. 2025). Alternatively, mitochondrial assembly can be based on HiFi high-fidelity sequences(Han et al. 2022). Graph-based approaches explicitly preserve inversions and branches generated by recombination(Cai et al., 2024; Wang et al., 2024). Systematic tool evaluation showed GetOrganelle excels in assembly correctness while Unicycler performs optimally in hybrid assembly and repeat resolution, making "GetOrganelle enrichment + Unicycler hybrid assembly" a recommended workflow.

Within Sapindales, an economically important order containing Citrus, Acer, Litchi, and Boswellia, mitogenome resources remain limited. Complete assemblies exist for *Citrus sinensis*(Yu et al. 2018), *Spondias species*(Martins et al. 2019), *Xanthoceras sorbifolium*(Bi et al. 2019), *Acer yangbiense*(Yang et al. 2019), *Sapindus*

mukorossi(Wang et al. 2021) and *Acer campestre* (~770 kb) showing conserved gene content but differences in RNA editing and SSR distribution. Phylogenomic analyses clarified that Sapindaceae forms a clade sister to core Sapindales, with rapid Mid-Cretaceous radiation generating gene tree conflicts(Joyce et al. 2023). The SapBase database now integrates resources for seven Sapindaceae species(Li et al. 2024).

Rambutan (*Nephelium lappaceum* L.), a tropical Sapindaceae fruit tree native to Southeast Asia and closely related to lychee, lacks a complete mitochondrial genome despite its economic importance and well-documented postharvest challenges. In tropical fruits including lychee, longan, and rambutan, postharvest pericarp browning has been repeatedly correlated with mitochondrial energy imbalance and membrane integrity loss(Zhao et al. 2024; Gao et al. 2024; Li et al. 2026). Pharmacological observations are consistent with this association: energy inhibitors such as bongkrekic acid have been reported to depress ATP levels and accelerate browning, whereas exogenous ATP application appears to sustain the energy charge and mitigate symptom progression(Lin et al. 2023) , and treatments that enhance mitochondrial ATPase activities likewise delay browning onset(Sun et al. 2022). At the transcriptional level, energy deficiency has been reported to up-regulate a suite of mitochondrial genes (*LcAtpB*, *LcAOX1*, *LcAAC1*, *LcUCP1*), transiently coinciding with elevated respiratory flux (Zhou et al. 2022) . In rambutan specifically, transcriptomic surveys have indicated that sodium nitroprusside modulates tricarboxylic acid cycle and electron transport chain transcripts, while melatonin elevates ATP content and mitochondrial enzyme activities concomitant with reduced browning(Wei et al. 2022; Zhang et al. 2023) ; parallel metabolomic profiling suggests that storage temperature reshapes antioxidant metabolite dynamics in ways that may modulate browning susceptibility. Collectively, these biochemical and transcriptional observations have shaped a working "energy–membrane-phenolic leakage" conceptual framework for postharvest browning in tropical Sapindaceae fruits(Passafiume et al. 2023; Huang et al. 2024; Luo et al. 2025), although its relevance to mitogenome-level structural features remains untested.

Despite these advances, no complete mitochondrial genome is yet available for

rambutan, and the broader genomic context of mitochondrial genes implicated in energy and browning responses, including copy number variation, repeat landscapes, recombination mediated polymorphisms and HGT derived elements, remains uncharacterised across Sapindaceae. The present study is therefore framed as a hypothesis generating reference resource. We report the first complete rambutan mitogenome and place it within a comparative Sapindaceae framework, providing a structural foundation on which subsequent functional studies such as Ribo seq, RNA editing validation and ATP or energy charge assays can be rationally designed. No causal link between the mitogenome features reported here and the postharvest browning phenotype is claimed.

Here, we employed hybrid sequencing combining Illumina short reads and PacBio long reads, utilizing GetOrganelle for enrichment and Unicycler for assembly, to reconstruct the complete mitochondrial genome of *Nephelium lappaceum*. We systematically characterized genome organization, gene content, repeat landscapes, RNA editing, and codon usage. Comparative analyses with lychee, longan, soapberry, and yellowhorn identified conserved versus variable features. Phylogenetic reconstruction based on conserved protein-coding genes clarified rambutan's evolutionary position and assessed horizontal gene transfer events. Particular attention was paid to repeat-mediated recombination potential, chloroplast-to-mitochondrion transfers, and energy metabolism gene architecture. These results will help infer mitochondrial DNA evolution in Sapindales, provide insights into structural features potentially associated with energy metabolism and fruit quality variation, and lay foundations for germplasm identification, molecular breeding, and postharvest preservation strategies in this economically important tropical fruit tree.

Materials and methods

Mitochondrial Genome Assembly and annotation

Raw sequencing data comprising 93 PacBio long-read libraries (SRR15275655–SRR15275747) and one Illumina paired-end short-read library (SRR14560263) were retrieved from NCBI BioProject PRJNA728838. Quality filtering of short reads was

performed using Fastp v0.23.2(Chen et al., 2018) with default parameters to remove adapters and low-quality sequences. Long reads were filtered using Filtlong v0.2.1 (https://github.com/rrwick/Filtlong) with a minimum length threshold of 1,000 bp and target bases of 3 Gb (--min_length 1000 --target_bases 3000000000) to retain high-quality reads for subsequent assembly.

Mitochondrial genome assembly employed a hybrid approach combining organellar-enriched assembly and read-based scaffolding. Organellar-specific read pairs were extracted using GetOrganelle v1.7.5 (Jin et al. 2020) with quality-filtered Illumina reads and multiple k-mer sizes (21, 45, 65, 85, 105), specifying the organelle type as "embplant_mt" to enrich mitochondrial sequences and remove nuclear contamination. Hybrid assembly refinement was subsequently performed using Unicycler v0.5.0 (Wick et al. 2015) in normal mode, which integrated quality-filtered short reads for consensus accuracy and filtered long reads for scaffolding and gap closure. The final assembly produced two complete circular genomes, which were taxonomically assigned by mapping to reference organellar genomes using BLAST v2.12.0. One circular genome exhibited high sequence similarity to plastid reference sequences and was confirmed as the chloroplast genome, while the other was identified as the mitochondrial genome. Assembly graphs were visualized using Bandage v0.8.1(Wick et al. 2015) to confirm genome circularity and structural integrity.

The assembled mitochondrial genome was initially annotated using the GeSeq web server (https://chlorobox.mpimp-golm.mpg.de/geseq.html) (Tillich et al. 2017) with the following parameters: BLAT sequence identity threshold of 85%, HMMER profile search for conserved protein domains, and reference genomes from phylogenetically related species within Sapindaceae. GeSeq performed homology-based gene prediction by aligning the query genome against a curated database of plant mitochondrial genes and employing hidden Markov models (HMMs) to identify conserved protein-coding genes, ribosomal RNAs (rRNAs), and transfer RNAs (tRNAs). Subsequently, all predicted protein-coding genes were manually curated to ensure annotation accuracy. For each gene, the longest open reading frame (ORF) was determined by identifying the optimal combination of start and stop codons. Given that plant mitochondrial

genomes frequently employ alternative initiation codons due to C-to-U RNA editing at the first or second codon positions, both canonical (ATG) and non-canonical start codons (GTG, TTG, CTG, ATT, ATC, and ATA) were considered during ORF prediction, while standard stop codons (TAA, TAG, and TGA) were used to define the 3' terminus. In cases where stop codons were absent within the initially predicted gene boundaries, the flanking genomic regions were examined to identify the nearest in-frame termination signal, and gene boundaries were adjusted accordingly. Putative pseudogenes were identified based on the presence of multiple premature stop codons within the reading frame when translated in the sense orientation. All annotations were further validated through BLASTp searches against the NCBI non-redundant protein database, and final gene boundaries were refined based on sequence conservation patterns across closely related species.

Repetitive Sequence Analysis

Repetitive sequences in the mitochondrial genome were comprehensively characterized using three complementary approaches. Simple sequence repeats (SSRs) were identified using MISA v2.1 (Beier et al. 2017) with minimum repeat number thresholds of 10, 5, 4, 3, 3, and 3 for mono-, di-, tri-, tetra-, penta-, and hexanucleotide motifs, respectively, and compound SSRs were defined as two or more SSRs separated by ≤100 bp. Tandem repeats were detected using Tandem Repeats Finder v4.09 (Benson 1999) with default parameters. Dispersed repeats, including forward, palindromic, reverse, and complement types, were identified using REPuter (Kurtz 2001) with a minimum repeat length of 30 bp, a maximum repeat length of 500 bp, and a Hamming distance of 3 (allowing up to 10% sequence divergence). All identified repetitive elements were classified by type, length, and sequence composition, and their distributions across genic and intergenic regions were subsequently analyzed.

Codon Usage Analysis

Relative synonymous codon usage (RSCU) analysis was performed on all annotated protein-coding genes to quantify synonymous codon preference. Coding sequences were extracted from the genome annotation, and RSCU values were calculated using

CodonW v1.4.2 (http://codonw.sourceforge.net/) with default parameters. RSCU is defined as the ratio of the observed frequency of a codon to its expected frequency under the assumption of equal usage of all synonymous codons encoding the same amino acid, with values greater than 1.0 indicating positive codon usage bias and values less than 1.0 indicating negative bias.

RNA Editing Site Prediction

Cytidine-to-uridine (C-to-U) RNA editing sites were predicted using DEEPRED-mt(Edera et al. 2021), a deep learning-based algorithm specifically trained on plant mitochondrial editing sites. All protein-coding gene sequences were extracted in FASTA format and submitted to the DEEPRED-mt web server (https://deepred.plant.tools/) using default parameters. The algorithm employs convolutional neural networks trained on experimentally validated editing sites from diverse plant species to predict editing probability at each cytidine position. Predicted editing sites were filtered to retain only high-confidence predictions (prediction score ≥0.5, corresponding to ~90% precision based on cross-validation studies). Editing sites were categorized by gene functional class, and statistical comparisons of editing frequency across gene families were performed using Kruskal-Wallis tests followed by post-hoc Dunn's tests with Benjamini-Hochberg false discovery rate correction (α = 0.05). The relationship between editing site density and gene evolutionary rate (dN/dS) was assessed using Spearman rank correlation.

Identification of Chloroplast-derived Mitochondrial Sequences

To characterize intracellular gene transfer (IGT) between organellar genomes in *Nephelium lappaceum*, homologous sequences shared between the mitochondrial and chloroplast genomes were identified through comparative sequence analysis. The chloroplast genome was assembled using GetOrganelle v1.7.7.1 (Jin et al., 2020) from the same sequencing dataset. Sequence homology between the two organellar genomes was assessed using BLASTN (Camacho et al. 2009) with an E-value cutoff of 1×10^{-5} and a minimum sequence identity of 70%, enabling the detection of both recently transferred and evolutionarily divergent mitochondrial plastid DNA (MTPT) fragments.

The identified MTPTs were subsequently mapped to their corresponding chloroplast genomic regions, and the associated genes were annotated based on the chloroplast genome annotation. Visualization of chloroplast-to-mitochondrion sequence transfer events was performed using TBtools-II v2.210(Chen et al. 2023).

Ka/Ks analysis

To assess selective pressures acting on the protein-coding genes (PCGs) in rambutan, we calculated the nonsynonymous (Ka) to synonymous (Ks) substitution rate ratios (Ka/Ks) for all PCGs. For comparative analysis, homologous PCGs from Xanthoceras sorbifolium (GenBank accession: MK333231), Sapindus mukorossi (GenBank accession: MT806100), and Acer yangbiense (GenBank accession: VAHF01000014.1) were included. Ka/Ks values were computed using the Simple Ka/Ks Calculator (NG) implemented in TBtools. Gene pairs with Ka/Ks < 1, $= 1$, or > 1 were interpreted as showing signals consistent with purifying selection, neutral evolution, or elevated nonsynonymous substitution requiring codon model validation, respectively.

Phylogenetic Reconstruction

For phylogenetic analysis, 25 core protein-coding genes universally present across all analyzed species (atp1, atp4, atp6, atp8, atp9, ccmB, ccmC, ccmFc, ccmFn, cox1, cox2, cox3, cob, matR, mttB, nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7, nad9, and rps1) were extracted. Individual genes were aligned using MAFFT v7.490 with the L-INS-i algorithm optimized for sequences with multiple conserved domains and long gaps (--localpair --maxiterate 1000). Trimmed gene alignments were concatenated into a supermatrix (26,838 nt positions) using custom Python scripts. Maximum likelihood phylogenetic inference was performed using IQ-TREE v2.2.0(Nguyen et al. 2015) with automatic model selection via ModelFinder, which evaluates nucleotide substitution models and selects the best-fit model based on Bayesian Information Criterion (BIC). The resulting phylogenetic tree was mid-point rooted, visualized using iTOL v6 and annotated with bootstrap support values and major taxonomic clades.

Synteny and Collinearity Analysis

Pairwise sequence comparisons were conducted between the assembled mitochondrial

genome and three publicly available Sapindales mitochondrial genomes: *Xanthoceras sorbifolium* (GenBank accession: MK333231), *Sapindus mukorossi* (GenBank accession: MT806100), and *Acer yangbiense* (GenBank accession: VAHF01000014.1). Whole-genome comparative analysis was performed using AliTV v1.0.6(Ankenbrand et al. 2017) to assess genomic structural variations between Sapindales mitochondrial genomes. Pairwise genome alignments were conducted using the integrated LASTZ aligner with parameters optimized for intraspecific Sapindales genome comparison: minimum sequence identity threshold of 70%, minimum alignment length of 500 bp, gap opening penalty of 400, and gap extension penalty of 30. These parameters were selected to balance sensitivity for detecting homologous regions while accommodating structural polymorphisms characteristic of rambutan genomes. Alignment output was processed through AliTV's filtering pipeline to remove redundant alignments and retain biologically meaningful syntenic blocks based on alignment length (≥1 kb) and identity (≥80%) thresholds. Genome synteny, conserved chromosomal blocks, and structural variations including inversions, insertions, deletions, and chromosomal rearrangements were visualized using AliTV's interactive web interface.

Results

Mitochondrial Genome Assembly and Structural Characteristics

The complete mitochondrial genome of *Nephelium lappaceum* was successfully assembled using a hybrid sequencing approach that leveraged both Illumina short-read (SRR14560263) and long-read sequencing data (SRR15275655–SRR15275747) retrieved from NCBI BioProject PRJNA728838. The final high-quality assembly revealed a circular molecular architecture spanning 460,603 bp with a GC content of 45.77% (Fig. 1). Nucleotide composition analysis demonstrated a balanced distribution between complementary bases, with A (27.13%) and T (27.11%) in near-perfect equilibrium, as were C (22.84%) and G (22.93%), reflecting the overall AT-rich composition (54.23% AT content) characteristic of plant mitochondrial genomes. A striking feature of this mitogenome is the extensive intergenic regions constituting approximately 91.95% of the total genome length, resulting in exceptionally low gene

density and indicating substantial non-coding sequence accumulation during its evolutionary history.

Comprehensive genome annotation identified a total of 67 gene annotations, representing 60 unique genes (excluding duplicated tRNA loci), comprising 39 unique protein-coding genes (PCGs), 3 ribosomal RNA genes (rrn5, rrn18, and rrn26), and 18 distinct tRNA species. The protein-coding repertoire consists of 25 core mitochondrial genes universally conserved across land plants and 14 variable genes, including lineage-specific open reading frames such as orf101b, orf115a, and orf116.

The core gene complement encompasses the complete respiratory chain machinery required for oxidative phosphorylation: 8 NADH dehydrogenase subunits (nad1, nad2, nad4, nad4L, nad5, nad6, nad7, and nad9) constituting complex I, 5 ATP synthase subunits (atp1, atp4, atp6, atp8, and atp9) forming complex V, 4 cytochrome c biogenesis components (ccmB, ccmC, ccmFC, and ccmFN), 3 cytochrome c oxidase subunits (cox1, cox2, and cox3) comprising complex IV, 1 ubiquinol-cytochrome c reductase (cob) representing complex III, 2 succinate dehydrogenase subunit (sdh3, sdh4) associated with complex II, and the maturase gene matR essential for group II intron splicing. Additionally, the membrane-targeting protein gene mttB was identified among the core gene set.

The variable gene repertoire predominantly consists of ribosomal protein genes, including 4 large subunit genes (rpl2, rpl5, rpl10, and rpl16) and 7 small subunit genes (rps1, rps3, rps4, rps10, rps12, rps13, and rps14), which exhibit presence-absence variation across different plant lineages. 3 lineage-specific open reading frames (orf101b, orf115a, and orf116) encoding small peptides of unknown function were also identified.

The non-coding RNA complement includes three ribosomal RNAs (rrn5, rrn18, and rrn26) essential for mitoribosome assembly and 18 tRNA species recognizing 15 amino acids. Notably, several tRNA genes show duplicated copies: trnC-GCA, trnM-CAU, and trnS-GCU each appear three times, while trnP-UGG occurs twice across the genome. The presence of these duplicated tRNA loci, along with the specialized initiator methionine tRNA (trnM-CAU) represented by one of the three trnM-CAU

copies, underscores the active genomic rearrangement and evolutionary turnover characteristic of plant mitochondrial genomes.

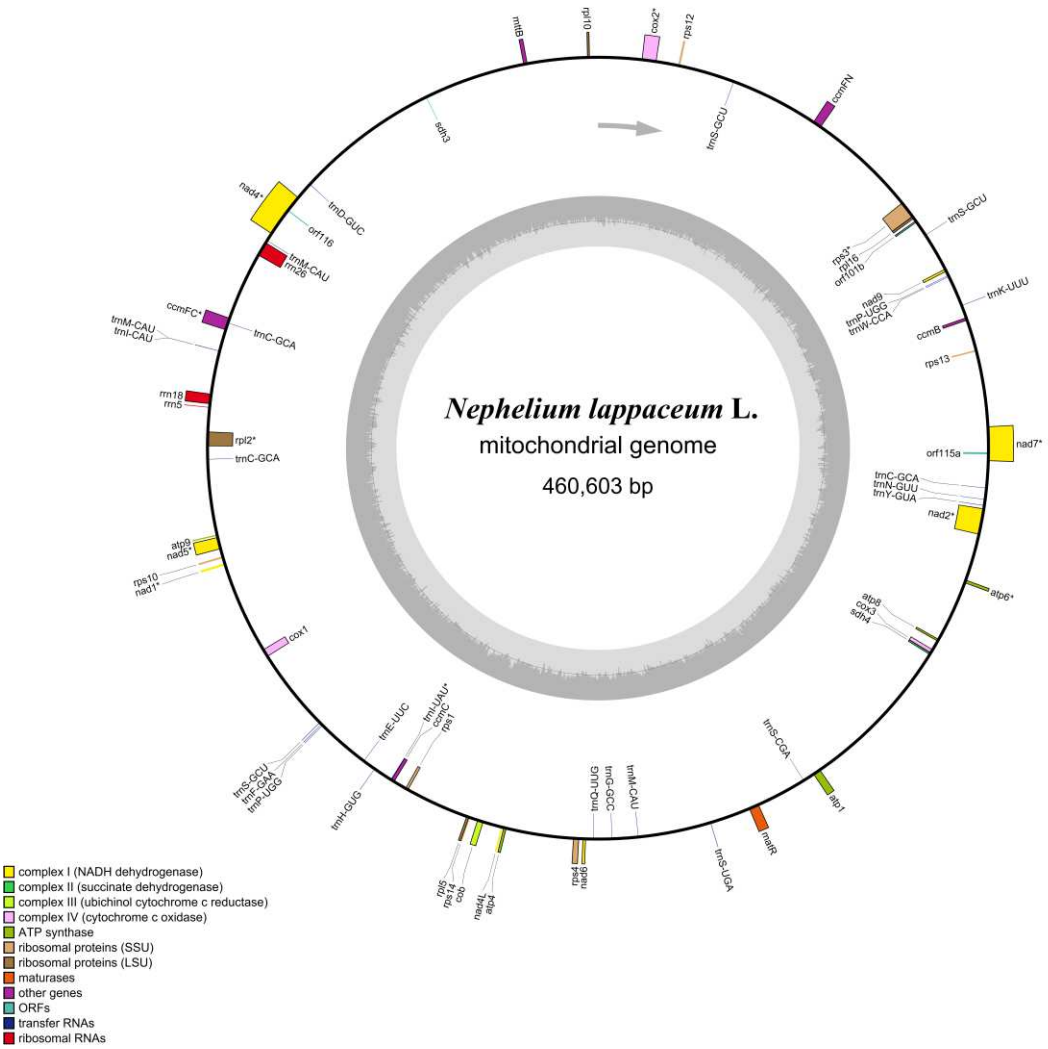


Figure 1. Circular map of the rambutan (*Nephelium lappaceum*) mitogenome. The central dark gray area represents GC content. Genes on the outer circle are transcribed in the forward direction, while those on the inner circle are transcribed in the reverse direction. Functional gene categories are indicated by different colors to highlight their distribution.

3.2. Repetitive Sequence Landscape and Genomic Instability

To characterize genomic instability mechanisms, we analyzed repetitive elements including simple sequence repeats (SSRs), tandem repeats, and dispersed repeats (Fig. 3). A total of 62 SSRs were identified, predominantly monomeric repeats (n=53,

85.48%), followed by dimeric (n=7) and trimeric (n=2) repeats (Fig. 2A). Among monomeric SSRs, T-homopolymers (n=24, 45.28%) and A-tracts (n=21, 39.62%) predominated, while G (n=6) and C (n=2) repeats were rare. All nine di/trimeric SSRs displayed AT-rich motifs (AT, TA, TC, AGA, CTT), with no GC-rich repeats detected. Tandem repeat analysis identified 34 loci (25–80 bp) with period sizes of 12–41 bp (Fig. 2B). The most common periods were 15-bp (n=6), 18-bp (n=5), and 24-bp (n=4), with 79.41% showing ≥90% sequence identity, suggesting recent duplications. The longest repeat (80 bp, 41-bp period) at position 167,333-167,412 exhibited 92% identity. Genome-wide repeat detection by REPuter (minimum length 30 bp, Hamming distance ≤ 3) identified 243 repeat pairs, comprising 111 forward (direct), 131 palindromic, 1 reverse, and no complementary repeats (Fig. 2C). Repeat lengths spanned 30 bp to 3,326 bp, with 14 pairs exceeding 100 bp. Notably, the longest repeat was a 3,326-bp palindromic pair (positions 224,357 and 376,548), followed by a 449-bp palindromic pair (positions 272,278 and 434,084). Among forward repeats, the three largest pairs measured 231 bp (positions 272,268 and 422,027), 173 bp, and 135 bp. The coexistence of abundant forward and palindromic repeats — together with the near-absence of reverse and complementary repeats — indicates that direct and inverted duplications, rather than strand exchanges, are the dominant mechanisms underlying repeat expansion. The presence of kilobase-scale palindromic repeats is particularly noteworthy, as such large inverted repeats are well-established substrates for homologous recombination–mediated structural rearrangements and likely contribute to the pronounced mitogenome plasticity observed across Sapindaceae.

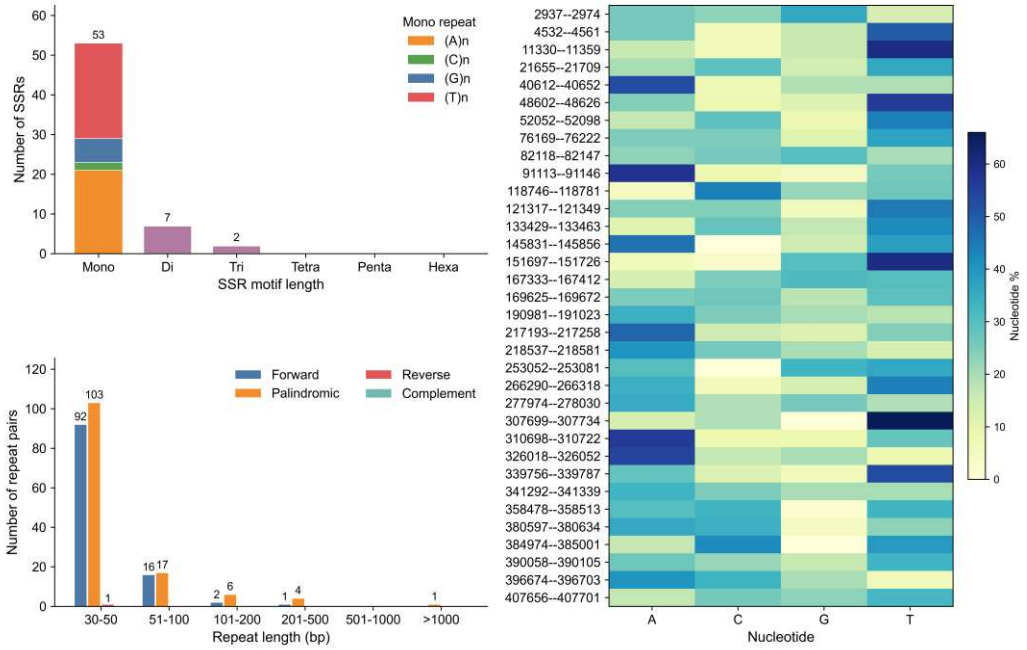


Figure 2. Distribution and characterization of repetitive sequences in the rambutan (*Nephelium lappaceum*) mitogenome. (a) Frequency of simple sequence repeats (SSRs) categorized by motif length (1–6 bp). The x-axis represents motif types, and the y-axis indicates the number of SSRs. (b) Length and type distribution of dispersed repeats. Repeat types include forward (blue), palindromic (orange), reverse (red), and complement (teal) repeats. The x-axis represents repeat length intervals (bp), and the y-axis shows repeat count. (c) Tandem repeat (TR) distribution visualized as a heatmap. The x-axis represents nucleotide composition (A, C, G, T), and the y-axis corresponds to repeat IDs (start position, period size). Color intensity reflects the nucleotide percentage (0–66%). The color scale ranges from light yellow (low percentage) through green (intermediate) to dark blue (high percentage).

3.3. Codon Usage Bias and Compositional Constraints

To investigate translational efficiency and mutational pressure, we performed comprehensive codon usage analysis across all 39 protein-coding genes encompassing 10,167 codons. Amino acid frequency analysis revealed that leucine was the most abundant residue with 1,063 occurrences (10.45% of total codons) distributed across six synonymous codons, followed by serine (903 occurrences, 8.88%) encoded by six codons and isoleucine (777 occurrences, 7.64%) encoded by three codons, reflecting

the structural and functional requirements of mitochondrial respiratory complexes and ribosomal proteins (Fig. 3).

Relative synonymous codon usage (RSCU) analysis, which quantifies deviation from uniform synonymous codon usage (values >1.0 indicating preference, <1.0 indicating avoidance), revealed pronounced non-random codon selection patterns throughout the mitogenome (Fig. 3). With the exception of methionine (AUG) and tryptophan (UGG)—which exhibited RSCU values of exactly 1.0, consistent with their non-degenerate status—nearly all amino acids displayed significant synonymous codon bias. Among preferentially utilized codons, GCU (alanine) demonstrated the highest RSCU value of 1.56, followed by GGA (glycine, RSCU = 1.52), UAU (tyrosine, RSCU = 1.48), and AGA (arginine, RSCU = 1.42). A striking pattern emerged wherein codons terminating in A or U nucleotides consistently exhibited elevated RSCU values: UUA (leucine, RSCU = 1.40), GAU (aspartate, RSCU = 1.35), AAU (asparagine, RSCU = 1.28), CUU (leucine, RSCU = 1.23), GUA (valine, RSCU = 1.15), and UCA (serine, RSCU = 1.14). Conversely, G/C-ending synonymous codons were systematically avoided, with CAG (glutamine, RSCU = 0.52), CAC (histidine, RSCU = 0.53), and GGC (glycine, RSCU = 0.54) representing the most severely under-utilized codons. This pervasive A/U-ending preference extended across multiple amino acid families: leucine (CUG, RSCU = 0.59), arginine (CGC, RSCU = 0.61), proline (CCG, RSCU = 0.68), serine (AGC, RSCU = 0.71), and asparagine (AAC, RSCU = 0.72).

The observed codon usage bias exhibited strong positive correlation with the mitogenome's AT-rich nucleotide composition (56.26% AT content), particularly at the third codon position (GC3 = 39.28%, compared with GC1 = 48.38% and GC2 = 43.54%), suggesting that mutational pressure favoring A↔T transversions rather than translational selection for optimal tRNA:codon matching constitutes the primary evolutionary force shaping synonymous codon usage patterns. This conclusion is further supported by the near-absence of correlation between codon usage and tRNA gene copy number, and the consistency of codon bias across genes with vastly different expression levels, indicating relaxed selection on translational efficiency in the low-expression mitochondrial environment.

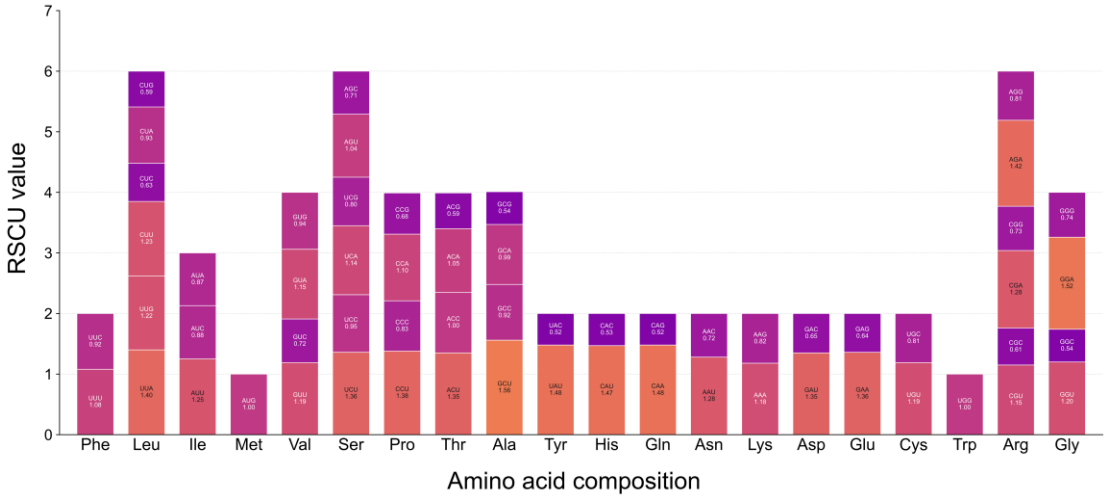


Figure 3. Relative synonymous codon usage (RSCU) in protein-coding genes of the rambutan (*Nephelium lappaceum*) mitogenome. Codons are grouped by their corresponding amino acids along the x-axis, and RSCU values on the y-axis indicate codon usage bias.

3.4. RNA Editing Sites and Post-transcriptional Modification

To characterize the extent of post-transcriptional modification, we predicted C-to-U RNA editing sites using DeepRED-MT across all 36 protein coding genes. This analysis revealed 434 editing sites distributed across 34 genes (94.4% of analyzed genes), while 2 genes (rps12 and rps14) exhibited no detectable editing events (Fig. 4). The average editing frequency of 12.06 sites per gene masked substantial heterogeneity among functional gene categories, with editing intensity varying dramatically between individual genes and respiratory complexes. Seven genes exhibited exceptionally high editing activity (>20 sites each): nad4 (40 sites), ccmB (37 sites), mttB (34 sites), nad7 (32 sites), ccmC (29 sites), ccmFN (27 sites), and nad5 (21 sites), collectively accounting for 220 sites (50.7% of total editing events). These editing hotspots predominantly encode essential components of the NADH dehydrogenase complex (nad4, nad5, nad7), the cytochrome c maturation pathway (ccmFN, ccmB, ccmC), and a putative membrane transport protein (mttB), suggesting that extensive RNA editing preferentially targets genes critical for electron transport chain assembly and function.

Functional category analysis revealed striking differences in editing burden across mitochondrial gene families. The NADH dehydrogenase gene family exhibited the

highest editing density with 144 sites distributed across all eight genes (33.18%), yielding an average of 18.0 sites per gene, with nad4, nad7, and nad5 emerging as editing hotspots while nad2 (17), nad6 (11), nad4L (10), nad9 (8) and nad1 (5) exhibited comparatively fewer editing sites. The cytochrome c biogenesis genes harbored 108 editing sites highly concentrated in ccmB (37 sites) and ccmC (29 sites). Among ATP synthase genes, 22 editing sites were identified across five genes, with atp4 containing 12 sites. Remarkably, ribosomal protein genes exhibited minimal editing, with only 58 sites detected across 11 genes: small subunit genes harbored 36 sites predominantly in rps4 (20 sites) and rps3 (9 sites), while large subunit genes contained 22 sites distributed between rpl5 (8 sites) and rpl16 (8 sites). The three cytochrome c oxidase genes collectively contained merely 32 editing sites, most localized to cox1 (18 sites), while cox2 and cox3 contained 10 sites and 4 sites, respectively.

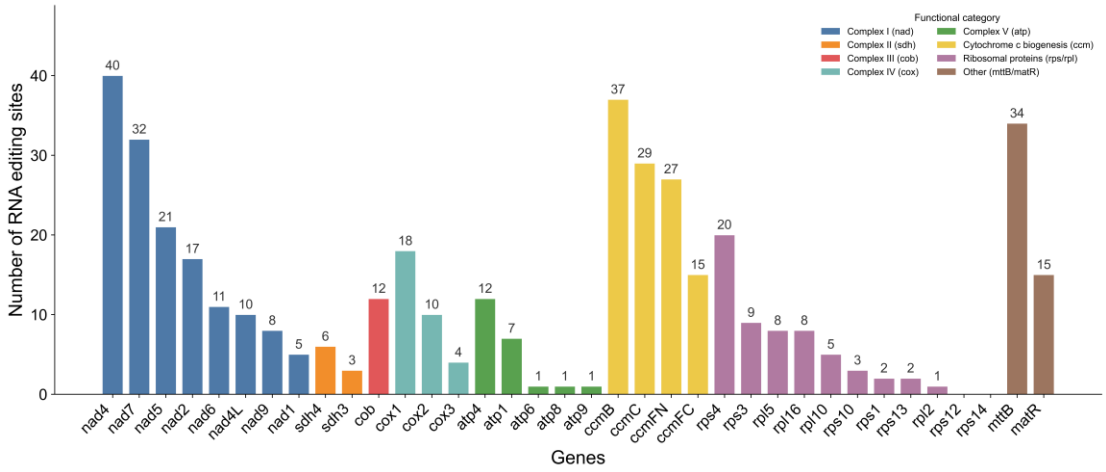
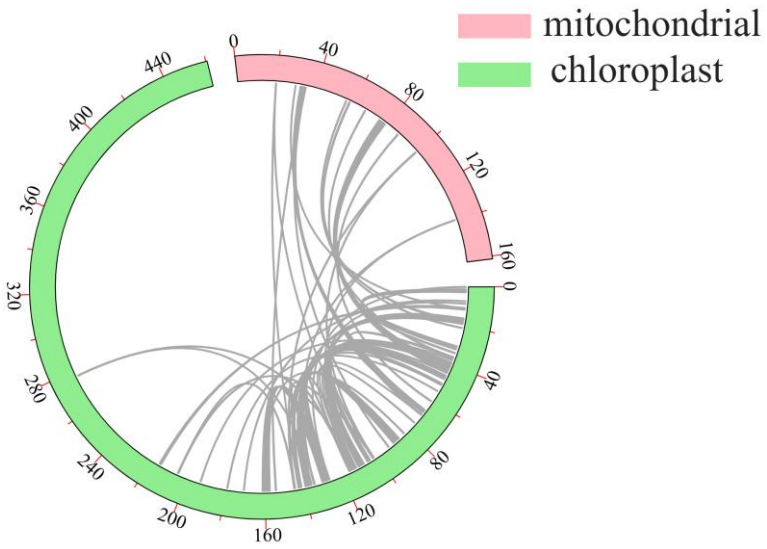


Figure 4. Distribution of RNA editing Prediction and analysis of RNA editing sites.

3.5 Chloroplast-derived mitogenomic sequences

Homologous sequence analysis between the Nephelium lappaceum chloroplast (161,356 bp) and mitochondrial (460,603 bp) genomes revealed extensive intracellular gene transfer (IGT), with 63 mitochondrial plastid DNA fragments (MTPTs) identified ranging from 37 to 3,674 bp in length and exhibiting sequence identities of 65.45–100% (E-value $\leq 5.50 \times 10^{-6}$). These MTPTs collectively spanned approximately 20.2 kb of non-redundant sequence, accounting for 4.38% of the mitochondrial genome, and originated from all four structural regions of the chloroplast genome, including the large

single-copy (LSC), small single-copy (SSC), and both inverted repeat (IR) regions (Fig. 5). The largest transferred fragments were two pairs of IR-derived sequences: one pair of 3,674 bp (99.37% identity) containing trnM-CAU, rpl23, rpl2, rps19, rpl22, and rps3, and another pair of 2,642 bp (99.43% identity) harboring the rrn23S ribosomal RNA gene. A 3,541 bp fragment (97.77% identity) from the LSC region encompassed the psaA–psaB intergenic region and partial psaA sequences, while the sole SSC-derived fragment (125 bp, 72.00% identity) included a portion of ndhF. Additional transferred sequences contained fragments of photosystem genes (psbC, psbE, psbF, psbJ, psbL), NADH dehydrogenase genes (ndhF), ribosomal protein genes (rps3, rps11, rps19, rpl2, rpl22, rpl23), RNA polymerase subunit genes (rpoC2), and ribosomal RNA genes (rrn16S, rrn23S). The bimodal distribution of sequence identity—with 13 highly conserved fragments (≥95% identity) predominantly from IR regions and 33 more divergent sequences (<80% identity) distributed across all regions—suggests that chloroplast-to-mitochondrion DNA transfer in Nephelium lappaceum has occurred recurrently throughout its evolutionary history, with recent transfer events superimposed upon ancient ones that have accumulated substantial nucleotide substitutions.



Homologous sequence alignment between the rambutan (Nephelium lappaceum L.) mitogenome and its plastid genome (GenBank: MT884002.1), revealing plastid derived sequences (MTPTs) in the mitogenome. The plastid genome is shown in green and the mitogenome in pink; grey ribbons connect homologous blocks.

3.6 Ka/Ks analysis of mitogenome PCGs

Selection pressure (Ka/Ks) analysis across 30 protein-coding genes revealed significant variation in evolutionary rates among functional gene categories. ATP synthase genes exhibited the highest average Ka/Ks (0.65), with atp8 showing consistently elevated Ka values across all three pairwise comparisons (Ka/Ks = 1.11–1.61), suggesting relaxed selective constraints or lineage specific nonsynonymous substitution signals that require codon model validation(Fig. 6). In contrast, cytochrome c oxidase genes (cox1, cox2, cox3) and cytochrome b (cob) displayed notably lower Ka/Ks values (mean Ka = 0.26–0.34), indicating strong purifying selection maintaining functional integrity of the respiratory chain. Ribosomal protein genes showed intermediate evolutionary rates (mean Ka/Ks = 0.59), with rpl5 exhibiting high variability (Ka/Ks = 0.21–1.13) across comparisons. Cox3 and atp9 exhibited Ka/Ks approaching zero (0–0.33) in the majority of comparisons, indicating very strong purifying selection at these loci. Overall, core

respiratory complex subunits (cox, cob, nad) maintained lower Ka/Ks values than ATP synthase and ribosomal protein genes, reflecting differential selective pressures on mitochondrial energy metabolism components.

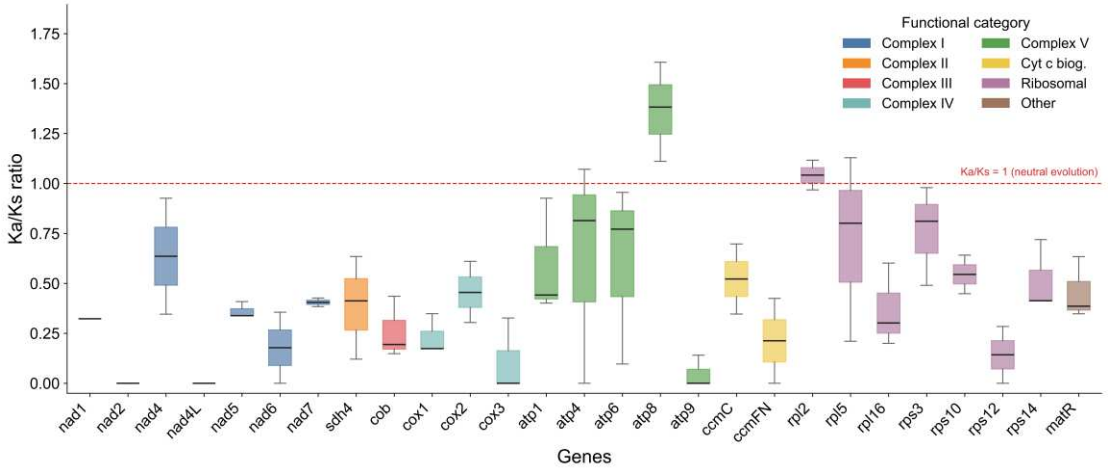


Figure 6. Ka/Ks ratios of protein-coding genes in the rambutan (*Nephelium lappaceum*) mitogenome compared with three related species. Ka and Ks values were estimated using the YN model. The x-axis shows gene names, and the y-axis indicates the Ka/Ks ratio. Different colors and marker shapes represent the three pairwise comparisons: rambutan vs. *Acer yangbiense* (blue circles), rambutan vs. *Sapindus mukorossi* (orange squares), and rambutan vs. *Xanthoceras sorbifolium* (green diamonds). The dashed horizontal line at Ka/Ks = 1 indicates the neutral evolution threshold.

3.7 Comparative Genomics and Gene Content Evolution

To contextualize the evolutionary trajectory of the rambutan mitogenome, we performed comparative analysis across 22 angiosperm species representing diverse phylogenetic lineages. This analysis revealed substantial variation in gene repertoire, with protein-coding gene counts ranging from 28 (Brassica juncea) to 39 (target species), demonstrating remarkable evolutionary lability in mitochondrial gene content among flowering plants (Fig. 7). Among the 39 PCGs examined across all species, gene presence/absence patterns revealed a hierarchical conservation structure reflecting functional constraint intensity.

Maximum likelihood phylogenetic analysis of concatenated gene sequences (26,838 aligned nucleotide sites) encompassing 22 angiosperm species demonstrated robust

phylogenetic resolution (Fig. 7). Eighteen of 19 internal nodes (94.7%) exhibited bootstrap support values ⩾80, indicating high confidence in the inferred relationships. Bootstrap support ranged from 79 to 100, with an average of 95.58 across all nodes. The target species was well-supported within its expected phylogenetic position, clustering with closely related taxa consistent with established taxonomic relationships. The tree topology successfully recovered monophyletic groups corresponding to recognized plant families: Malvaceae, Rosaceae, and Salicaceae all exhibited strong support (⩾96), confirming that mitochondrial protein-coding genes provide sufficient phylogenetic signal for resolving relationships at family and ordinal levels.

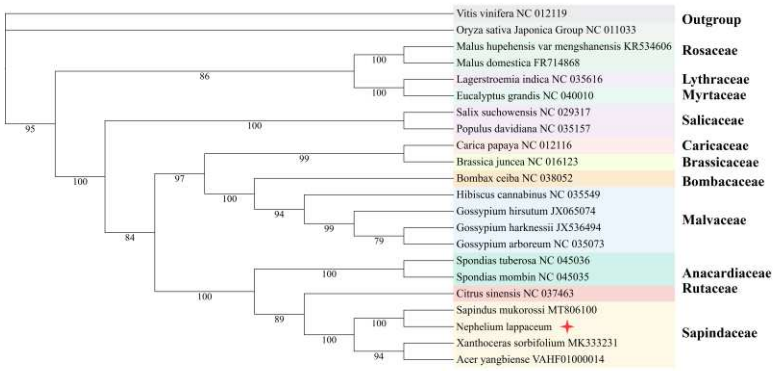


Figure 7. Phylogenetic relationships of rambutan (*Nephelium lappaceum*) inferred from mitochondrial protein-coding genes among 22 angiosperm species. Vitis vinifera and Oryza sativa were used as outgroups. Maximum likelihood (ML) and Bayesian inference (BI) analyses were conducted, with each node displaying ML bootstrap support and BI posterior probability values. Colored squares indicate family-level classification, and red squares with asterisks denote species included in this study. Families such as Malvaceae, Rosaceae, and Salicaceae showed exceptionally strong support (99.6–100, 96.1–100, and 100, respectively), confirming that mitochondrial protein-coding genes provide sufficient phylogenetic signal to resolve relationships at both family and ordinal levels.

3.8 Genomic Synteny

Synteny analysis between the target mitogenome and three reference species (*Acer*

yangbiense, *Sapindus mukorossi*, and *Xanthoceras sorbifolium*) revealed extensive genomic rearrangement despite strong sequence conservation (Fig. 8). A total of 283 collinear blocks ⩾500 bp were identified, with 180 blocks exceeding 1,000 bp and 14 large-scale segments surpassing 10,000 bp, indicating substantial retention of genic sequences despite dramatic structural reorganization. Collinear sequences exhibited high sequence identity (90.33% average), reflecting intense purifying selection on functionally critical genes. The largest collinear block measured 24,051 bp (with Sapindus mukorossi, 99.56% identity), followed by 18,741-bp and 15,168-bp blocks, demonstrating preservation of extended multi-gene syntenic regions. However, comparative visualization using ribbon diagrams revealed that while homologous sequences are highly conserved, their genomic positions, orientations, and arrangements vary dramatically among species, with syntenic blocks frequently appearing in scrambled order and orientation.

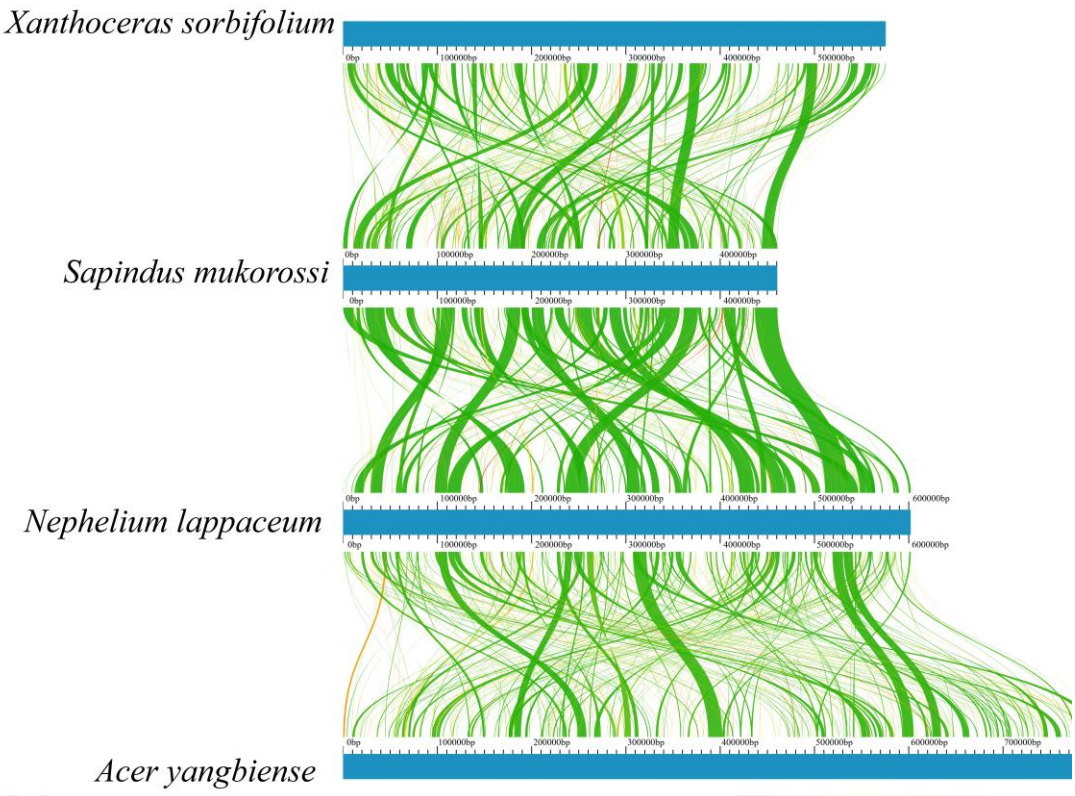


Figure 8. Synteny analysis of the rambutan (*Nephelium lappaceum*) mitogenome with three reference species (*Acer yangbiense*, *Sapindus mukorossi*, and *Xanthoceras*

sorbifolium) visualized using AliTV.

Discussion

Structural Characteristics and Gene Content

Plant mitochondria serve as central hubs for energy metabolism and cellular signaling, participating in oxidative phosphorylation, programmed cell death, cytoplasmic male sterility, and stress responses, thereby forming a highly interactive tripartite genome system with nuclear and plastid genomes(Wang et al. 2025; Zhang et al. 2026). Unlike compact and conserved animal mitogenomes, plant mitochondrial genomes exhibit remarkable diversity in size (ranging from <100 kb to several Mb), characterized by abundant repeats, AT-rich intergenic regions, and extensive foreign DNA insertions from nuclear and chloroplast genomes(Ni et al. 2025). As of February 2025, the NCBI database contains complete mitochondrial genome data for seven species in Sapindales, with genome sizes ranging from 575.6 kb (*Xanthoceras sorbifolium*) to 803.3 kb (*Acer yangbiense*), GC content from 43.98% to 45.71%, and intergenic regions comprising 92.18-95.14% of the genomes, encoding 32-45 protein-coding genes (PCGs), 3-4 rRNA genes, and 21-31 tRNA genes, demonstrating substantial genomic variability across the order(Yu et al. 2018; Bi et al. 2019; Yang et al. 2019; Martins et al. 2019; Wang et al. 2021). The complete mitochondrial genome of rambutan (*Nephelium lappaceum*) spans 460,603 bp with 45.77% GC content, higher than *Acer yangbiense* (~45.46%) and *Xanthoceras sorbifolium* (~45.71%), but lower than Sapindus mukorossi (45.84%). The genome exhibits exceptionally low gene density, with intergenic regions constituting 91.95% of total length, similar to other Sapindaceae families. Notably, rambutan exhibits the smallest mitogenome size among sequenced Sapindales species, suggesting lineage-specific genome contraction events that warrant further investigation.

Mitochondrial genomes exhibit variability, but the genes themselves are highly conserved(Ni et al. 2025). Comprehensive annotation identified 60 genes comprising 39 protein-coding genes, 3 rRNAs, and 18 tRNAs. The protein-coding repertoire encompasses 25 core mitochondrial genes universally conserved across land plants, comprehensively covering the complete respiratory chain machinery. Notably,

rambutan retains all core subunits of complexes I, III, IV, and V, highly consistent with strong energy metabolism demands observed in postharvest physiology. Previous studies demonstrated that rambutan pericarp upregulates tricarboxylic acid cycle and electron transport chain genes under energy deficiency(Zhang et al. 2023), while melatonin treatment enhances mitochondrial enzyme activities and ATP content(Wei et al. 2022). Our genomic-level confirmation of complete retention of these key energy metabolism genes provides important support for understanding the molecular foundation underlying the "energy-membrane-phenolic leakage" pathway governing postharvest browning.

Repetitive Elements

In the study of plant mitochondrial genomes, the widespread presence of repetitive sequences and the frequent recombination mediated by them are among the most striking features. These sequences play multiple crucial roles in maintaining genomic structural stability, driving genome expansion, regulating gene expression, and promoting species evolution(Davila et al. 2011; Cole et al. 2018). In this study, Repetitive element characterization identified 62 SSRs (85.48% monomeric), 34 tandem repeats (25-80 bp). Thymine homopolymers dominated monomeric SSRs (45.28%), directly reflecting AT-rich composition and indicating SSR formation primarily driven by replication slippage. Repeat element analysis using REPuter identified 243 repeat pairs (≥30 bp) comprising 111 forward (45.7%), 131 palindromic (53.9%), and 1 reverse repeat (0.4%), with lengths spanning 30–3,326 bp. Contrary to a pattern of pure tandem duplication, palindromic repeats constituted the predominant class and included the longest element detected (a 3,326 bp palindrome), suggesting that recombination-mediated rearrangements—particularly inverted-repeat-driven intramolecular recombination—rather than simple direct tandem duplications, constitute a major mechanism shaping repeat expansion and structural plasticity of the rambutan mitogenome. Compared to Trachelospermum jasminoides (42 repeats, largest 7,460 bp, recombination frequency 48-51%)(Cai et al. 2024) and Rubus idaeus (three key repeats driving eight conformations)(Zhang et al. 2025), rambutan's moderately-

sized repeats suggest intermediate structural rearrangement frequency. However, the abundance of dispersed repeats throughout extensive intergenic regions 91.95% indicates substantial potential for homologous recombination-mediated rearrangements characteristic of plant mitogenomes, highlighting the evolutionary plasticity underlying mitochondrial genome architecture in rambutan.

Codon Usage

Codon usage bias reflects the non-random utilization of synonymous codons, shaped by the interplay between mutational pressure and natural selection for translational efficiency, with plant mitochondrial genomes exhibiting unique patterns due to their AT-rich composition and organellar-specific evolutionary constraints(Iriarte et al. 2021; Zhang et al. 2024). To elucidate the forces driving codon preferences in rambutan, we performed comprehensive codon usage analysis across the mitochondrial genome. Codon usage analysis across 10,167 codons revealed pronounced A/U-ending codon preference, with GCU (alanine, RSCU = 1.56), GGA (glycine, RSCU = 1.52), and UAU (tyrosine, RSCU = 1.48) most favored, while G/C-ending codons—most notably CAG (glutamine, RSCU = 0.52), CAC (histidine, RSCU = 0.53), and GGC (glycine, RSCU = 0.54)—were systematically avoided. This pattern strongly correlates with the AT-rich composition of the coding regions (56.26% AT, GC3 = 39.28% vs. GC1 = 48.38% and GC2 = 43.54%), suggesting that mutational bias toward A/T transversions, rather than translational selection, is the predominant evolutionary force, aligning with trends in other angiosperm mitogenomes(Wang et al. 2011). This is also consistent with the codon preferences of most plants, where the number of codons ending in A/T is far greater than the number of codons ending in G/C (Xu et al. 2015).

RNA Editing and Post-transcriptional Modification

At the post-transcriptional level, plant mitochondrial protein-coding genes undergo widespread C-to-U RNA editing, which reshapes codons and restores conserved amino acid residues to maintain the functional stability of respiratory chain subunits and metabolic enzymes, playing critical roles in stress adaptation and reproductive development(Yu et al. 2025). Recent advances in programmable RNA base editing tools

have enabled precise manipulation of specific editing sites in plant mitochondria and chloroplasts, providing new avenues for engineering cytoplasmic male sterility, enhancing crop stress tolerance, and optimizing energy metabolism(Tang and Luo 2018). RNA editing analysis revealed 434 sites distributed across 34 genes (94.4%), with seven genes exhibiting high editing activity (>20 sites): nad4 (40), ccmB (37), mttB (34), nad7 (32), ccmC (29), ccmFN (27), and nad5 (21), collectively accounting for 50.7% of total events. The NADH dehydrogenase gene family showed the highest editing density (144 sites, average 18.0 per gene), while ribosomal proteins exhibited comparatively low editing activity (58 sites across 11 genes). In summary, the RNA editing sites identified in this study exhibit a clear gene preference, being highly concentrated in core genes related to oxidative phosphorylation and cytochrome c biogenesis while avoiding evolutionarily stable ribosomal proteins. This "uneven distribution" reveals the precise regulatory role of RNA editing in plant adaptive evolution.

This pattern aligns with upregulation of mitochondrial genes (LcAtpB, LcAOX1, LcAAC1, LcUCP1) under energy deficiency in lychee(Zhou et al. 2022; Zhao et al. 2024; Gao et al. 2024). High editing intensity in NADH dehydrogenase (nad4, nad5, nad7) and ATP synthase (atp4) genes likely optimizes protein sequences for electron transport efficiency, maintains multisubunit complex interactions, and compensates for constraints imposed by uniparental inheritance. The concentration in complex I genes may reflect particularly complex assembly requirements. Rambutan's 434 editing sites represent moderate levels compared to P. hunanensis (455)(Chen et al. 2024a), possibly reflecting balanced evolutionary rates in Sapindaceae.

Intragenomic Gene Transfer Drives Mitochondrial Genome Expansion and Complexity

IGT (Intragenomic Gene Transfer) is a core driving force behind the expansion and structural complexity of plant mitochondrial genomes(Alverson et al. 2010; Donnelly et al. 2017). Through the continuous acquisition of exogenous fragments, mitochondrial genomes accumulate large amounts of non-coding sequences, resulting in extremely

high diversity in their size and organizational structure(Wang et al. 2024b). Synteny analysis identified 283 collinear blocks ≥500 bp with remarkably high sequence identity (90.33%), reflecting strong purifying selection. However, genomic positions, orientations, and arrangements vary dramatically, with syntenic blocks appearing in scrambled order. This paradox—high sequence conservation coupled with extensive positional rearrangement—underscores plant mitochondrial genome evolution: sequences evolve slowly while gene order evolves rapidly through recombination between dispersed repeats and substoichiometric shifting. Compared to high-frequency recombination in T. jasminoides and R. idaeus (>45%)(Cai et al. 2024; Zhang et al. 2025), rambutan shows intermediate rearrangement, reflecting balance between stability and plasticity in woody Sapindaceae. This intermediate recombination frequency suggests that while structural rearrangements are common in rambutan, core respiratory genes remain largely positionally conserved, maintaining functional stability essential for cellular energy production across diverse environmental conditions.

Ka/Ks analysis

The ratio of nonsynonymous to synonymous substitution rates (Ka/Ks or ω) serves as a fundamental metric for inferring the mode and intensity of natural selection acting on protein-coding genes, where Ka/Ks < 1 indicates purifying selection, Ka/Ks = 1 suggests neutral evolution, and Ka/Ks > 1 reflect elevated nonsynonymous substitution that requires codon model validation before being interpreted as positive selection(Zhang 2022). In comparative mitochondrial genomics, Ka/Ks analysis provides critical insights into the evolutionary constraints shaping organellar gene function and adaptation across lineages. In the present study, we observed that the Ka/Ks ratio of the atp8, rpl5, rpl2, and atp4 genes in *Nephelium lappaceum* exhibited pronounced lineage-dependent variation across pairwise comparisons, indicating that these four genes harbor lineage-specific elevated nonsynonymous substitution signals that warrant codon model validation. Both atp4 and atp8 encode subunits of the mitochondrial F_1F_o-ATP synthase, with atp1 forming part of the catalytic F_1 head and

atp8 constituting a small, rapidly evolving F_0 membrane component(Yang et al. 2024; Chen et al. 2024b). The rpl5 gene encodes the mitochondrial ribosomal protein L5 of the large subunit, a key component of the central protuberance that participates in rRNA binding and ribosome assembly, and is essential for mitochondrial protein translation and respiratory chain function(Jung et al. 2023). The Ka/Ks analysis revealed that the vast majority of mitochondrial protein-coding genes in *Nephelium lappaceum* exhibited Ka/Ks values well below 1 across all pairwise comparisons, indicating that these genes are evolving under strong purifying selection that eliminates deleterious mutations to maintain protein structural integrity and function. atp8 consistently exhibited Ka/Ks > 1 across all three pairwise comparisons (1.11–1.61), whereas rpl5, rpl2 and atp4 exceeded this threshold in only a single comparison each, remaining under purifying selection in the majority of comparisons. Therefore, this anomaly more likely reflects lineage specific variation or analytical artefacts, and codon model validation is required before invoking positive selection. Overall, the predominant signature of purifying selection across the mitochondrial genome reflects a high degree of evolutionary conservation in N. lappaceum, providing a robust foundation for phylogenetic inference and taxonomic classification within Sapindaceae.

Comparative Genomics and Phylogenetic Insights

Plant mitochondrial genomes exhibit remarkable variation in gene content, reflecting lineage-specific gene losses, functional transfers to the nucleus, and adaptations to diverse life histories. Comparative analysis across 22 angiosperms revealed substantial gene content variation (28-39 PCGs), with 18 genes highly conserved (≥21/23 species) constituting the functional core, while ribosomal proteins and succinate dehydrogenase genes showed pronounced variability. Rambutan retains a relatively complete gene set (39 PCGs), moderately high within Sapindaceae and significantly exceeding rapidly-evolving lineages like Brassica (17 genes lost), possibly related to life history as a long-lived woody tree. This genomic stability may provide foundations for robust energy metabolism crucial for fruit development and postharvest stress responses. Furthermore, the retention of a complete respiratory gene complement likely contributes to the

extended shelf life and metabolic resilience observed in tropical fruit crops like rambutan.

Phylogenetic analysis based on 26,838 aligned nucleotide sites demonstrated robust resolution (94.7% nodes with bootstrap ≥80), confirming rambutan's sister relationship with lychee within a highly supported Sapindaceae clade, consistent with nuclear and chloroplast phylogenies(Dong et al. 2021; Zhang et al. 2021). High congruence suggests no significant cytonuclear discordance within Sapindaceae.

Conclusion

We successfully assembled the complete mitochondrial genome of rambutan (*Nephelium lappaceum*), spanning 460,603 bp with 45.77% GC content and comprising 60 genes (39 PCGs, 18 tRNAs, 3 rRNAs). Comprehensive analyses revealed complete oxidative phosphorylation core genes and relatively conservative genomic structure, potentially supporting robust energy metabolism capacity. High RNA editing intensity in NADH dehydrogenase and ATP synthase genes aligns with mitochondrial energy metabolism's central role in postharvest browning. Phylogenetic analysis confirmed rambutan's sister relationship with lychee, while synteny analysis revealed extensive genomic rearrangement despite high sequence conservation. This study provides the first complete mitogenome resource for rambutan, establishing foundations for germplasm identification, molecular breeding, and investigating mitogenome-phenotype associations in economically important Sapindaceae species.

Data availability

All sequencing data analyzed in this study were obtained from publicly available repositories. Raw sequencing reads, including 93 PacBio long-read libraries (SRA accessions: SRR15275655–SRR15275747) and one Illumina paired-end short-read library (SRR14560263), were retrieved from NCBI BioProject PRJNA728838. The complete mitochondrial genome assembled in this study has been deposited in the NCBI GenBank database (BioProject/GenBank Submission ID: SUB16264403). The accession number will be provided upon assignment and released upon publication of this article.

References

Alverson AJ, Wei X, Rice DW, Stern DB, Barry K, Palmer JD (2010) Insights into the Evolution of Mitochondrial Genome Size from Complete Sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). Mol Biol Evol 27(6):1436–1448. https://doi.org/10.1093/molbev/msq029

Ankenbrand MJ, Hohlfeld S, Hackl T, Förster F (2017) AliTV—interactive visualization of whole genome comparisons. PeerJ Comput Sci 3:e116. https://doi.org/10.7717/peerj-cs.116

Beier S, Thiel T, Münch T, Scholz U, Mascher M (2017) MISA-web: a web server for microsatellite prediction. Bioinformatics 33(16):2583–2585. https://doi.org/10.1093/bioinformatics/btx198

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27(2):573–580. https://doi.org/10.1093/nar/27.2.573

Bi Q, Cui Y, Zhao Y, Liu X, Wang L, Yu H (2019) Complete mitochondrial genome of a Chinese oil tree yellowhorn, *Xanthoceras sorbifolium* (Sapindales, Sapindaceae). Mitochondrial DNA Part B 4(1):1492–1493. https://doi.org/10.1080/23802359.2019.1601038

Cai Y, Chen H, Ni Y, Li J, Zhang J, Liu C (2024) Repeat-mediated recombination results in Complex DNA structure of the mitochondrial genome of *Trachelospermum jasminoides*. BMC Plant Biol 24(1):966. https://doi.org/10.1186/s12870-024-05568-6

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC Bioinformatics 10(1):421. https://doi.org/10.1186/1471-2105-10-421

Chen C, Wu Y, Li J, Wang X, Zeng Z, Xu J, Liu Y, Feng J, Chen H, He Y, Xia R (2023) TBtools-II: A “one for all, all for one” bioinformatics platform for biological big-data mining. Mol Plant 16(11):1733–1742. https://doi.org/10.1016/j.molp.2023.09.010

Chen L, Dong X, Huang H, Xu H, Rono PC, Cai X, Hu G (2024a) Assembly and comparative analysis of the initial complete mitochondrial genome of *Primulina hunanensis* (Gesneriaceae): a cave-dwelling endangered plant. BMC Genomics

25(1):322. https://doi.org/10.1186/s12864-024-10247-9

Chen S, Zhou Y, Chen Y, Gu J fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34(17):i884-i890. https://doi.org/10.1093/bioinformatics/bty560

Chen T, Guo B, Gao K, Zhang R-G, Jia K-H, Wu R, Li Y, Qin D, An X (2024b) Comprehensive analysis of the mitochondrial genome of Populus tomentosa 'GM107' and the dynamic expression of mitochondrial ATP synthase genes. Ind Crops Prod 221:119382. https://doi.org/10.1016/j.indcrop.2024.119382

Cole LW, Guo W, Mower JP, Palmer JD (2018) High and Variable Rates of Repeat-Mediated Mitochondrial Genome Rearrangement in a Genus of Plants. Mol Biol Evol 35(11):2773-2785. https://doi.org/10.1093/molbev/msy176

Davila JI, Arrieta-Montiel MP, Wamboldt Y, Cao J, Hagmann J, Shedge V, Xu Y-Z, Weigel D, Mackenzie SA (2011) Double-strand break repair processes drive evolution of the mitochondrial genome in Arabidopsis. BMC Biol 9(1):64. https://doi.org/10.1186/1741-7007-9-64

Dong F, Lin Z, Lin J, Ming R, Zhang W (2021) Chloroplast Genome of Rambutan and Comparative Analyses in Sapindaceae. Plants 10(2):283. https://doi.org/10.3390/plants10020283

Donnelly K, Cottrell J, Ennos RA, Vendramin GG, A'Hara S, King S, Perry A, Wachowiak W, Cavers S (2017) Reconstructing the plant mitochondrial genome for marker discovery: a case study using Pinus. Mol Ecol Resour 17(5):943–954. https://doi.org/10.1111/1755-0998.12646

Edera AA, Small I, Milone DH, Sanchez-Puerta MV (2021) Deepred-Mt: Deep representation learning for predicting C-to-U RNA editing in plant mitochondria. Comput Biol Med 136:104682. https://doi.org/10.1016/j.compbiomed.2021.104682

Gao Z, Zhao K, Zhang Z, Nizamani MM, Li S, Li M, Gong D, Wang J, Hu M (2024) Hydrogen sulfide alleviates pericarp browning in lichi fruit by modulating energy and sugar metabolisms. Front Plant Sci 15:1421203. https://doi.org/10.3389/fpls.2024.1421203

Guo L, Shi W, Luo Y, Gao K, Huang J, Wei H, Liang P, He L, Xiao D, Zhan J, Zeng G, Wang A (2025) De Novo Assembly of First Mitochondrial Genome in *Melicope*

pteleifolia (Rutaceae): Resolving Inter-Organellar Gene Transfer Events Through Integrated Chloroplast Analysis. Horticulturae 11(6):628. https://doi.org/10.3390/horticulturae11060628

Han F, Qu Y, Chen Y, Xu L, Bi C (2022) Assembly and comparative analysis of the complete mitochondrial genome of Salix wilsonii using PacBio HiFi sequencing. Front Plant Sci 13:1031769. https://doi.org/10.3389/fpls.2022.1031769

Hou Z, Wang M, Jiang Y, Xue Q, Liu W, Niu Z, Ding X (2024) Mitochondrial genome insights into the spatio-temporal distribution and genetic diversity of *Dendrobium hancockii* Rolfe (Orchidaceae). Front Plant Sci 15:1469267. https://doi.org/10.3389/fpls.2024.1469267

Huang CC, Paull RE, Wang T (2024) Litchi postharvest physiology and handling. Crop Sci 64(4):2014–2063. https://doi.org/10.1002/csc2.21274

Iriarte A, Lamolle G, Musto H (2021) Codon Usage Bias: An Endless Tale. J Mol Evol 89(9–10):589–593. https://doi.org/10.1007/s00239-021-10027-z

Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi T-S, Li DZ (2020) GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol 21(1):241. https://doi.org/10.1186/s13059-020-02154-5

Joyce EM, Appelhans MS, Buerki S, Cheek M, De Vos JM, Pirani JR, Zuntini AR, Bachelier JB, Bayly MJ, Callmander MW, Devecchi MF, Pell SK, Groppo M, Lowry PP, Mitchell J, Siniscalchi CM, Munzinger J, Orel HK, Pannell CM, Nauheimer L, Sauquet H, Weeks A, Muellner-Riehl AN, Leitch IJ, Maurin O, Forest F, Nargar K, Thiele KR, Baker WJ, Crayn DM (2023) Phylogenomic analyses of Sapindales support new family relationships, rapid Mid-Cretaceous Hothouse diversification, and heterogeneous histories of gene duplication. Front Plant Sci 14:1063174. https://doi.org/10.3389/fpls.2023.1063174

Jung L, Schleicher S, Alsaied Taha F, Takenaka M, Binder S (2023) The MITOCHONDRIAL TRANSCRIPT STABILITY FACTOR 4 (MTSF4) is essential for the accumulation of dicistronic *rpl5-cob* MRNAS in *Arabidopsis thaliana*. Plant J 113(2):375–386. https://doi.org/10.1111/tpj.16053

Kitazaki K, Kubo T (2010) Cost of Having the Largest Mitochondrial Genome:

Evolutionary Mechanism of Plant Mitochondrial Genome. J Bot 2010:1–12. https://doi.org/10.1155/2010/620137

Kurtz S (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res 29(22):4633–4642. https://doi.org/10.1093/nar/29.22.4633

Li J, Chen C, Zeng Z, Wu F, Feng J, Liu B, Mai Y, Chu X, Wei W, Li X, Liang Y, Liu Y, Xu J, Xia R (2024) SapBase: A central portal for functional and comparative genomics of Sapindaceae species. J Integr Plant Biol 66(8):1561–1570. https://doi.org/10.1111/jipb.13680

Li M, Lin H, Zhuo R, Li B, Lin Y, Chen Y, Fan Z, Wang H (2026) The changes of antioxidant system, membrane integrity and phenolics metabolism participate in bongkrekic acid and adenosine triphosphate-mediated energy status regulating longan pericarp browning. Postharvest Biol Technol 232:114001. https://doi.org/10.1016/j.postharvbio.2025.114001

Lin D, Shao B, Gao Z, Li J, Li Z, Li T, Huang W, Zhong X, Xu C, Chase MW, Jin X (2025) Phylogenomics of angiosperms based on mitochondrial genes: insights into deep node relationships. BMC Biol 23(1):45. https://doi.org/10.1186/s12915-025-02135-9

Lin Y, Lin H, Zeng L, Lin M, Chen Y, Fan Z, Wang H, Lin Y (2023) DNP and ATP regulate the breakdown occurrence in the pulp of Phomopsis longanae Chi-infected longan fruit through modulating the metabolism of membrane lipid. Food Chem 409:135330. https://doi.org/10.1016/j.foodchem.2022.135330

Luo H, Tang J, Zeng Z, Xu J, Chen B, Huang X, Yang H, Huang R, Chen C, Song C, Jin S (2025) Metabolic dynamics of litchi pericarp and pulp during browning: Unraveling differential profiles through temporal clustering and untargeted metabolomics. Food Chem Mol Sci 11:100277. https://doi.org/10.1016/j.fochms.2025.100277

Mariault L, Puginier C, Keller J, El Baidouri M, Delaux PM. (2025). Mechanisms, detection, and impact of horizontal gene transfer in plant functional evolution. Plant Cell 37(9):koaf195. https://doi.org/10.1093/plcell/koaf195

Martins G, Balbino E, Marques A, Almeida C (2019) Complete mitochondrial genomes of the Spondias tuberosa Arr. Cam and Spondias mombin L. reveal highly repetitive

DNA sequences. Gene 720:144026. https://doi.org/10.1016/j.gene.2019.144026

Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ (2015) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol 32(1):268–274. https://doi.org/10.1093/molbev/msu300

Ni Y, Li J, Tan Y, Shen G, Liu C (2025) Advance in the assembly of the plant mitochondrial genomes using high-throughput DNA sequencing data of total cellular DNAS. Plant Biotechnol J 23(11):4944–4965. https://doi.org/10.1111/pbi.70249

Ouyang L, Li X, Wang R, Chen Y, Wang S, Wang J, Tian Y (2025) Comprehensive analysis of the mitochondrial genome of *Iris domestica* emphasizing multichromosomal organization and repeat-mediated homologous recombination. Front Plant Sci 15:1520033. https://doi.org/10.3389/fpls.2024.1520033

Passafiume R, Roppolo P, Tinebra I, Pirrone A, Gaglio R, Palazzolo E, Farina V (2023) Reduction of Pericarp Browning and Microbial Spoilage on Litchi Fruits in Modified Atmosphere Packaging. Horticulturae 9(6):651. https://doi.org/10.3390/horticulturae9060651

Putintseva YA, Bondar EI, Simonov EP, Sharov VV, Oreshkova NV, Kuzmin DA, Konstantinov YM, Shmakov VN, Belkov VI, Sadovsky MG, Keech O, Krutovsky KV (2020) Siberian larch (*Larix sibirica* Ledeb.) mitochondrial genome assembled using both short and long nucleotide sequence reads is currently the largest known mitogenome. BMC Genomics 21(1):654. https://doi.org/10.1186/s12864-020-07061-4

Skippington E, Barkman TJ, Rice DW, Palmer JD (2015) Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes. Proc Natl Acad Sci 112(27). https://doi.org/10.1073/pnas.1504491112

Soe T, Kong J, Xu B, Kan J, Tembrock LR, Nie L, Jia M (2025) Mitochondrial genomes of four *Dendrobium* species (Orchidaceae): a comprehensive analysis of structural diversity, synteny, and plastid-derived sequences. Genomics Commun 2(1):0–0. https://doi.org/10.48130/gcomm-0025-0018

Sun J, Chen H, Chen Y, Lin M, Hung Y-C, Jiang Y, Lin H (2022) ε-Poly-l-Lysine Enhances Fruit Disease Resistance in Postharvest Longans (*Dimocarpus longan* Lour.)

by Modulating Energy Status and ATPase Activity. Foods 11(5):773. https://doi.org/10.3390/foods11050773

Tang W, Luo C (2018) Molecular and Functional Diversity of RNA Editing in Plant Mitochondria. Mol Biotechnol 60(12):935–945. https://doi.org/10.1007/s12033-018-0126-z

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S (2017) GeSeq-versatile and accurate annotation of organelle genomes. Nucleic Acids Res 45(W1):W6–W11. https://doi.org/10.1093/nar/gkx391

Wang B, Yuan J, Liu J, Jin L, Chen J-Q (2011) Codon Usage Bias and Determining Forces in Green Plant Mitochondrial Genomes: Codon Bias in Plant Mitochondrial Genomes. J Integr Plant Biol 53(4):324–334. https://doi.org/10.1111/j.1744-7909.2011.01033.x

Wang H, Wu Z, Li T, Zhao J (2024a) Highly active repeat-mediated recombination in the mitogenome of the aquatic grass *Hygroryza aristata*. BMC Plant Biol 24(1):644. https://doi.org/10.1186/s12870-024-05331-x

Wang J, Kan S, Liao X, Zhou J, Tembrock LR, Daniell H, Jin S, Wu Z (2024b) Plant organellar genomes: much done, much more to do. Trends Plant Sci 29(7):754–769. https://doi.org/10.1016/j.tplants.2023.12.014

Wang T, He H, Sun J, Man J, Li J, Sun C, Zhao Y, Kuang Z, Liu Y, Huang Z, Zhao Y (2025) From genome assembly to functional interpretation: an epistemological shift in understanding plant mitochondrial architecture. Front Plant Sci 16:1723739. https://doi.org/10.3389/fpls.2025.1723739

Wang X, Zhang R, Yun Q, Xu Y, Zhao G, Liu J, Shi S, Chen Z, Jia L (2021) Comprehensive analysis of complete mitochondrial genome of Sapindus mukorossi Gaertn.: an important industrial oil tree species in China. Ind Crops Prod 174:114210. https://doi.org/10.1016/j.indcrop.2021.114210

Wei D, Yang J, Xiang Y, Meng L, Pan Y, Zhang Z (2022) Attenuation of postharvest browning in rambutan fruit by melatonin is associated with inhibition of phenolics oxidation and reinforcement of antioxidative process. Front Nutr 9:905006. https://doi.org/10.3389/fnut.2022.905006

Wick RR, Schultz MB, Zobel J, Holt KE (2015) Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31(20):3350–3352. https://doi.org/10.1093/bioinformatics/btv383

Xu W, Xing T, Zhao M, Yin X, Xia G, Wang M (2015) Synonymous codon usage bias in plant mitochondrial genes is associated with intron number and mirrors species evolution. Plos One 10(6):e0131508. https://doi.org/10.1371/journal.pone.0131508

Yang F, Vincis Pereira Sanglard L, Lee C-P, Ströher E, Singh S, Oh GGK, Millar AH, Small I, Colas Des Francs-Small C (2024) Mitochondrial *atp1* mRNA knockdown by a custom-designed pentatricopeptide repeat protein alters ATP synthase. Plant Physiol 194(4):2631–2647. https://doi.org/10.1093/plphys/kiae008

Yang J, Wariss HM, Tao L, Zhang R, Yun Q, Hollingsworth P, Dao Z, Luo G, Guo H, Ma Y, Sun W (2019) De novo genome assembly of the endangered *Acer yangbiense*, a plant species with extremely small populations endemic to Yunnan Province, China. GigaScience 8(7):giz085. https://doi.org/10.1093/gigascience/giz085

Yong Y, Hu S, Zhong M, Wen Y, Zhou Y, Ma R, Jiang X, Zhang Q (2024) Horizontal gene transfer from chloroplast to mitochondria of seagrasses in the yellow–Bohai seas. Genomics 116(5):110940. https://doi.org/10.1016/j.ygeno.2024.110940

Yu F, Bi C, Wang X, Qian X, Ye N (2018) The complete mitochondrial genome of Citrus sinensis. Mitochondrial DNA Part B 3(2):592–593. https://doi.org/10.1080/23802359.2018.1473738

Yu R, Liu L, Jost M, Zhao R, Wanke S, Jiao Y (2025) Evolution of mitochondrial RNA editing sites and stop codon-lacking transcripts in angiosperms. Commun Biol 8(1):977. https://doi.org/10.1038/s42003-025-08418-9

Zhang H, Yan M, Li L, Jiang Z, Xiong Y, Wang Y, Akogwu CO, Tolulope OM, Zhou H, Sun Y, Wang H (2025) Assembly and comparative analysis of the complete mitochondrial genome of red raspberry (*Rubus idaeus* L.) revealing repeat-mediated recombination and gene transfer. BMC Plant Biol 25(1):85. https://doi.org/10.1186/s12870-024-05969-7

Zhang R, Jiang F, He X, Yuan Z, Chen P, Zheng Z (2023) Transcriptome profiling of *Nephelium lappaceum*: Sodium nitroprusside-induced delays of postharvest browning.

Physiol Plant 175(3):e13932. https://doi.org/10.1111/ppl.13932

Zhang W, Lin J, Li J, Zheng S, Zhang X, Chen S, Ma X, Dong F, Jia H, Xu X, Yang Z, Ma P, Deng F, Deng B, Huang Y, Li Z, Lv X, Ma Y, Liao Z, Lin Z, Lin J, Zhang S, Matsumoto T, Xia R, Zhang J, Ming R (2021) Rambutan genome revealed gene networks for spine formation and aril development. Plant J 108(4):1037–1052. https://doi.org/10.1111/tpj.15491

Zhang Y, Li P, Yi B, He D (2026) Current advances in plant mitochondria: Application revolution of cytoplasmic male sterility. New Crops 3:100081. https://doi.org/10.1016/j.ncrops.2025.100081

Zhang Y, Ma Y, Yu H, Han Y, Yu T (2024) Deciphering Codon Usage Patterns in the Mitochondrial Genome of the Oryza Species. Agronomy 14(11):2722. https://doi.org/10.3390/agronomy14112722

Zhang Z (2022) KaKs_Calculator 3.0: Calculating Selective Pressure on Coding and Non-Coding Sequences. Genomics Proteomics Bioinformatics 20(3):536–540. https://doi.org/10.1016/j.gpb.2021.12.002

Zhao K, Gao Z, Nizamani MM, Hu M, Li M, Li X, Wang J (2024) Mechanisms of Litchi Response to Postharvest Energy Deficiency via Energy and Sugar Metabolisms. Foods 13(14):2288. https://doi.org/10.3390/foods13142288

Zhou Y, Li Z, Zhu H, Jiang Y, Jiang G, Qu H (2022) Energy homeostasis mediated by the LCSNRK1A-LCBZIP1/3 signaling pathway modulates litchi fruit senescence. Plant J 111(3):698–712. https://doi.org/10.1111/tpj.15845

Funding

This work was supported by the Central Public-interest Scientific Institution Basal Research Fund (Grant NO. 1630032025024).

Author information

Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences/Key Laboratory of Crop Gene Resources and Germplasm Enhancement in Southern China, Ministry of Agriculture and Rural Affairs, Haikou, 571101, China

Jun Wang

Key Laboratory of Agro-Forestry Environmental Processes and Ecological Regulation of Hainan Province, Hainan University, Haikou, 570228, China

Dan Feng & Jun Wang

Key Laboratory of Quality Regulation of Tropical Horticultural Crop in Hainan Province, School of Tropical Agriculture and Forestry, Hainan University, Haikou, 570228, China

Dong Yu, Hailin Qiu & Wuqiang Ma

Sanya Institute of Breeding and Multiplication, Hainan University, Sanya, 572025, China

Dong Yu, Hailin Qiu & Wuqiang Ma

Contributions

JW and WQM designed the project. DF, DY and HLQ analyzed data. JW and DF wrote the manuscript. All authors read and approved the final manuscript.

Corresponding authors

Correspondence to Wuqiang Ma or Jun Wang.

Ethics declarations

Conflict of interest

The authors declare that they have no conflict of interest.

Additional information

Communicated by Jun Wang.

Publisher's Note

Supplementary Information

Not applicable. All data supporting the findings of this study are presented within the main manuscript.