BMC Plant Biology

RESEARCH Open Access

Complete mitochondrial genome of *Houttuynia cordata* sheds light on the evolution of complex structural features in early angiosperms

Yahui Huang[1], Bingyue Zhu[1], Siqi Hu[1], Wenbo Shi[1], Yuxue Liu[1], Qianhui Lin[1], Shuo Wang[1], Jing Yang[1*] and Chao Shi[1*]

Abstract

Background *Houttuynia cordata* is a well-known medicinal and edible plant with important economic value. It belongs to the magnoliid clade, a basal lineage of angiosperms. Currently, the limited number of reported mitochondrial genomes from magnoliids severely restricts our understanding of early mitochondrial genome evolution. To address this gap, we present the complete mitochondrial genome of *H. cordata*.

Results The *H. cordata* mitochondrial genome spans 534,194 bp, and predominantly adopts a bicyclic structure among four possible configurations. It encodes 45 protein-coding genes (PCGs) (40 unique), 29 transfer RNAs (21 unique), and 3 ribosomal RNAs. Repeat analysis revealed significant differences in tandem repeats, dispersed repeats, and simple sequence repeat (SSR) types and densities between the two chromosomes, with Chromosome 1 exhibiting higher repeat density and diversity. Additionally, the *H. cordata* mitochondrial genome harbors 18 potential mitochondrial plastid DNAs (MTPTs), most of which originate from the inverted repeats (IR) regions of the plastid genome. Variant detection at the individual-level revealed significant heteroplasmy across most regions, while large repeat regions and MTPTs regions exhibited relative conservation. Phylogenetic and collinearity analyses further indicated substantial differences between the mitochondrial genome structure of *H. cordata* and 15 other magnoliid species, while conserved regions were maintained.

Conclusion This study elucidates the unique evolutionary characteristics of the *H. cordata* mitochondrial genome and provides valuable insights into its genetics and evolution. Comparative analysis with other magnoliid species reveals both dynamic evolutionary changes and the conservation of mitochondrial genome structure and gene content in early angiosperms.

Keywords *Houttuynia cordata*, Mitochondrial genome, Homologous recombination, Heteroplasmy, Evolutionary analysis

*Correspondence:
Jing Yang
yangjing@qust.edu.cn
Chao Shi
chaoshi@qust.edu.cn
[1]College of Biological Engineering, Qingdao University of Science and Technology, Qingdao 266042, China

Introduction

Houttuynia cordata Thunb. (*H. cordata*), a perennial herbaceous plant in the family Saururaceae (Piperales, magnoliids), is widely distributed across China, Japan, Korea, Southeast Asia, and the Himalayan region, thriving in shady and humid habitats [1, 2]. *H. cordata* is well known for its dual role as a medicinal and edible plant, with a long history of traditional use spanning centuries. Its medicinal properties are well documented, particularlyin the treatment of respiratory diseases, such as pneumonia and lung abscesses [3]. Recent studiess have identified a diverse array of bioactive compounds, including alkaloids, essential oils, phenolic acids, and flavonoids, which contribute to its antimicrobial, antiallergic, anti-inflammatory, and antioxidant activities [4]. Additionally, *H. cordata* has shown notable therapeutic potential against epidemic pneumonia, with significant inhibitory effects on the SARS virus [5, 6] and COVID-19 [7]. Beyond its medicinal value, *H. cordata* is cultivated as a traditional vegetable crop, with its tender stems, leaves, and rhizomes serving as key ingredients in the cuisines of several Chinese provinces (e.g., Hubei, Hunan, Guizhou, Sichuan) and northeastern India. Despite extensive research on the physiological and biochemical traits of *H. cordata* mitochondrial genomic studies remain limited.

The plant mitochondrion, a semi-autonomous organelle with maternally inherited genetic material, is crucial for cellular energy production by generating adenosine triphosphate (ATP) for various biological processes [8, 9]. According to the endosymbiotic theory, mitochondria originated from an α-proteobacterium [10, 11]. Unlike the relatively conserved circular structure of animal mitochondrial and plant chloroplast genomes, plant mitochondrial genomes exhibit remarkable structural diversity, existing in circular, linear, branched, or mixed forms [12]. These genomes are often multipartite, with multiple chromosomes undergoing homologous recombination, leading to heterogeneous mitochondrial genome populations [12–14]. The size of plant mitochondrial genomes varies dramatically across species, ranging from 66 kb in *Viscum scurruloideum* [15] to 12 Mb in *Larix sibirica* [16]. This variation is primarily attributed to extensive non-coding sequences of unknown function, with coding regions accounting for only 1–10% of the genome [17]. Despite slow sequence evolution, plant mitochondrial genomes undergo rapid structural rearrangements. For example, in *Brassica* and *Raphanus* species, the point mutation rate of mitochondrial DNA is approximately four times slower than that of chloroplast DNA in land plants and 100 times slower than that of animal mitochondrial DNA. Conversely, the rearrangement rate of plant mitochondrial DNA is significantly higher than that of both chloroplast DNA and animal mitochondrial DNA [18]. Post-transcriptional RNA editing, which is widespread in plant mitochondria, may regulate gene expression and increase protein complexity and diversity [19]. These unique features make plant mitochondrial genomes invaluable for species identification, evolutionary studies, and trait inheritance research. For instance, mitochondrial genome analysis has been used to identify species within the genus *Calypogeia* [20], trace hybridization events in *Pinus* [21], and study cytoplasmic male sterility in crops such as wheat and rice [22, 23].

The assembly of plant mitochondrial genomes remains challenging due to factors such as mitochondrial plastid DNAs (MTPTs), nuclear mitochondrial DNAs (NUMTs), long repeats, and frequent genomic recombination events [17]. The advent of HiFi sequencing technology, offering both long read lengths and high accuracy, has helped overcome many of these obstacles [24]. This technology has facilitated the characterization of multipartite genome structures and enabled advanced comparative studies of plant mitochondrial architecture and its evolutionary implications. Although the whole genome [25, 26] and complete chloroplast genome [27] of *H. cordata* have been reported, its mitochondrial genome remains poorly characterized, leaving significant gaps in our understanding of mitochondrial functions in this species.

In this study, we assembled the first complete mitochondrial genome of *H. cordata* using HiFi sequencing data. We characterized the predominant mitochondrial genome structure and performed comprehensive annotations and analyses, including gene content, codon usage preferences, repetitive sequences, RNA editing, MTPTs, and phylogenetic relationships. Using complementary Illumina sequencing data, we identified sequence variants across the mitochondrial genome, revealing significant heteroplasmy in most genic regions while detecting conservation in specific repetitive elements. This study provides a foundation for understanding the evolutionary dynamics of the *H. cordata* mitochondrial genome and contributes to broader insights into evolution of mitochondrial genomes in early angiosperms.

Materials and methods

Plant materials and DNA sequencing

Fresh samples of *H. cordata* were collected from Xishan District, Kunming, Yunnan Province, China (24.97° N, 102.63° E). A voucher specimen (QUST20240816) was identified and deposited in the Qingdao University of Science and Technology (QUST) by Yahui Huang and Siqi Hu. The samples were snap-frozen with liquid nitrogen and then stored at − 80 °C. Genomic DNA was extracted from fresh leaf tissue using the CTAB method [28]. DNA quality was assessed via agarose gel electrophoresis to confirm purity and integrity. The SMRTbell library was prepared using the SMRTbell Express Template Prep Kit v2.0 following the manufacturer's instructions. Fragment

distribution was assessed using Femto Pulse to confirm that the majority of fragments exceeded 10 kb. SMRTbell libraries were sequenced on the PacBio Sequel IIe platform of Novogene (Beijing, China). Paired-end libraries of approximately 500 bp for DNA inserts were created using an Illumina TruSeq library preparation kit (Illumina, San Diego, CA, USA). The libraries were sequenced on the Illumina HiSeq 6000 platform at Novogene (Beijing, China). Finally, we obtained approximately 9.35 Gb of HiFi reads and 28 Gb of 2 × 150 bp paired-end raw reads.

Genome assembly and annotation

To obtain the complete mitochondrial genome assembly of *H. cordata*, we performed Flye v2.9.2 [29] by using HiFi sequencing reads, with the default parameters. Mitochondrial genome contigs were identified via BLAST search in Bandage v0.8.1 [30], with the mitochondrial genome of *Saururus chinensis* (NCBI accession numbers: OQ539548–50) serving as the reference. For acquiring the plastid genome of *H. cordata*, we conducted GetOrganelle v1.7.7.1 [31] by using Illumina reads, with the parameters "-R 15 -k 21,45,65,85,105 -F embplant_pt".

The mitochondrial genome of *H. cordata* were annotated using the IPMGA [32] online tools (http://www.1kmpg.cn/ipmga/). The transfer RNA (tRNA) genes were predicted with the tRNAscan-SE v2.0.11 [33], and the ribosomal RNA (rRNA) genes were conducted identified using the BLASTn v2.13.0 [34]. We further manually adjusted the annotation results through Apollo v2.7.0 [35], with *S. chinensis* (OQ539548–50) and *Aristolochia fimbriata* (OP649454–6) as references. Then, the plastid genomes were annotated using GeSeq v2.03 [36]. The visualization of the mitochondrial genome of *H. cordata* was performed using the PMGmap [37] online tools (http://47.96.249.172:16086/home/).

Mitochondrial genome structure analysis

To analyze the structure of the mitochondrial genome of *H. cordata*, the long reads were aligned to genome contigs using minimap2 v2.24 [38]. Reads associated with mitochondrial sequences were extracted with Samtools v1.20 [39]. Bandage v0.8.1 [30] was then utilized to identify contig coverage through its "querypaths" function. We used the custom script to count entries (≥ 95% identity), with a particular focus on analyzing multi-link contigs in the assembly graph.

Analyses of RSCU and repeated sequences

The codon usage differs in various species. To assess codon usage bias in *H. cordata*, we performed relative synonymous codon usage (RSCU) analysis. First, we utilized the Biopython v1.78 [40] to extract the protein-coding sequences of the mitochondrial genome of *H. cordata*. Subsequently, we employed the CAI v1.0.3 [41] module to calculate the RSCU values.

To identify forward, reverse, palindromic, and complementary repeats in the *H. cordata* mitochondrial genome, we conducted the online tools REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer), with the following parameters: Hamming distance to 0, minimum repeat size to 30 bp, and e-value cutoff to 1e-5. Tandem repeats in the *H. cordata* mitochondrial genome were identified using the online website Tandem Repeats Finder (https://tandem.bu.edu/trf/basic_submit). In addition, to identify simple sequence repeats (SSRs), we used the online tools MISA (https://webblast.ipk-gatersleben.de/misa/). We employed the R package ggplot2 [42] to draw the bar plots.

RNA-editing sites prediction and collinearity analysis

RNA editing sites within *H. cordata* mitochondrial genome protein-coding genes (PCGs) were determined through the online tools PREPACT3 v3.12.0 (http://www.prepact.de/), using the *Amborella trichopoda* mitochondrial genome as a reference. To increase the prediction accuracy, a cut-off value was sensibly set at 0.2 [43].

To compare the differences between the structure of *H. cordata* mitochondrial genomes and that of two other Piperales species (*S. chinensis* and *A. fimbriata*), collinearity analyses were firstly performed using the GetTwoGenomeSyn.pl script in the software package NGenomeSyn v1.41 [44]. Alignments were generated with MUMmer v4.0.0 [45] as the aligner, with a minimum collinear block size threshold of 500 bp to ensure alignment stringency. We then employed the main program in NGenomeSyn by using the generated files to finish the visualization.

Mitochondrial plastid DNAs and nuclear mitochondrial DNAs analyses

Sequence similarity comparisons between plastids and mitochondria of *H. cordata* were executed using BLASTn v2.13.0 (e-value < 1e-5) [34]. For analyzing the greater conservatism, we filtered the sequence similarities > 95% as MTPTs. To identify putative NUMTs, the complete genome sequence of *H. cordata* (GenBank accession: CM077502–15) was retrieved from the GenBank database and subsequently analyzed using NUMTfinder v0.5.5 [46]. Stringent alignment criteria were implemented to reduce false positives from short nuclear repeats: BLASTn parameters included an E-value cutoff of 1e-10 and a minimum local alignment length of 100 bp. We performed the visualisation using the "Advance Circos" module in TBtools v2.069 [47].

Evaluation of heteroplasmy in the mitochondrial genome

To quantify the heteroplasmy of the *H. cordata* mitochondrial genome, we utilized BWA-MEM v0.7.17 [48] to map Illumina reads to the mitochondrial genome assembly. We employed the software Samtools v1.20 [39] to perform the quality check, filter the short reads and remove the duplicates. Variant calling was implemented through VarDict v1.82 [49] with parameters -X 5 -f 0.05 to detect single nucleotide variants (SNVs), insertions/deletions (Indels), and complex variants. The complex variant refers to a special variant in which nearby indels (within 10 bp) or mismatches (within 5 bp) in the same read are combined. To minimize false positives, only variants with an allele frequency > 0.05 (SNVs, Indels, and complex variants) were retained. To illustrate the distribution of each variant in mitochondrial genome assembly, the visualizations were performed using TBtools v2.069 [47].

Phylogenetic analysis

Based on the sequences of PCGs from the mitochondrial genome of *H. cordata*, we inferred the phylogenetic relationships between *H. cordata* and 14 closely related species (magnoliids). Firstly, mitochondrial genome sequences for 14 magnoliid species and two Chloranthales species (*Chloranthus spicatus* and *Hedyosmum orientale*) were retrieved from GenBank (Table S1). Chloranthales was selected as the sister group to root the phylogenetic tree. Next, the shared mitochondrial PCGs from all 17 species were extracted. The analyzed genes include *atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFc*, *ccmFn*, *cob*, *cox1*, *cox2*, *cox3*, *mttB*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*, *rpl16*, *rps1*, *rps10*, *rps11*, *rps12*, *rps13*, *rps14*, *rps19*, *rps3*, *rps4*, *rps7*, and *sdh4*. These genes were aligned with MAFFT v7.407 [50]. Maximum likelihood (ML) analysis was conducted using IQ-TREE v2.2.2.7 [51], with the GTR + F + R2 model selected based on the Bayesian Information Criterion (BIC). Bootstrap support values for each branch in the phylogenetic tree were calculated based on 1,000 replicates. Bayesian phylogenetic analysis was conducted using MrBayes v3.2.7 [52], with the GTR + I + G substitution model selected via jModelTest v2.1.10 [53] under the BIC. Posterior probabilities for branch support were estimated from two independent Markov chains run for 2 million generations, sampled every 1,000 iterations, and discarding the initial 25% of trees as burn-in. Consensus trees were visualized in iTOL v6 (https://itol.embl.de/).

Result

Multiple structures in *H. cordat* a mitochondrial genome

We initially identified nine contigs associated with the mitochondrial genome. These contigs exhibited an average depth of 346.2x and a total length of 526,458 bp (Table S2). Based on the features of the assembly graph and the previously proposed structural models of the minimum master circle (MMC) and minimum secondary circle (MSC) [54], the genome was resolved into four structural configurations: MMC1, MMC2, MSC1-2, and MSC3-4 (Fig. 1). To validate these structures, HiFi reads were mapped to contigs followed by mitochondrial-derived read extraction and contig connection using Bandage v0.8.1. We then counted the number of reads spanning regions R1, ctg1, and R2, that were linked to the contigs at both ends. The results indicated that the proportions of reads supporting these four structures were 29.9%, 2.6%, 63.6%, and 3.9%, respectively (Fig. 1, Table S3). These results suggest that the *H. cordata* mitochondrial genome likely adopts multiple structures mediated by homologous recombination between repetitive regions R1 and R2. Among these, MSC1-2 was the predominant form, while MMC1 was a suboptimal structure, and the other two structures were less prevalent (Fig. 1).

General feature of *H. cordata* the mitochondrial genome

The total length of the MSC1-2 of the *H. cordata* mitochondrial genome was measured 534,194 bp, exhibiting a GC content of 46.4%. The protein-coding regions encompassed 37,311 bp, which constitutes approximately 6.98% of the genome's total length. For chromosomes 1 and 2, the sizes were 340,739 bp and 193,455 bp, with corresponding GC contents of 46.5% and 46.3%, respectively. The protein-coding regions for these chromosomes were 20,673 bp and 16,638 bp, representing 6.07% and 8.60% of their total lengths, respectively (Table 1). The mitochondrial genome contains 45 PCGs (40 unique), 29 transfer RNAs (21 unique), and 3 ribosomal RNAs. Key genes include those involved in ATP synthesis, cytochrome c biogenesis, NADH dehydrogenase, and ribosomal proteins (Fig. 2, Table 2). Notably, exons of *nad1* and *nad5* are distributed across both chromosomes, indicating trans-chromosomal gene organization in this mitochondrial genome (Fig. 2, Fig. S1).

Analyses of RSCU and repeated sequences

Analysis of relative synonymous codon usage (RSCU) in the *H. cordata* mitochondrial genome identified 26 codons with significant usage bias (RSCU > 1.1). Among these codons, GCU (Ala), CAA (Gln), and UAU (Tyr) exhibited particularly high preferences, with RSCU values greater than 1.5. Notably, all 26 codons, except for UUG (Leu), end in either A or U (Fig. 3). This finding suggested a preference for A-ended or U-ended codons in the *H. cordata* mitochondrial genome.

Tandem repeat analysis identified 19 repeats on Chromosome 1 and 12 on Chromosome 2. The percentage of sequence matches ranged from 68 to 100% and lengths

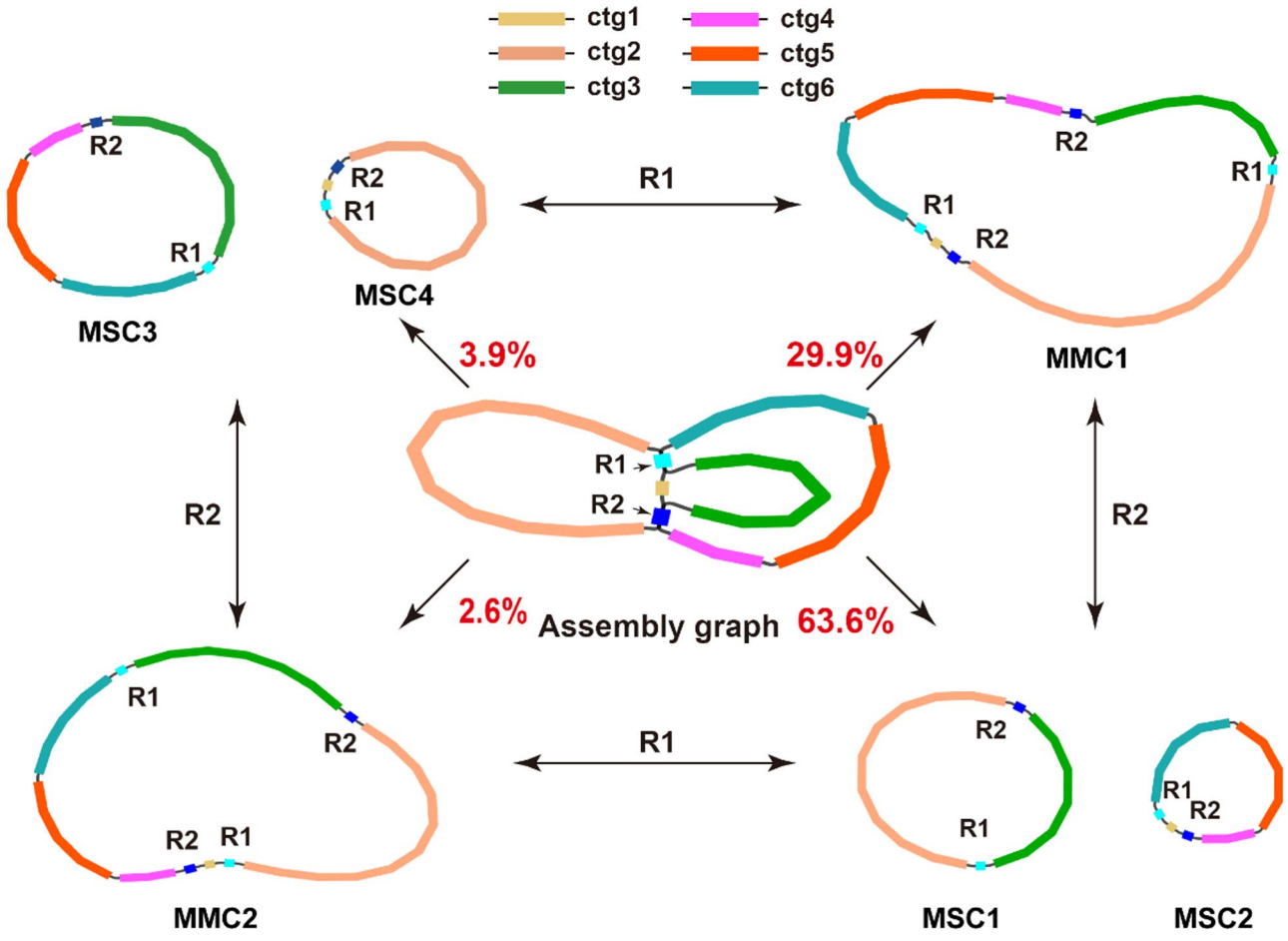

Fig. 1 *H. cordata* assembly graph and schematic diagram of repeat-mediated multiple structures. The middle one is the assembly graph, pointing in four directions to represent that it can be disassembled into four structures. The double arrows and text on the outside represent the recombination that occurs mediated by R1 or R2. Red text represents the percentage of this structure within the species. The varied colors denote distinct contigs, as detailed in the legend above

Table 1 The basic genomic features of the *H. cordata* mitochondrial genome

Genome features	Chromosome 1	Chromosome 2	Chromosome 1–2
Length	340,739	193,455	534,194
Total GC content (%)	46.5	46.3	46.4
PCG size (bp)	20,673	16,638	37,311
PCG proportion (%)	6.07	8.60	6.98
Number of genes [unique]	55 [45]	24 [23]	77 [64]
Protein genes [unique]	29 [26]	18 [18]	45 [40]
tRNA genes [unique]	23 [16]	6 [5]	29 [21]
rRNA genes [unique]	3 [3]	0 [0]	3 [3]
Accession numbers in GenBank	PQ214805	PQ214806	PQ214805-PQ214806

spanning 25–101 bp (Fig. 4A, Additional file 1). Additionally, dispersed repeat analysis identified 402 repeat elements in Chromosome 1 compared to 93 in Chromosome 2, indicating a higher density in the former. Specifically, Chromosome 1 comprised 206 forward repeats and 196 palindromic repeats, while Chromosome 2 contained 51 and 42, respectively. Notably, neither chromosome displayed reverse and complementary repeats. SSR analyses identified 137 and 64 microsatellite sequences on Chromosome 1 and 2, respectively. All six types of SSRs—monomeric, dimeric, trimeric, tetrameric, pentameric, and hexameric repeats—were identified on both chromosomes (Fig. 4B, Additional file 1). The distribution of the six types of SSRs across the two chromosomes was as follows: 58, 20, 13, 43, 2, 1 for Chromosome 1 and 34, 8, 2, 16, 3, 1 for Chromosome 2. Interestingly, both chromosomes were predominantly composed of monomers and tetramers, which collectively accounted for approximately 73.4% and 78.1% of the total SSRs, respectively.

RNA-editing sites analysis

RNA editing analysis identified 35 protein-coding genes (PCGs) with editing sites among the 40 unique PCGs in the *H. cordata* mitochondrial genome, revealing 510 total editing events. Significantly, no editing sites

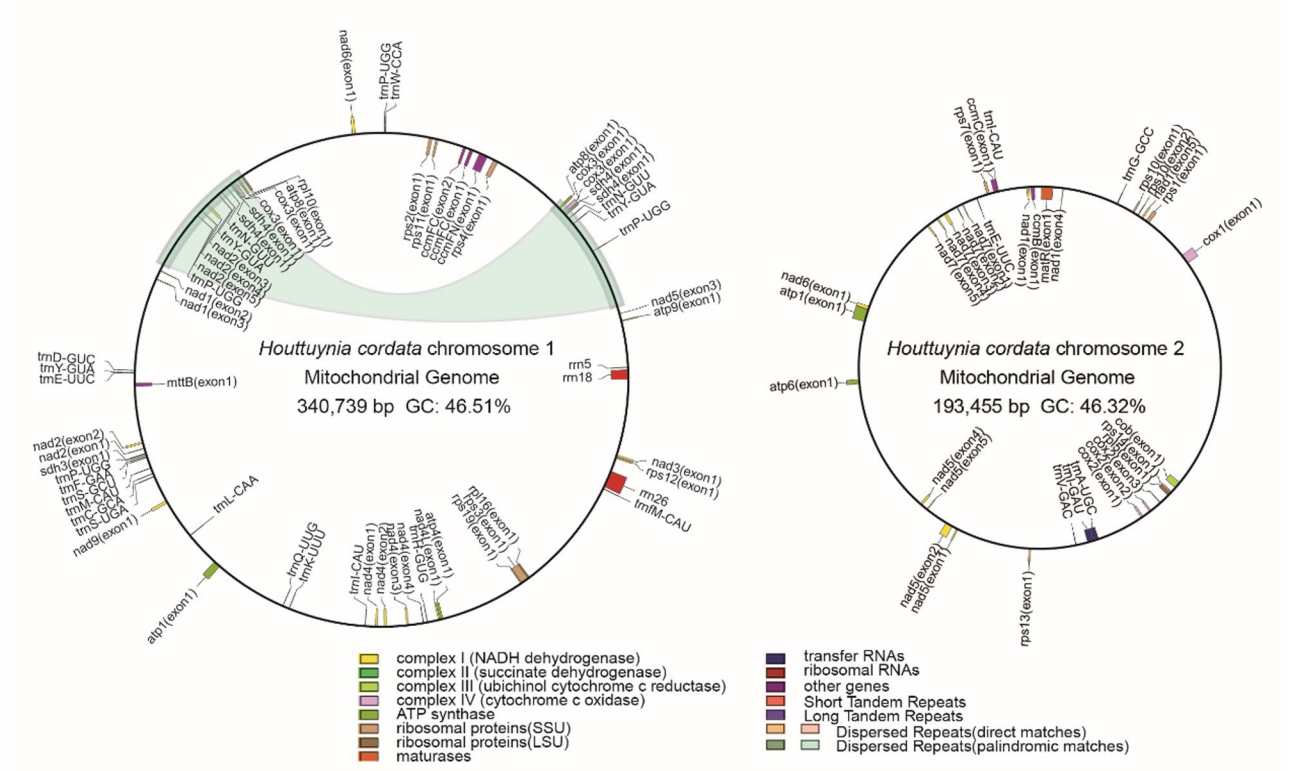

Fig. 2 Schematic of the *H. cordata* mitochondrial genome. Genes depicted on the inside represent those on the negative strand, while those on the outside are on the positive strand. The varied colors denote distinct functional categories, as detailed in the legend below

Table 2 The coding genes of the *H. cordata* mitochondrial genome

Group of genes	Gene name
ATP synthase	*atp1(2) atp4 atp6 atp8(2) atp9*
Cytohrome c biogenesis	*ccmB ccmC ccmFC* ccmFN*
Ubichinol cytochrome c reductase	*cob*
Cytochrome c oxidase	*cox1 cox2** cox3(2)*
Maturases	*matR*
Transport membrance protein	*mttB*
NADH dehydrogenase	*nad1**** nad2**** nad3 nad4*** nad4L nad5**** nad6(2) nad7**** nad9*
Ribosomal proteins (LSU)	*rpl10 rpl16 rpl5*
Ribosomal proteins (SSU)	*rps1 rps10* rps11 rps12 rps13 rps14 rps19 rps2 rps3 rps4 rps7*
Succinate dehydrogenase	*sdh3 sdh4(2)*
Ribosomal RNAs	*rrn18 rrn26 rrn5*
Transfer RNAs	*trnA-UGC* trnC-GCA trnD-GUC trnE-UUC(2) trnF-GAA trnG-GCC trnH-GUG trnI-CAU(2) trnI-GAU* trnK-UUU trnL-CAA trnM-CAU trnN-GUU(2) trnP-UGG(4) trnQ-UUG trnS-GCU trnS-UGA trnV-GAC trnW-CCA trnY-GUA(3) trnfM-CAU*

Notes: *:intron number; #Gene: Pseudo gene; Gene(2):Number of copies of multi-copy genes

were predicted in *atp1*, *atp9*, *rps2*, or *rps3*, while *nad4* contained the highest number of edits ($n = 53$). All the detected RNA editing events involved C-to-U (cytosine to uracil) transitions. Furthermore, in *nad3*, *nad7*, *rps19*, *nad4*, *nad6*, *ccmFN*, *cob*, *nad5*, and *matR*, dual editing events occurred within single codons, modifying both first and second nucleotide positions to convert proline (P) codons to phenylalanine (F) (Fig. 5, Table S4). Crucially, RNA editing converted the *nad1* start codon from ACG to ATG and introduced stop codons in *ccmFC* and *ccmFN*.

Intracellular gene transfer

Intracellular gene transfer events between organellar and nuclear genomes predominantly occur from organelle genomes to the nuclear genome or from nuclear and plastid genomes to the mitochondrial genome [55, 56]. BLASTn analysis comparing mitochondrial chromosomes with the plastid genome identified 11 and 7 mitochondrial plastid DNAs (MTPTs) on Chromosomes 1 and 2, respectively (Fig. 6A, Table S5). Six fragments from each chromosome originated from the plastid IR regions (IRA and IRB). The longest fragment on Chromosome 1 was 5,687 bp, with three fragments exceeding 500 bp, while the longest fragment on chromosome 2 was 9,630 bp, with two fragments were longer than

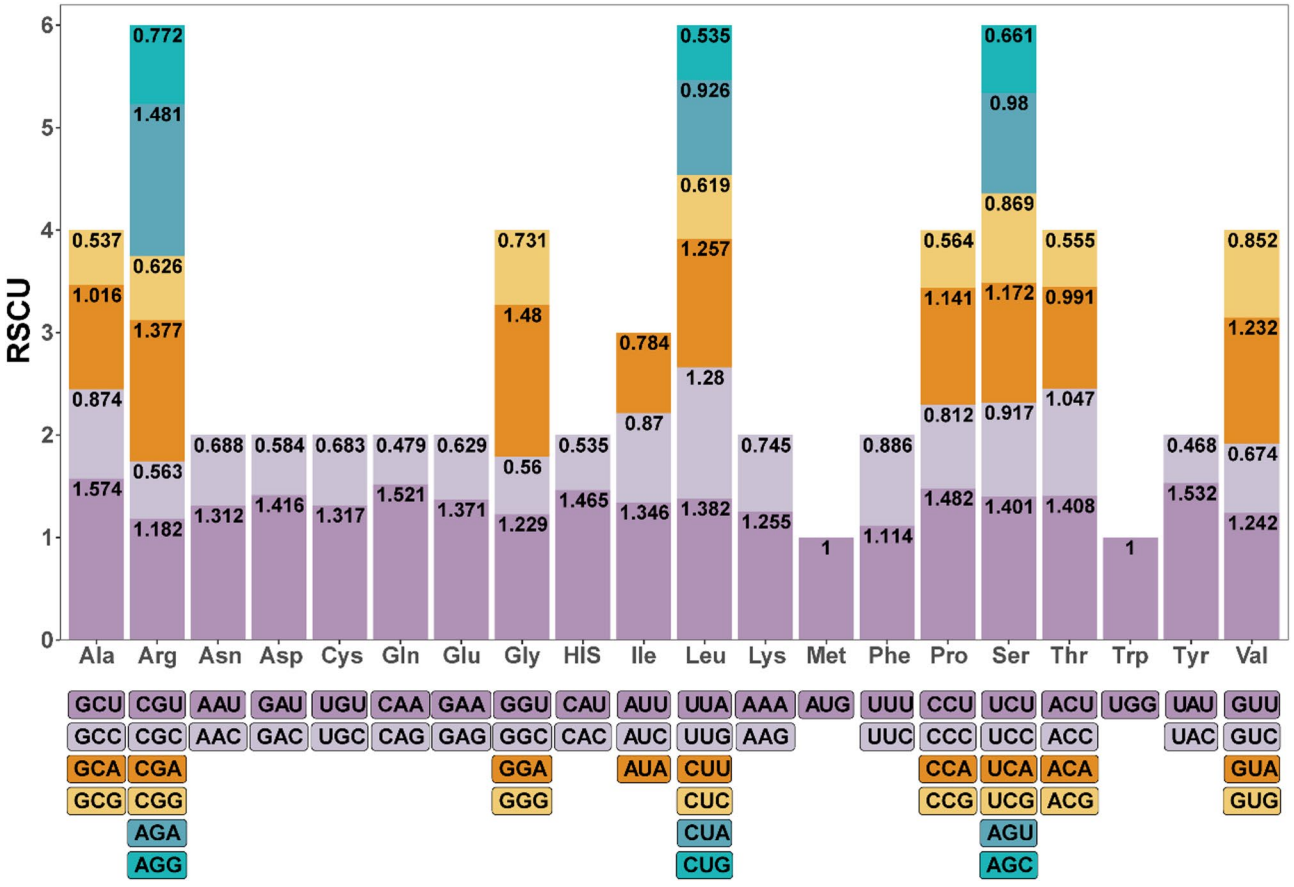

Fig. 3 Analysis of codon preference in the *H. cordata* mitochondrial genome. The chart illustrates the Relative Synonymous Codon Usage (RSCU) values for each codon in relation to their designated amino acids. Distinct bars correspond to specific codons for every amino acid, with color differentiation enabling prompt codon identification. An RSCU value exceeding 1 indicates preferential codon usage above its synonymous counterparts, while a value below 1 implies a lesser frequency of usage

500 bp. Total MTPTs length reached 20,029 bp (3.75% of the mitochondrial genome), distributed as 9,047 bp on Chromosome 1 and 10,982 bp on Chromosome 2. Most MTPTs originated from the plastid IR regions, accounting for 84.90% of the total MTPTs length. All MTPTs included nine tRNAs (*trnL-CAA*, *trnW-CCA*, *trnP-UGG*, *trnD-GUC*, *trnH-GUG*, *trnN-GUU* from chromosome 1; *trnV-GAC*, *trnI-GAU*, *trnI-CAU* from chromosome 2), four PCGs (*psbL*, *psbF*, *psbE* from Chromosome 1; rps7 from Chromosome 2), and three rRNAs (*rrn4.5*, *rrn5*, *rrn16* from Chromosome 2). In eukaryotes, mitochondrial genome DNA (mtDNA) can be translocated to the nucleus and forms nuclear mitochondrial DNAs (NUMTs) [57]. Our analysis identified 4,472 mtDNA-derived sequences homologous to the nuclear genome in *H. cordata* (Fig. 6B, Additional File 2), with integration lengths ranging 100–7,735 bp. Annotation of the NUMTs revealed that the NUMTs collection contained only one intact protein-coding gene (*nad6*), present in NUMT3089 and NUMT4380. Fifteen fragments contained complete tRNA genes, including *trnD-GUC* (NUMT963, NUMT1736, NUMT1739), *trnG-GCC* (NUMT3769), *trnP-UGG* (NUMT2024), *trnY-GUA* (NUMT1738), *trnI-CAU* (NUMT1042), *trnI-GAU* (NUMT3860), *trnL-CAA* (NUMT221, NUMT227, NUMT966, NUMT1839), and *trnV-GAC* (NUMT3845, NUMT3860, NUMT4465). Collectively, these data demonstrate extensive mitochondrial DNA transfer to the nuclear genome in *H. cordata.*

Extremely high heteroplasmy in the *H. cordata* mitochondrial genome

Mapping short reads to the two chromosomes of *H. cordata* mitochondrial genome revealed substantial sequence heterogeneity, including 2,010 variants classified as SNVs (672), insertions (407), deletions (68), and complex variants (863). Variant frequencies ranged from 0.05 to 0.37, with over 90% of insertion and complex variants exhibiting low frequencies (0.05–0.1). Approximately 16.8% of SNVs and over 25% of deletions showed higher frequencies (> 0.1). Notably, 638 insertions, deletions, and complex variants involved sequence lengths > 100 bp, with the longest being 540 bp, 642 bp,

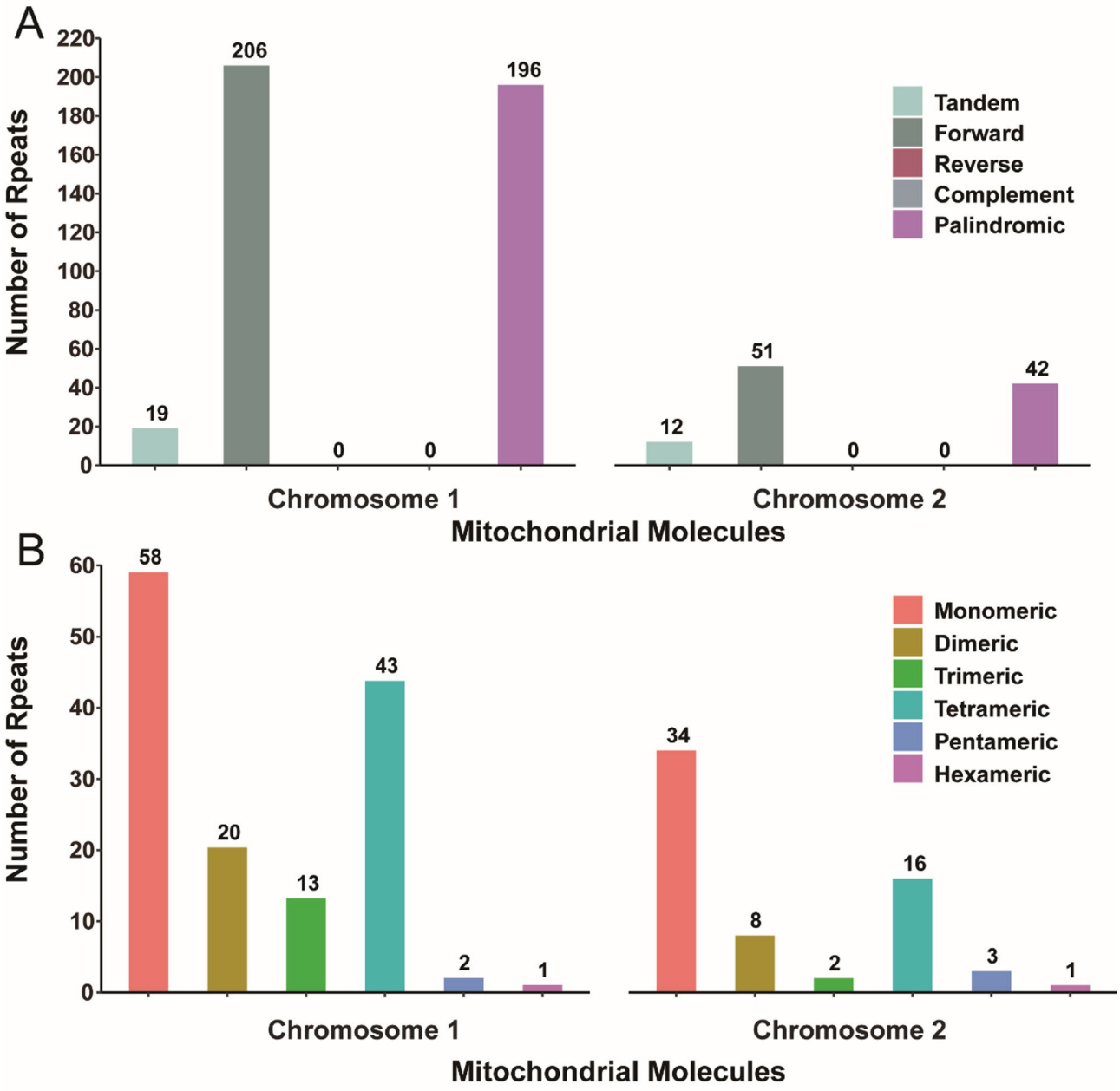

Fig. 4 Distribution and classification of repeat sequences in the *H. cordata* mitochondrial genome. **(A)** Distribution of various repeat types, including tandem, palindromic, forward, reverse, and complementary repeats in the same genomic regions. **(B)** Histogram depicting the number of monomeric to hexameric repeat sequences across Chromosome 1, and Chromosome 2. Each bar color corresponds to a specific type of repeat, as indicated in the legend. The counts above or within each bar represent the number of repeats for that particular category

and 375 bp, respectively. In contrast, the chloroplast genome exhibited far fewer variants (29 in total), with no complex variants detected (Additional file 3F).

Variants were distributed across most *H. cordata* mitochondrial genome regions but were nearly absent in conserved regions (>5 kb) associated with larger MTPTs and repetitive sequences (Fig. 6A). The MTPT regions contained 19 variants, predominantly featuring short deletions and complex variants spanning 1–9 bp. Among MTPTs > 1,000 bp, only MTPT2 had substantial variants, while MTPT1, MTPT12, and MTPT13 showed none (Fig. 6A, Additional file 3C). Similarly, 34 variants were identified in repeats > 100 bp, with few found in repeats > 1,000 bp, such as LR and R1 regions, while R2 had none (Additional file 3D). This variant distribution pattern showed no correlation with smaller MTPTs or repeats, establishing that genomic conservation correlates with large MTPTs and repetitive domains.

In protein-coding regions and noncoding RNA regions, 232 variants were detected, including 87 SNVs, 43

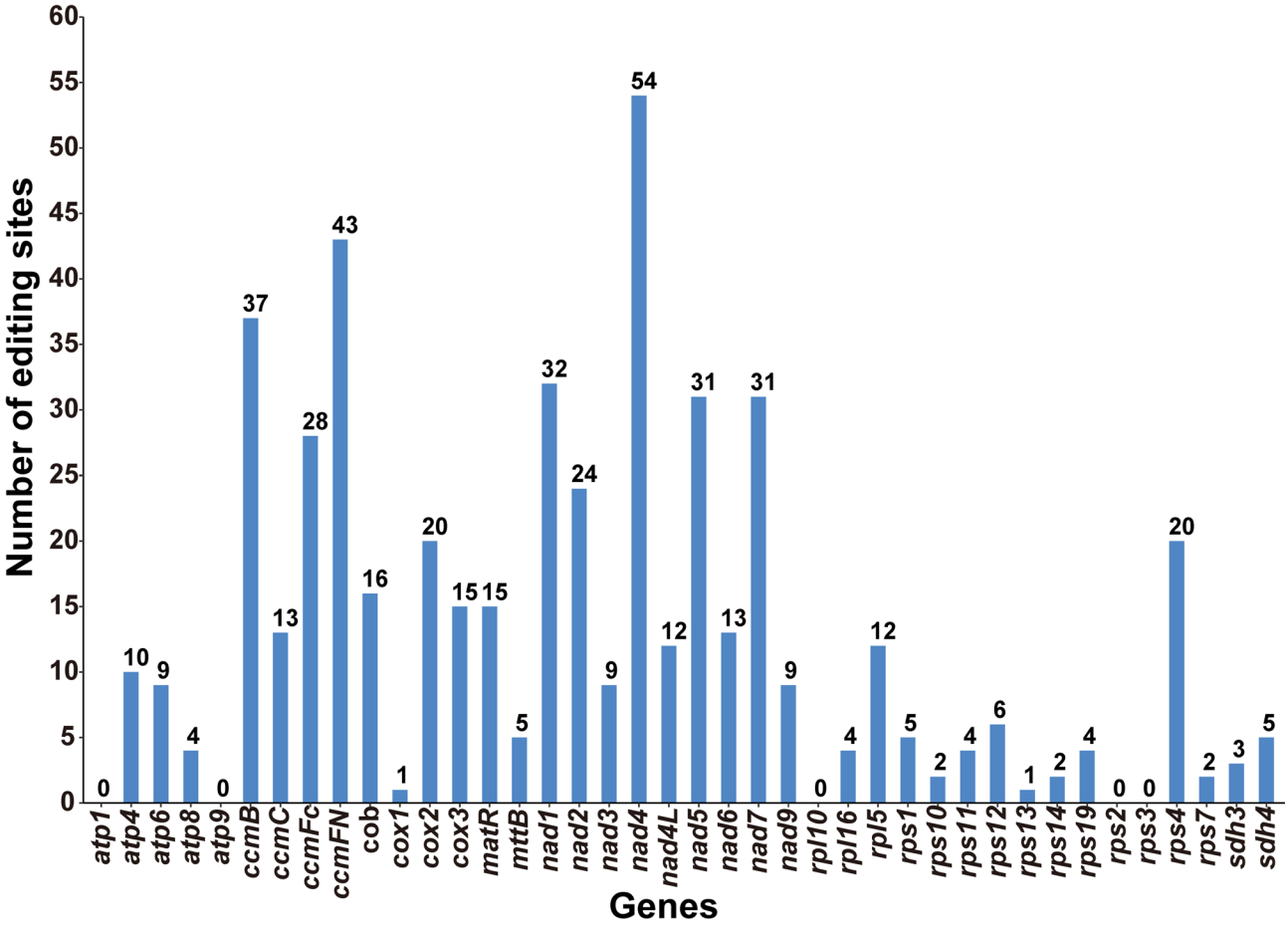

Fig. 5 Distribution of predicted RNA-editing sites across protein-coding genes (PCGs) in the mitochondria of *H. cordata*. The vertical bars represent the count of editing sites for each gene, with the gene names displayed along the x-axis. The y-axis quantifies the number of editing sites. Due to the absence of the *rpl10* gene in the mitochondrial genome of *Amborella trichopoda* (the reference genome), no RNA editing sites were detected for this gene

insertions, 9 deletions, and 93 complex variants (Additional file 3E). Among the 45 PCGs, 33 contained variants, with *rps3* and *nad5* showing more than 20 variants, indicating high heteroplasmy. All three rRNA genes had variants, with *rrn26* containing 28 variants. Only five out of 29 tRNA genes had variants, likely due to their shorter sequences. Over 60% of insertions, deletions, and complex variants in protein-coding regions induced frameshift mutations, potentially producing nonfunctional proteins. Additionally, 108 of 145 such variants resulted in changes to gene lengths of more than 10 bp, potentially significantly affecting gene function.

Phylogenetic and synteny analyses

Based on 33 conserved mitochondrial PCGs, we constructed both maximum likelihood and Bayesian phylogenetic trees that includes two outgroup species and 15 magnoliid species. The resulting phylogeny demonstrated robust nodal support, with maximum likelihood bootstrap values reaching 100 for all branches except one 76, and Bayesian posterior probabilities of 1.0 across all nodes. This topology resolved the 15 magnoliid species into three strongly supported monophyletic clades: Piperales, Magnoliales, and Laurales. Piperales emerged as the basal lineage within magnoliids, while Magnoliales and Laurales formed sister groups. Our analysis identified *H. cordata* and *S. chinensis* (both Saururaceae) as closest relatives within the Piperales clade, with *A. fimbriata* forming their sister group (Fig. 7A). Within Magnoliales, the four *Magnolia* species (*M. biondii*, *M. liliflora*, *M. officinalis*, and *M. figo*) clustered with *Liriodendron tulipifera* as sister taxa. Notably, the branch lengths among *Magnolia* species are very short, and the bootstrap value of one clade is less than 100, potentially reflecting evolutionary conservation in plant mitochondrial PCGs. In the Laurales clade, *Chimonanthus praecox* occupied the basal position. The three species of the genus *Cinnamomum* (*C. chekiangense*, *C. insularimontanum*, and *C. camphora*) exhibited closer phylogenetic affinity to *Machilus pauhoi* than to the other three species. This indicates that these four species share a most recent common ancestor. These finding were consistent

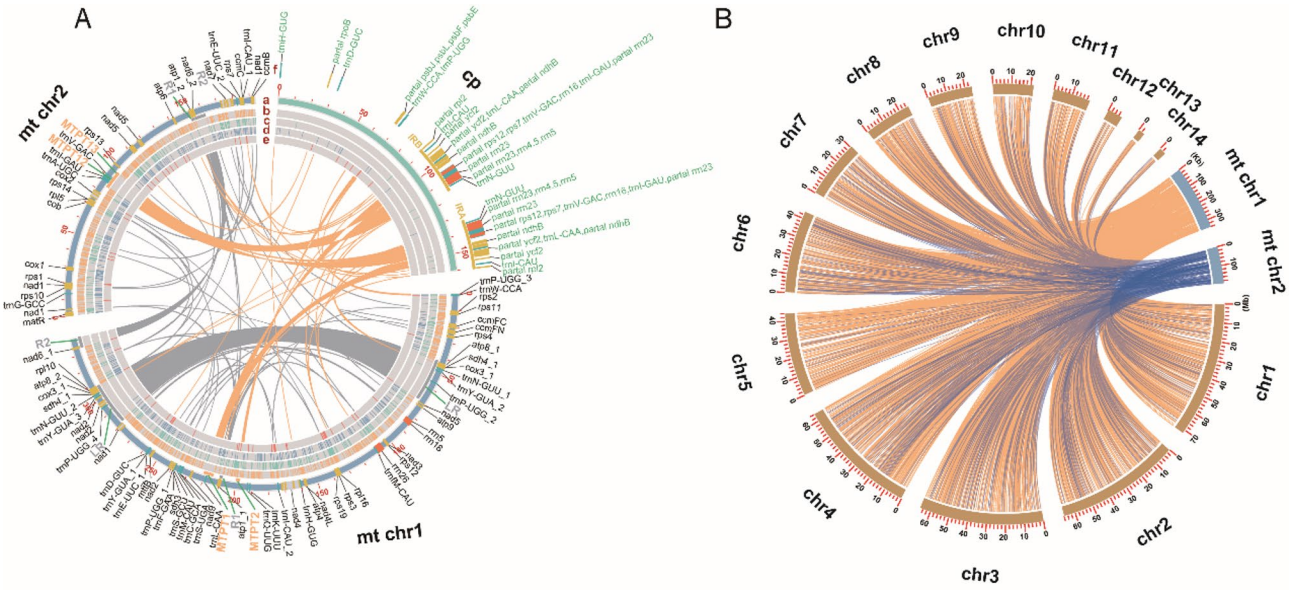

Fig. 6 Analysis of homologous fragments and heteroplasmy in *H. cordata* mitochondrial genome. **(A)** Distribution of homologous fragments and four variants on *H. cordata* mitochondrial genome and chloroplast genome. a represents the karyotype, blue represents the mitochondrial genome, and green represents the chloroplast genome. b-e, represents complex, single nucleotide variants (SNVs), insertions, and deletions variants, respectively. Different colours represent different gene types, and the yellow bar below represents the IR regions. The circled grays and orange arcs represent mitochondrial plastid DNA sequences (MTPTs) and mitochondrial repeat sequences, respectively. Black text around the circos plot represents mitochondrial genes, grey and orange text represent MTPTs and repeat regions larger than 1000 bp, respectively, and green text represents genes contained in MTPTs. **(B)** Homologous fragments shared between the mitochondrial genome and 14 chromosomes of the nuclear genome

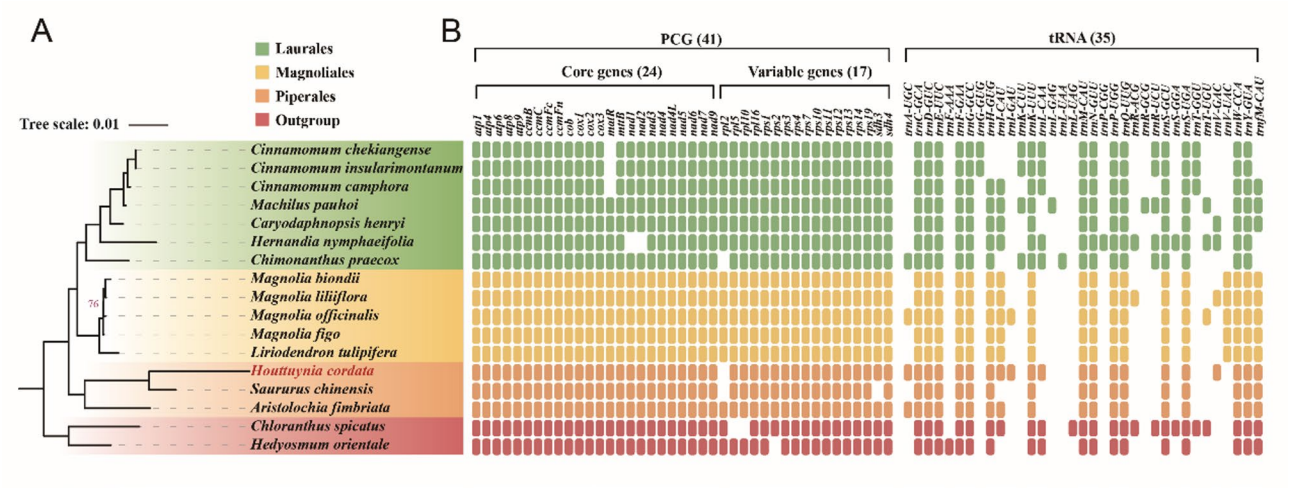

Fig. 7 The phylogenetic relationships and gene map of magnoliids. **(A)** Phylogenetic reconstruction of magnoliids based on 33 conserved mitochondrial protein-coding genes (PCGs). Maximum likelihood (ML) and Bayesian inference (BI) analyses yielded concordant topologies with robust nodal support: ML bootstrap values = 100 for all branches except red-colored numbers and BI posterior probabilities = 1.0 at all nodes. **(B)** Mitochondrial gene map of magnoliids where color-filled squares represent the presence of at least one complete copy. Species belonging to different order are color-coded for easy identification

with the latest Angiosperm Phylogeny Group IV (APG IV) [58].

We also analyzed the presence and absence of PCGs and tRNAs across these 17 species (Fig. 7B). The mitochondrial genomes of angiosperms actually always possess the same set of 24 core protein genes, which mainly encode respiratory proteins, but they often have significant differences in the composition of the other 17 protein genes that mainly encode ribosomal proteins [15, 59, 60]. Generally, these species exhibited a maximum loss of two PCGs, even in variable genes. This stands in stark contrast to monocotyledonous Poales plants, which typically lack > 2 variable genes [61]. Within Piperales, both *H. cordata* and *S. chinensis* mitochondrial genomes

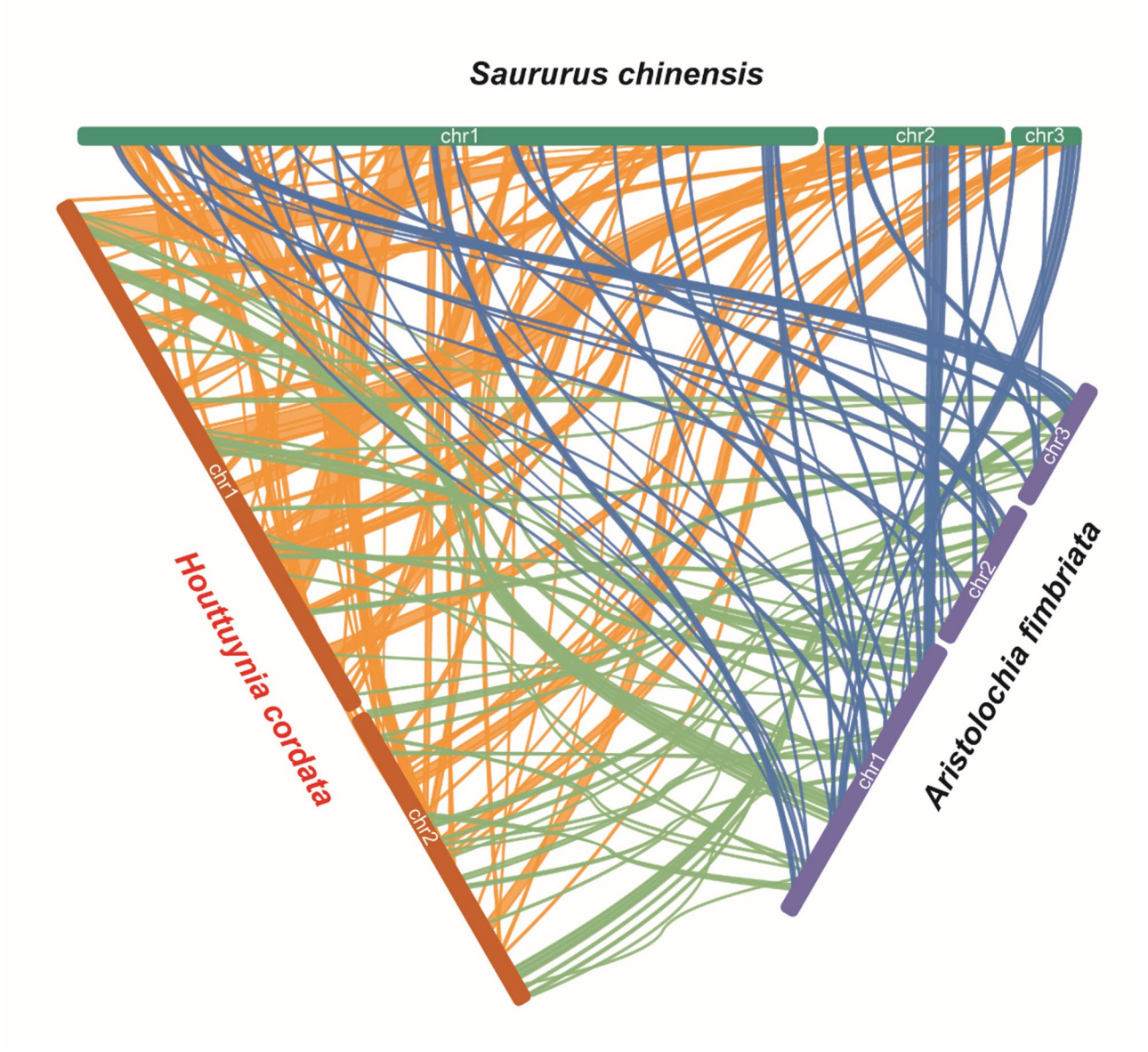

Fig. 8 Collinearity between *H. cordata* and two Piperales species. Collinear segments less than 500 bp in length are not shown

lost the *rpl2* gene, while *S. chinensis* additionally lost the *sdh3* gene. In contrast, *A. fimbriata* retained all PCGs. Regarding tRNA genes, these 17 species shared the following genes: *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnG-GCC*, *trnK-UUU*, *trnM-CAU*, *trnN-GUU*, *trnP-UGG*, *trnQ-UUG*, *trnS-GCU*, *trnS-UGA*, *trnW-CCA*, *and trnY-GUA*.

We conducted a collinearity analysis of the *H. cordata* mitochondrial genome in comparison with the mitochondrial genomes of *S. chinensis* and *A. fimbriata*. The analysis revealed significantly greater genome structure conservation within family members, with *H. cordata* and *S. chinensis* sharing 100 syntenic blocks (> 500 bp) spanning 46.3% of the *H. cordata* mitochondrial genome. The largest conserved segment measured 14,836 bp with 96.06% sequence identity. In contrast, *A. fimbriata* shared only 49 syntenic blocks, covering 11.05%, with the largest block at 2,963 bp and 94.56% similarity. Gene cluster analysis identified four conserved clusters across Piperales (n*ad3-rps12*, *rps14-rpl5*, *atp4-nad4L* and *atp8-cox3-sdh4*) and 13 shared clusters between *H. cordata* and *S. chinensis*. These findings indicate fewer structural variations within the same family compared to different families, with closely related species sharing more gene clusters (Fig. 8, Additional file 4).

Discussion

Complex structural features of the *H. cordata* mitochondrial genome

Plant mitochondrial genomes exhibit remarkable variation in size and structural complexity, ranging from compact genomes, such as the 66–187 kb in *Marchantia polymorpha* [62], to highly expanded genomes like the 11.6 Mb in *Caryodaphnopsis henryi* [63]. The mitochondrial genome of *H. cordata* (534,194 bp) represents a relatively compact form among magnoliids (349,849-1,168,029 bp), comparable in size to its closest relative *S. chinensis* (580,630 bp) (Table S1).

Structural analysis revealed that *H. cordata* has a predominantly bicircular structure (MSC1-2), and interconverts between various structures through homologous recombination mediated by repeating elements R1 and R2 (Fig. 1). This dynamic architecture aligns with the multi-chromosomal organizations observed in approximately 10% of sequenced species [14], where homologous recombination drives structural diversification. For instance, repetitive sequence-mediated chromosome number variation has been documented among *Silene noctiflora* germplasms [64]. Our findings reinforce that mitochondrial genome plasticity in plants arises primarily from repetitive element activity and recombination processes rather than genome size variations. While structural dynamics are governed by recombination mechanisms, genome size fluctuations result from repetitive element expansion, horizontal DNA transfer, and large indels [14]. Previous studies have demonstrated that plant mitochondria cope with high oxidative damage environments (caused by reactive oxygen species, ROS) by enhancing double-strand break repair (DSBR). This repair mechanism relies on template-dependent replication, but when template mismatches occur, it triggers gene recombination, duplication, and chimeric DNA formation, ultimately leading to drastic genomic structural rearrangements [65]. Despite the variability in genomes size and structure, the core gene number remains conserved across plant mitochondrial genomes (Fig. 7B). Similarly, *H. cordata* retains a core gene content consistent with other species within magnoliid species, demonstrating that genome size variation has minimal impact on functional gene conservation.

In the annotation of the *H. cordata* mitochondrial genome, we observed an interesting phenomenon: the apparent loss of *rpl2*, a nuclear-encoded ribosomal protein critical for translation initiation [66] (Fig. 2, Table 2). This observation aligns with findings by Yu et al. [67] demonstrating complete *rpl2* pseudogenization across Saururaceae species, suggesting lineage-specific gene loss. Despite its absence, the function of *rpl2* may be compensated by other genes or mechanisms, such as the increased expression of other ribosomal proteins [68, 69] or the transfer of *rpl2* to the nuclear genome [59], ensuring normal plant life activities. Additionally, mitochondrial genomes of seed plants are rich in repetitive sequences, including SSRs, tandem repeats, and dispersed repeats. Among these, SSRs have significant potential as molecular markers, aiding in the analysis of genetic diversity and species identification [70]. Analysis of repetitive sequences in the *H. cordata* mitochondrial genome identified 201 unique SSRs (Fig. 4B), providing a wealth of reference markers for assessing *H. cordata* diversity and facilitating species identification. These findings highlight the complexity and adaptability of the *H. cordata* mitochondrial genome, which, despite losing certain genes, can maintain functional integrity through compensatory mechanisms and abundant repetitive elements.

Codon usage patterns often reflect gene expression levels and evolutionary adaptation strategies. The *H. cordata* mitochondrial genome exhibits a typical plant genome feature, with a strong preference for A/U-ending codons. This bias likely enhances transcription and translation efficiency, thus improving protein synthesis and environmental adaptability [71]. Furthermore, RNA editing, defined as post-transcriptional single-base substitutions in mitochondrial genes, plays an essential role in gene expression [72]. We predicted 510 RNA editing sites in the *H. cordata* mitochondrial genome, including editing events that generate the start codon for the *nad1* gene and introduce stop codons for the *ccmFC* and *ccmFN* genes (Table S4). Similar phenomena have been reported in other plant species [73, 74]. Through RNA editing, plant mitochondria restore conserved protein sequences in post-transcriptional RNA, maintaining functional integrity [75].

MTPTs in *H. cordata* mitochondrial genome trace early angiosperm evolution

Our finding reveals the remarkable dynamism and evolutionary plasticity of *H. cordata* mitochondrial genomes, demonstrating through the identification of extensive MTPTs that mitochondrial genomes can actively integrate genetic information from other organelles. MTPTs are present in most plants with multiple polyplastid cells [76], and significant differences exist among various species [77]. Previous studies have revealed that MTPTs not only exhibit diversity across different plants but also demonstrate distinct transfer hotspots, with most transfers concentrated in the plastid IR regions [78, 79]. Consistent with these observations, our analysis of *H. cordata* mitochondria revealed that the majority of detected MTPTs originated from the hotspot region, and most of MTPTs located in the plastid IR regions (Fig. 6A, Table S5).

Previous research has demonstrated that certain MTPTs possess ancient evolutionary origins. For instance, the transfer hotspot region *trnP_UGG-trnW_CCA* was integrated into seed plant mitochondria approximately 300 million years ago (Mya), predating the divergence of gymnosperms and angiosperms [80]. Recent comparative analyses of 42 angiosperm species identified 10 conserved transfer hotspots [79]. Our results reveal shared hotspot regions (*ycf2*, *trnP_UGG-trnW_CCA*, *rrn23*, *trnI_GAU*) with *Liriodendron tulipifera* from the magnoliid clade. According to the APG IV classification system [58], magnoliids represent a basal angiosperm lineage originating approximately 120 Mya [81]. We hypothesize that MTPTs sharing hotspot regions likely emerged in early angiosperm mitochondria, reflecting a long-standing evolutionary process. Our study supports the Nhat's hypothesis [79] that initial transfer of the entire plastid DNA, followed by independent fragmentation, contributes to the current diversity of MTPTs in angiosperms.

Heteroplasmy in different regions of the mitochondrial genome of *H. cordata*

The mitochondrial genome of *H. cordata* exhibits remarkable variants and demonstrates high heteroplasmy. Variant detection indicates that the distribution of variants is consistent across functional regions and non-function region (Additional file 3E), with variant frequency showing strong positive correlation with gene length (Fig. S2). By sequencing a single plant, we effectively eliminated interindividual variation, suggesting that high genomic heteroplasmy may be an intrinsic characteristic of this species mitochondrial genome. Additionally, during variant detection, prevalent NUMTs in the *H. cordata* nuclear genome (Fig. 6B) posed a potential source of interference and false positives. To minimize such effect, we applied stringent quality control measures during Illumina reads mapping, including the use of a high-quality score threshold (q = 30), duplicate reads removal, and conservative variant calling parameters (-X 5 -f 0.05) in Vardict v1.82 [49] to enhance detection accuracy. On the other hand, according to NUMT detection results, although NUMTs were nearly derived from all regions of the mitochondrial sequence (Fig. 6B), the heteroplasmy analysis (Fig. 6A) revealed that certain regions (e.g., LR) exhibited minimal variants. This results indicates that NUMTs had limited interference on the outcomes of heteroplasmy analysis in this study. Previous studies have indicated that the elevated heteroplasmy in the holoparasitic plant *Rhopalocnemis phalloides* results from the loss of expression of two mitochondria-targeted DNA replication, repair, and recombination (DNA-RRR) genes, namely *WHY2* and *OSB2/3/4* [82]. Thus, we infer that the pervasive heteroplasmy observed in the *H. cordata* mitochondrial genome may originate from compromised DNA repair mechanisms.

Notably, no variants were found in certain larger repeat regions and MTPTs, indicating relatively conservative genomic segments with low heteroplasmy (Fig. 6). This finding aligns with reports of conserved repetitive region in the *R. phalloides* mitochondrial genome [82]. Additionally, we observed that MTPT2, positioned outside the transfer hotspot region, exhibits high heteroplasmy (Fig. 6, Table S5, Additional file 3C). Conversely, MTPTs located within the transfer hotspot displayed minimal variation, with only MTPT6 containing a single SNV. This observation suggests that transfer hotspot-associated MTPTs may be stably maintained within the *H. cordata* mitochondrial genome. However, the underlying mechanism for the reduced heteroplasmy in larger repeat regions and MTPTs remains enigmatic but may be attributed to the normal expression of the associated homologous recombination pathways [13]. Further research is necessary to elucidate these genomic dynamics.

Genome collinearity and evolution

To date, only a limited number of studies have exclusively utilized mitochondrial genes to address plant phylogenetic questions, with most research integrating mitochondrial genomic data with plastid genome or nuclear genome data into combined datasets [83, 84]. This study reconstructed a robust tree (Fig. 6A) using all 33 conserved mitochondrial protein-coding genes from 15 magnoliid species and two outgroups. The resulting topology groups these taxa into three monophyletic clades that align with APG IV classifications [58]. This demonstrates that mitochondrial-encoded genes possess certain feasibility as genetic barcodes for phylogenetic studies. Therefore, in future research, mitochondrial genomes may effectively resolve topological controversies at deep-level nodes in plant phylogeny, providing novel evidence and alternative hypotheses for research on early angiosperm evolution and diversification.

Based on the robust phylogenetic tree we obtained, which includes 15 magnoliid species, we systematically investigated the loss patterns of protein-coding genes and tRNA genes. This comprehensive analysis provides insights into mitochondrial genome variation and the evolutionary dynamics. The magnoliid lineage exhibited a minimal degree of PCG loss (Fig. 7), consistent with previous research [67]. In contrast, tRNA genes demonstrated significant depletion across examined species (Fig. 7), suggesting a conserved pattern of mitochondrial tRNA gene reduction within this lineage. Recent studies indicates that cytoplasmic tRNA can compensate for the losses of the mitochondrial tRNA gene [85], demonstrating cellular compensatory mechanisms that

optimize energy expenditure and enhance environmental adaptability.

Plant mitochondrial genomes display substantial structural variation across species due to rapid evolutionary dynamics [75]. However, researchers have identified certain conserved sequences within plant mitochondrial genomes, including gene coding regions [18] and specific genome segments critical for growth, development, mitochondrial morphology, and composition [86]. Comparative synteny analysis of *H. cordata* with two closely related Piperales species revealed four conserved gene clusters, including *atp8-cox3-sdh4* and *atp4-nad4L*, among others. Detailed analysis demonstrated that these gene clusters are also present across other magnoliid species within our study framework (Fig. S3, Fig. S4). Previous studies suggest that the *cox3-sdh4* gene cluster may enhance gene expression efficiency and play a synergistic role in cellular respiration and energy metabolism [87, 88]. Notably, this gene cluster is prevalent in dicots but frequent absence in most monocots [89]. Similarly, the conserved gene cluster *atp4-nad4L* demonstrates wide distribution across dicot species [89]. Our findings underscore the remarkable similarity between magnoliids and dicots, suggesting a potentially evolutionary affinity between the two.

Conclusion

In this study, we accomplished the first comprehensive mitochondrial genome assembly of *H. cordata* utilizing PacBio HiFi long-read sequencing. We identified four potential mitochondrial genome conformations, particularly focusing on the characterization of the predominant bicyclic structure (MSC1-2). Our analysis revealed the structural complexity of the *H. cordata* mitochondrial genome. Furthermore, most MTPTs were originated from the IR regions of the plastid genome. A preliminary exploration of individual-level heteroplasmy within the mitochondrial genome indicated that the *H. cordata* mitochondrial genome exhibits substantial heteroplasmy levels, interspersed with relatively conserved regions. Phylogenetic and collinearity analyses demonstrated that, despite significant structural differences between the *H. cordata* mitochondrial genome and those of closely related species, conserved regions are still present. This study elucidates the distinctive characteristics of plant mitochondrial genomes and addresses a critical knowledge gap in understanding the *H. cordata* mitochondrial genome.

Abbreviations

ATP	Adenosine triphosphate
APG IV	Angiosperm Phylogeny Group IV
BIC	Bayesian Information Criterion
IR	Inverted Repeat
ML	Maximum Likelihood
MTPTs	Mitochondrial plastid DNAs
NUMTs	Nuclear mitochondrial DNAs
PCGs	Protein-Coding Genes
RSCU	Relative Synonymous Codon Usage
rRNA	Ribosomal RNA Genes
SNVs	Single nucleotide variants
SSRs	Simple sequence repeats
tRNA	Transfer RNA Genes

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06801-6.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Author contributions

Y.H.H., J.Y., and B.Y.Z. designed the work. Y.H.H., S.Q.H., and Y.X.L. assembled and annotated the mitogenome. W.B.S. performed bioinformatics analyses. Q.H.L., and Y.X.L. prepared figures and tables. Y.H.H., S.Q.H., J.Y., and B.Y.Z. wrote the manuscript. Y.H.H., J.Y., B.Y.Z., S.W., and C.S. revised the manuscript. All authors approved the final manuscript.

Funding

This work was supported by the Taishan Scholar Project (No. tsqn202306214 and No. tsqn202312205).

Data availability

The mitogenome and plastid genome sequence have been deposited in National Center for Biotechnology Information with accession number: PQ214805, PQ214806 and PQ643225 (https://www.ncbi.nlm.nih.gov).

Declarations

Ethics approval and consent to participate

The plant materials used in this study were collected from farmland in the suburbs of Kunming, Yunnan, China. The landowners provided explicit consent for the collection of these materials. And the sampling was conducted in accordance with local regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 10 December 2024 / Accepted: 28 May 2025
Published online: 07 June 2025

References

1. Bhattacharyya N, Sarma S. Assessment of availability, ecological feature, and habitat preference of the medicinal herb *Houttuynia cordata* thunb in the Brahmaputra Valley of assam, India. Environ Monit Assess. 2010;160:277.
2. Gupta S, Bharalee R. Genetic diversity and population structure Of a medicinal herb *Houttuynia cordata* thunb. Of north-east India. Plant Mol Biol Rep. 2021;39:434–42.
3. Wu Z, Deng X, Hu Q, Xiao X, Jiang J, Ma X, et al. *Houttuynia cordata* thunb: an ethnopharmacological review. Front Pharmacol. 2021;12:714694.
4. Yin D, Song J, Xiang Z, Tian S, Li G, Qi S. Hydroponic cultivation enhances biomass and medicinal safety of *Houttuynia cordata* thunb: an integrated physiological and multi-omics analysis. Ind Crops Prod. 2024;220:119284.

5. Lau K-M, Lee K-M, Koon C-M, Cheung CS-F, Lau C-P, Ho H-M, et al. Immunomodulatory and anti-SARS activities of *Houttuynia cordata*. J Ethnopharmacol. 2008;118:79–85.
6. Liu J, Yuan S, Yao Y, Wang J, Scalabrino G, Jiang S, et al. Network Pharmacology and molecular Docking elucidate the underlying Pharmacological mechanisms of the herb *Houttuynia cordata* in treating pneumonia caused by SARS-CoV-2. Viruses. 2022;14:1588.
7. Lin C-K, Chen B-Y, Ting JU, Rogio KGG, Tsai P-W, Liu Y-C. Deciphering *Houttuynia cordata* extract as electron shuttles with anti-COVID-19 activity and its performance in microbial fuel cells. J Taiwan Inst Chem Eng. 2023;145:104838.
8. Birky CW. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. Annu Rev Genet. 2001;35:125–48.
9. Osellame LD, Blacker TS, Duchen MR. Cellular and molecular mechanisms of mitochondrial function. Best Pract Res Clin Endocrinol Metab. 2012;26:711–23.
10. Sagan L. On the origin of mitosing cells. J Theor Biol. 1967;14:225–IN6.
11. Gabaldón T. Origin and early evolution of the eukaryotic cell. Annu Rev Microbiol. 2021;75:631–47.
12. Kozik A, Rowan BA, Lavelle D, Berke L, Schranz ME, Michelmore RW, et al. The alternative reality of plant mitochondrial DNA: one ring does not rule them all. Plos Genet. 2019;15:e1008373.
13. Chevigny N, Schatz-Daas D, Lotfi F, Gualberto JM. DNA repair and the stability of the plant mitochondrial genome. Int J Mol Sci. 2020;21:328.
14. Wu Z, Liao X, Zhang X, Tembrock LR, Broz A. Genomic architectural variation of plant mitochondria—a review of multichromosomal structuring. J Sytematics Evol. 2022;60:160–8.
15. Skippington E, Barkman TJ, Rice DW, Palmer JD. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all *Nad* genes. Proc Natl Acad Sci. 2015;112.
16. Putintseva YA, Bondar EI, Simonov EP, Sharov VV, Oreshkova NV, Kuzmin DA, et al. *Siberian larch* (*larix sibirica* ledeb.) mitochondrial genome assembled using both short and long nucleotide sequence reads is currently the largest known mitogenome. BMC Genomics. 2020;21:654.
17. Wang J, Kan S, Liao X, Zhou J, Tembrock LR, Daniell H, et al. Plant organellar genomes: much done, much more to do. Trends Plant Sci. 2024;29:754–69.
18. Palmer JD, Herbon LA. Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. J Mol Evol. 1988;28:87–97.
19. Sloan DB, Wu Z. Molecular evolution: the perplexing diversity of mitochondrial RNA editing systems. Curr Biol. 2016;26:R22–4.
20. Ślipiko M, Myszczyński K, Buczkowska K, Bączkiewicz A, Sawicki J. Supermitobarcoding in plant species identification? It can work! The case of leafy liverworts belonging to the genus calypogeia. Int J Mol Sci. 2022;23:15570.
21. Szczepański S, Łabiszak B, Lasek M, Wachowiak W. Hybridization has localized effect on genetic variation in closely related pine species. BMC Plant Biol. 2024;24:1007.
22. Brownfield L. Plant breeding: revealing the secrets of cytoplasmic male sterility in wheat. Curr Biol. 2021;31:R724–6.
23. Eckardt NA. Cytoplasmic male sterility and fertility restoration. Plant Cell. 2006;18:515–7.
24. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. Genome Biol. 2020;21:1–16.
25. Huang P, Li Z, Wang H, Huang J, Tan G, Fu Y et al. A genome assembly of decaploid *Houttuynia cordata* provides insights into the evolution of *Houttuynia* and the biosynthesis of alkaloids. Hortic Res. 2024;uhae203.
26. Yang Z, He F, Mai Y, Fan S, An Y, Li K, et al. A near-complete assembly of the *Houttuynia cordata* genome provides insights into the regulatory mechanism of flavonoid biosynthesis in Yuxingcao. Plant Commun. 2024;5:101075.
27. Zhu B, Feng Q, Yu J, Yu Y, Zhu X, Wang Y, et al. Chloroplast genome features of an important medicinal and edible plant: *Houttuynia cordata* (saururaceae). PLoS ONE. 2020;15:e0239823.
28. Jinlu L, Shuo W, Jing Y, Ling W, Shiliang Z. A modified CTAB protocol for plant DNA extraction: A modified CTAB protocol for plant DNA extraction. Chin Bull Bot. 2013;48:72–8.
29. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37:540–6.
30. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de Novo genome assemblies. Method Biochem Anal. 2015;31:3350–2.
31. Jin J-J, Yu W-B, Yang J-B, Song Y, DePamphilis CW, Yi T-S, et al. GetOrganelle: a fast and versatile toolkit for accurate de Novo assembly of organelle genomes. Genome Biol. 2020;21:1–31.
32. Li J, Ni Y, Lu Q, Chen H, Liu C. PMGA: a plant mitochondrial genome annotator. Plant Commun. 2025;6:101191.
33. Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Res. 2021;49:9077–96.
34. Chen Y, Ye W, Zhang Y, Xu Y. High speed BLASTN: an accelerated megablast search tool. Nucleic Acids Res. 2015;43:7762–8.
35. Dunn NA, Unni DR, Diesh C, Munoz-Torres M, Harris NL, Yao E, et al. Apollo: democratizing genome annotation. PLoS Comput Biol. 2019;15:e1006790.
36. Zhang X, Chen H, Ni Y, Wu B, Li J, Burzyński A, et al. Plant mitochondrial genome map (PMGmap): a software tool for the comprehensive visualization of coding, noncoding and genome features of plant mitochondrial genomes. Mol Ecol Resour. 2024;24:e13952.
37. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq–versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45:W6–11.
38. Li H. Minimap2: pairwise alignment for nucleotide sequences. Method Biochem Anal. 2018;34:3094–100.
39. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of samtools and BCFtools. GigaScience. 2021;10:giab008.
40. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available python tools for computational molecular biology and bioinformatics. Method Biochem Anal. 2009;25:1422.
41. Lee BD. Python implementation of codon adaptation index. JOSS. 2018;3:905.
42. Wickham H. ggplot2. WIREs Comput Stat. 2011;3:180–5.
43. Lenz H, Hein A, Knoop V. Plant organelle RNA editing and its specificity factors: enhancements of analyses and new database features in PREPACT 3.0. BMC Bioinformatics. 2018;19:255.
44. He W, Yang J, Jing Y, Xu L, Yu K, Fang X. NGenomeSyn: an easy-to-use and flexible tool for publication-ready visualization of syntenic relationships across multiple genomes. Bioinformatics. 2023;39:btad121.
45. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: A fast and versatile genome alignment system. PLoS Comput Biol. 2018;14:e1005944.
46. Edwards RJ, Field MA, Ferguson JM, Dudchenko O, Keilwagen J, Rosen BD, et al. Chromosome-length genome assembly and structural variations of the primal Basenji dog (canis lupus familiaris) genome. BMC Genomics. 2021;22:188.
47. Chen C, Wu Y, Li J, Wang X, Zeng Z, Xu J, et al. TBtools-II: A one for all, all for one bioinformatics platform for biological big-data mining. Mol Plant. 2023;16:1733–42.
48. Vasimuddin M, Misra S, Li H, Aluru S. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In: 2019 IEEE international parallel and distributed processing symposium (IPDPS). IEEE; 2019. pp. 314–24.
49. Lai Z, Markovets A, Ahdesmaki M, Chapman B, Hofmann O, McEwen R, et al. VarDict: A novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic Acids Res. 2016;44:e108–108.
50. Katoh K, Asimenos G, Toh H. Multiple alignment of DNA sequences with MAFFT. In: Posada D, editor. Bioinformatics for DNA sequence analysis. Totowa, NJ: Humana; 2009. pp. 39–64.
51. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4.
52. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61:539–42.
53. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. 2012;9:772–772.
54. He W, Xiang K, Chen C, Wang J, Wu Z. Master graph: an essential integrated assembly model for the plant mitogenome based on a graph-based framework. Brief Bioinform. 2023;24:bbac522.
55. Martin W, Herrmann RG. Gene transfer from organelles to the nucleus: how much, what happens, and why? Plant Physiol. 1998;118:9–17.
56. Mower JP, Jain K, Hepburn NJ. The role of horizontal transfer in shaping the plant mitochondrial genome. Advances in botanical research. Elsevier; 2012. pp. 41–69.
57. Hazkani-Covo E, Zeller RM, Martin W. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genet. 2010;6:e1000834.
58. Byng JW, Smets EF, van Vugt R, Bidault E, Davidson C, Kenicer G, et al. The phylogeny of angiosperms poster: a visual summary of APG IV family relationships and floral diversity. Global Flora. 2018;4:1–4.

59. Adams KL, Ong HC, Palmer JD. Mitochondrial gene transfer in pieces: fission of the ribosomal protein gene rpl2 and partial or complete gene transfer to the nucleus. Mol Biol Evol. 2001;18:2289–97.
60. Adams KL, Qiu Y-L, Stoutemyer M, Palmer JD. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc Natl Acad Sci USA. 2002;99:9905–12.
61. Wang H, Wu Z, Li T, Zhao J. Highly active repeat-mediated recombination in the mitogenome of the aquatic grass hygroryza aristata. BMC Plant Biol. 2024;24:644.
62. Ohyama K, Oda K, Yamoto K, Ohta E, Takemura M, Akashi K. The mitochondrial genome of a liverwort, marchantia polymorpha. Mol Biology Plant Mitochondria. 1995;:597–633.
63. Yang S, Huang J, Qu Y, Zhang D, Tan Y, Wen S, et al. Phylogenetic incongruence in an Asiatic species complex of the genus *caryodaphnopsis* (lauraceae). BMC Plant Biol. 2024;24:616.
64. Wu Z, Cuthbert JM, Taylor DR, Sloan DB. The massive mitochondrial genome of the angiosperm *silene noctiflora* is evolving by gain or loss of entire chromosomes. Proc Natl Acad Sci USA. 2015;112:10185–91.
65. Wynn EL, Christensen AC. Repeats of unusual size in plant mitochondrial genomes: identification, incidence and evolution. G3: genes. Genomes Genet. 2019;9:549–59.
66. Kubo N, Ozawa K, Hino T, Kadowaki K. A ribosomal protein L2 gene is transcribed, spliced, and edited at one site in rice mitochondria. Plant Mol Biol. 1996;31:853–62.
67. Yu R, Chen X, Long L, Jost M, Zhao R, Liu L, et al. *De Novo* assembly and comparative analyses of mitochondrial genomes in piperales. Genome Biol Evol. 2023;15:evad041.
68. Adams KL, Qiu Y-L, Stoutemyer M, Palmer JD. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc Natl Acad Sci. 2002;99:9905–12.
69. Adams KL, Palmer JD. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol Phylogenet Evol. 2003;29:380–95.
70. Morgante M, Hanafey M, Powell W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet. 2002;30:194–200.
71. Majeed A, Kaur H, Bhardwaj P. Selection constraints determine preference for a/U-ending codons in *taxus contorta*. Genome. 2020;63:215–24.
72. Aphasizhev R. RNA editing. Mol Biol. 2007;41:227–39.
73. Zheng P, Wang D, Huang Y, Chen H, Du H, Tu J. Detection and analysis of C-to-U RNA editing in rice mitochondria-encoded ORFs. Plants. 2020;9:1277.
74. Zhang J, Liu G, Wei J. Assembly and comparative analysis of the first complete mitochondrial genome of Setaria italica. Planta. 2024;260:23.
75. Mower JP, Sloan DB, Alverson AJ. Plant mitochondrial genome diversity: the genomics revolution. Plant Genome Divers Volume. 2012;1:123–44.
76. Smith DR. Extending the limited transfer window hypothesis to inter-organelle DNA migration. Genome Biol Evol. 2011;3:743–8.
77. Sloan DB, Wu Z. History of plastid DNA insertions reveals weak deletion and AT mutation biases in angiosperm mitochondrial genomes. Genome Biol Evol. 2014;6:3210–21.
78. Wang X-C, Chen H, Yang D, Liu C. Diversity of mitochondrial plastid DNAs (MTPTs) in seed plants. Mitochondrial DNA Part A. 2018;29:635–42.
79. Nhat Nam N, Pham Anh Thi N, Do HDK. New insights into the diversity of mitochondrial plastid DNA. Genome Biol Evol. 2024;16:evae184.
80. Wang D, Wu Y-W, Shih AC-C, Wu C-S, Wang Y-N, Chaw S-M. Transfer of Chloroplast genomic DNA to mitochondrial genome occurred at least 300 MYA. Mol Biol Evol. 2007;24:2040–8.
81. Massoni J, Couvreur TL, Sauquet H. Five major shifts of diversification through the long evolutionary history of magnoliidae (angiosperms). BMC Evol Biol. 2015;15:49.
82. Yu R, Sun C, Zhong Y, Liu Y, Sanchez-Puerta MV, Mower JP, et al. The minicircular and extremely heteroplasmic mitogenome of the holoparasitic plant rhopalocnemis phalloides. Curr Biol. 2022;32:470–e4795.
83. Qiu Y, Dombrovska O, Lee J, Li L, Whitlock BA, Bernasconi-Quadroni F, et al. Phylogenetic analyses of basal angiosperms based on nine plastid, mitochondrial, and nuclear genes. Int J Plant Sci. 2005;166:815–42.
84. Qiu Y-L, Lee J, Bernasconi-Quadroni F, Soltis DE, Soltis PS, Zanis M, et al. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. Nature. 1999;402:404–7.
85. Warren JM, Salinas-Giegé T, Triant DA, Taylor DR, Drouard L, Sloan DB. Rapid shifts in mitochondrial tRNA import in a plant lineage with extensive mitochondrial tRNA gene loss. Mol Biol Evol. 2021;38:5735–51.
86. Liberatore KL, Dukowic-Schulze S, Miller ME, Chen C, Kianian SF. The role of mitochondria in plant development and stress tolerance. Free Radic Biol Med. 2016;100:238–56.
87. Giegé P, Knoop V, Brennicke A. Complex II subunit 4 (sdh4) homologous sequences in plant mitochondrial genomes. Curr Genet. 1998;34:313–7.
88. Siqueira S, Dias S, Hardouin P, Pereira F, Lejeune B, De Souza A. Transcription of succinate dehydrogenase subunit 4 (sdh4) gene in potato: detection of extensive RNA editing and co-transcription with cytochrome oxidase subunit III (cox3) gene. Curr Genet. 2002;41:282–90.
89. Liao X, Zhao Y, Kong X, Khan A, Zhou B, Liu D, et al. Complete sequence of Kenaf (*Hibiscus cannabinus*) mitochondrial genome and comparative analysis with the mitochondrial genomes of other plants. Sci Rep. 2018;8:12714.

Publisher's note