



OPEN ACCESS

EDITED BY

Xiaozeng Yang,
Beijing Academy of Agricultural and
Forestry Sciences, China

REVIEWED BY

Qiang Fan,
Sun Yat-sen University, China
Jorge Maurício Mondego,
Instituto Agronômico de Campinas (IAC),
Brazil

*CORRESPONDENCE

Chang Liu

✉ cliu6688@yahoo.com;

✉ cliu@implad.ac.cn

†These authors have contributed equally to
this work

RECEIVED 18 July 2023

ACCEPTED 25 September 2023

PUBLISHED 11 October 2023

CITATION

Ni Y, Zhang X, Li J, Lu Q, Chen H, Ma B
and Liu C (2023) Genetic diversity of
Coffea arabica L. mitochondrial genomes
caused by repeat-mediated
recombination and RNA editing.
Front. Plant Sci. 14:1261012.
doi: 10.3389/fpls.2023.1261012

COPYRIGHT

© 2023 Ni, Zhang, Li, Lu, Chen, Ma and Liu.
This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Genetic diversity of *Coffea arabica* L. mitochondrial genomes caused by repeat-mediated recombination and RNA editing

Yang Ni[†], Xinyi Zhang[†], Jingling Li[†], Qianqi Lu, Haimei Chen,
Binxin Ma and Chang Liu*

Center for Bioinformatics, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China

Background: *Coffea arabica* L. is one of the most important crops widely cultivated in 70 countries across Asia, Africa, and Latin America. Mitochondria are essential organelles that play critical roles in cellular respiration, metabolism, and differentiation. *C. arabica*'s nuclear and chloroplast genomes have been reported. However, its mitochondrial genome remained unreported. Here, we intended to sequence and characterize its mitochondrial genome to maximize the potential of its genomes for evolutionary studies, molecular breeding, and molecular marker developments.

Results: We sequenced the total DNA of *C. arabica* using Illumina and Nanopore platforms. We then assembled the mitochondrial genome with a hybrid strategy using Unicycler software. We found that the mitochondrial genome comprised two circular chromosomes with lengths of 867,678 bp and 153,529 bp, encoding 40 protein-coding genes, 26 tRNA genes, and three rRNA genes. We also detected 270 Simple Sequence Repeats and 34 tandem repeats in the mitochondrial genome. We found 515 high-scoring sequence pairs (HSPs) for a self-to-self similarity comparison using BLASTn. Three HSPs were found to mediate recombination by the mapping of long reads. Furthermore, we predicted 472 using deep-mt with the convolutional neural network model. Then we randomly validated 90 RNA editing events by PCR amplification and Sanger sequencing, with the majority being non-synonymous substitutions and only three being synonymous substitutions. These findings provide valuable insights into the genetic characteristics of the *C. arabica* mitochondrial genome, which can be helpful for future study on coffee breeding and mitochondrial genome evolution.

Conclusion: Our study sheds new light on the evolution of *C. arabica* organelle genomes and their potential use in genetic breeding, providing valuable data for

developing molecular markers that can improve crop productivity and quality. Furthermore, the discovery of RNA editing events in the mitochondrial genome of *C. arabica* offers insights into the regulation of gene expression in this species, contributing to a better understanding of coffee genetics and evolution.

KEYWORDS

Coffea arabica, mitochondrial genome, repetitive sequence analysis, homologous recombination, RNA editing

1 Introduction

Coffea arabica L. belongs to the family Rubiaceae (Noir et al., 2004). It was widely distributed in tropical and subtropical regions of Asia, Africa, and Latin America (Ferreira T. et al., 2019). *C. arabica* is an essential agricultural crop mainly grown as a cash crop in tropical countries (Vidal et al., 2010). Caffeine is the main active ingredient of coffee and has many bioactive effects, such as neuroprotection, improving vascular function, reducing blood sugar, and protecting the liver (Cano-Marquina et al., 2013; Bhupathiraju et al., 2014; Kolahdouzan and Hamadeh, 2017). The first draft genome sequence of *C. canephora*, a close relative of *C. arabica*, was assembled to provide insights into the evolution of caffeine biosynthesis (Denoeud et al., 2014). Lashermes et al. studied the molecular characterization and origin of the *C. arabica* genome, revealing specific sequences and conserved regions. Mekbib et al. found that the SNPs might contribute to the genetic variations associated with important agronomic traits such as caffeine content, yield, disease, and pest in *C. arabica* (Mekbib et al., 2022). Min et al. (2019) and Park et al. (2019) reported the chloroplast genome of *C. arabica* and developed the molecular markers. However, the *C. arabica* mitochondrial genome remains unreported.

Mitochondria are semi-autonomous, membrane-bound organelles (Margulis and Bermudes, 1985). According to the endosymbiosis theory, they originated from an engulfed alpha-proteobacterium, which eventually formed a symbiotic relationship with the host cells (Poole and Penny, 2007). Their primary functions involve providing ATP to cells via oxidative phosphorylation (Shtolz and Dan, 2019) and synthesizing metabolic precursors (Scott and Logan, 2007). Moreover, mitochondria play vital roles in cell differentiation, growth, division, and programmed cell death (Zamzami et al., 1997; Valero, 2014).

The mitochondrial genome structure of angiosperms varies significantly among species (Handa, 2003; Sloan et al., 2012). Higher plant mtDNA is abundant in repeat sequences, which mediate homologous recombination and contribute to the evolution of plant mitochondrial genomes (Chevigny et al., 2020). Homologous recombination is essential for maintaining genomic stability, enhancing genomic diversity, driving genomic evolution, and adapting to environmental changes. The location, frequency, and type of recombination can influence genome structure and evolution (Gualberto and Newton, 2017). This widespread homologous recombination results in the complexity of plant mitochondrial genomes (Chevigny et al., 2020).

Earlier studies have demonstrated that mtDNA exhibits intricate and dynamic structures, including linear and branched chromosomes, which may be intermediates in replication or recombination and represent multiple genome isoforms (Backert et al., 1997). In some cases, mtDNAs were found as a master circular molecule (Cheng et al., 2021; Fang et al., 2021; Ni et al., 2022), such as in *Panax ginseng* (Jang et al., 2021) and *Vitex rotundifolia* (Bendich, 1996). In the *Coriandrum sativum*, the mitochondrial genomes consist of two circular molecules (lengths 82,926 bp and 224,590 bp) (Kozik et al., 2019), while experiments by Kozik et al. revealed that the predominant form of mitochondrial DNA molecules in *Lactuca sativa* (Kozik et al., 2019) is simple and branched linear.

RNA editing refers to the phenomenon that the base changes occur at the molecular level of the mRNA produced by transcription, including the insertion, deletion, and replacement of nucleotides and other different ways, resulting in its sequence cannot complement the gene coding sequence and the amino acid composition of the protein produced by translation also changes, which is a supplement to the central dogma (Gott and Emeson, 2000). Three study groups first documented RNA editing in flowering plant mitochondria 30 years ago (Covello and Gray, 1989; Araya et al., 1995). It was soon discovered that RNA editing in plants also occurs in chloroplasts (Hoch et al., 1991), but there are generally fewer RNA base changes in angiosperms than in mitochondria. Comparison of organelle RNA sequences with their corresponding mtDNA sites shows that editing events most often occur at the first or second codon position and may affect the amino acids defined by the mitochondrial genome (Lenz et al., 2018a). Many of these nucleotide changes lead to codon changes that specify amino acids highly conserved in evolution (Mower, 2005; Sloan, 2017; Edera et al., 2018; Lenz et al., 2018b; Brenner et al., 2019). RNA editing predominantly affects non-synonymous positions of protein-coding regions (Picardi et al., 2010; Grewe et al., 2014; Grimes et al., 2014), changing the resulting amino-acid sequences (Gualberto et al., 1989).

The Rubiaceae family, a diverse and significant group in the plant kingdom, has garnered attention for its mitochondrial genome variations and peculiarities. *Damnacanthus indicus* showcases dynamic evolution with unique gene and intron content, serving as the first reference mitochondrial genome for the Rubiaceae family (Han et al., 2021). The *Neolamarckia cadamba* mitochondrial genome not only revealed its phylogenetic position within the Gentianales order but also shed light on its taxonomic

relationship within the Rubiaceae family (Wang et al., 2021). *Oldenlandia corymbosa*'s smaller mitochondrial genome size compared to other family members indicates variations in size and content within the Rubiaceae species (Julca et al., 2023). The intricacies of the *Psychotria viridis* mitochondrial genome, presented through multiple mitogenome structures and evidence of heteroplasmy, enrich our understanding of mitochondrial genome organization (Varani et al., 2022). Meanwhile, the mitogenome of *Scyphiphora hydrophyllacea* stands out due to its intricate intron content and its phylogenetic positioning within the Gentianales (Chen Y. et al., 2020). These studies collectively offer invaluable insights into the mitochondrial genomes of Rubiaceae species, furthering the understanding of their evolution and diversity.

Here, we first sequenced the complete *C. arabica* mitochondrial genome. The coffee mitochondrial genome contained multiple conformations resulting from recombination mediated with repetitive elements. In addition, we identified 54 fragments that were likely to originate from the chloroplast genome. Lastly, we identified 90 RNA editing sites. These results demonstrated multiple mechanisms that led to the diversity of *C. arabica* mitochondrial genome. This study provides a theoretical basis for the evolution of *C. arabica* organelle genomes and molecular breeding.

2 Materials and methods

2.1 Plant materials, DNA extraction, and sequencing

We collected *C. arabica* plants from the Shunlong Nursery Farm, located in Baojingyuan Village, Hongyang Town, Puning City, Jieyang City, Guangdong Province, China. The geographic coordinates are Longitude: 116.25001 and Latitude: 23.43801. The altitude of the collection site is 55 meters above sea level. The soil at the collection site of *C. arabica* plants is classified as "Wuni Ditian", or "Black Clay Paddy Soil", primarily found in the regions of Guangdong province such as Huiyang, Zhanjiang, Maoming, Zhaoqing, Guangzhou, and Shaoguan (<http://vdb3.soil.csdb.cn/>) and cleaned the fresh leaves with DPEC water. For NGS sequencing, we extracted DNA using the Magnetic Plant Genomic DNA Kit (Tiangen, China) and constructed a short-read DNA library with an insert size of 350 bp using the TIANSeq Fast DNA Library Kit (catalog number NG102, Illumina, California, USA). We sequenced this library on an Illumina HiSeq X sequencer (Illumina, USA).

For long-read sequencing using Oxford Nanopore technology, we extracted DNA with the NEB Monarch HMW DNA Extraction Kit (catalog number: T3060L, New England Biolabs, Massachusetts, USA). We then constructed a DNA library with 10 kb fragment insert sizes using the DNA Library Kit (catalog number: SQK-LSK110, TIANGEN) and sequenced it on a PromethION sequencer (Novogene Co., Ltd., Beijing, China).

2.2 Organelle genome assembly and annotation

The organelle genomes were assembled using a hybrid assembly strategy (Figure S1). In step 1, we extracted the cpgenome reads with the parameters "-R 15 -k 21,45,65,85,105 -F embplant_pt" and the mitochondrial genome reads with the parameters "-R 20 -k 21,45,65,85,105 -P 1000000 -F embplant_mt" using GetOrganelle software (Jin et al., 2020) from Illumina data (SRA Accession Number: SRR17345023). In step 2, the short reads were *de novo* assembled using the SPAdes software (Bankevich et al., 2012) embedded in Unicycler software (Wick et al., 2017) into a unitig graph. In step 3, the double bifurcating structures (DBS) in the unitig graph were resolved by mapping the Nanopore long reads (SRA Accession Number: SRR17345007) using Unicycler software (Wick et al., 2017). The hybrid assembly strategy could minimize the false assemblies generated in the polishing step resulting from the interference from Nuclear Mitochondrial DNAs (NUMTs) and Mitochondrial Plastid DNAs (MTPTs) sequences (Timmis et al., 2004; Hazkani-Covo et al., 2010).

We annotated the cpgenome using the CPGAVAS2 web server with database 2 (Shi et al., 2019), and the annotation was checked by the CPGView web server (Liu S. et al., 2023). The *C. arabica* mitochondrial genome was annotated with the Geseq web server (Michael et al., 2017) and IPMGA webserver (<http://www.1kmpg.cn/mga/>). The mitochondrial genome annotation results were visualized using the OGdraw web server (Stephan et al., 2019). The annotation errors were manually corrected with Apollo software (Lewis et al., 2002).

2.3 Tandem repeat elements analysis

We identified two kinds of tandem repeat elements. The Simple Sequence Repeat (SSRs) were detected using the Misa web server (Beier et al., 2017) with default parameters. The tandem repeats were identified using the TRF (Tandem repeats finder) webserver with default parameters (Benson, 1999). The distribution of these repeat elements was visualized by the Circos package (Zhang et al., 2013) embedded in the TBtools (Chen C. et al., 2020).

2.4 Identification and validation of repeats able to mediate homologous recombination

We used bioinformatic analysis and experimental methods to verify the possible presence of repeat-mediated recombination products of the mitochondrial genome. For the bioinformatics method, we mapped the long reads (Nanopore data of WGS) to the sequences corresponding to the hypothetical recombination products. We first identified the high-scoring sequence pairs (HSPs) using the BLASTn program (Chen et al., 2015). It should be pointed out that the sequences from a HSP can be considered a pair of

dispersed repeat units with low level of sequence similarity. Then, we extracted 1,000 bp long flanking sequences on both sides of each HSP sequence. The resulting two sequences corresponded to two conformations (c1 and c2). Then we recombined the two sequences in silico to create sequences corresponding to the two recombined conformations (c3 and c4). We mapped the Nanopore reads to the sequences corresponding to c1-4. If there were long reads spanned the repeat region of a conformation, we considered the corresponding conformation present.

The presence of recombination products around these repeats was further validated using the PCR amplification and Sanger sequencing methods. The IDT SciTools (Owczarzy et al., 2008) was used for designing the primers specific to amplify each conformation using PCR. The primer sequences are listed in Table S1. We used approximately 1 μ l DNA, 1 μ l 10 μ M each of the forward and reverse primer, 13 μ l 2 \times Taq PCR Master Mix, and 10 μ l ddH₂O for PCR with the following conditions: 94°C for 3 min; 35 cycles of 94°C for 30 s, 60°C for 30 s and 72°C for 1 min; 72°C for 10 min. Lastly, the PCR products with the expected size were further sequenced using the Sanger method.

2.5 Identification of mitochondrial plastid DNAs and phylogenetic analysis

To identify potential MTPTs, we conducted a reciprocal BLASTn search between the complete plastome (OL789882) and mitochondrial genomes (OL789880-OL789881). The objective was to discern segments with sequence similarity. In the initial analysis, the entire plastome served as the “query” sequence, while the mitochondrial genome, comprising both chromosome sequences, acted as the “subject” sequence. For the reciprocal analysis, we reversed these roles, using the mitochondrial genomes as the “query” and the plastome as the “subject”. Results from both searches were then combined. The specific parameters we employed for this BLASTn search were “-evalue 1e-5 -outfmt 6” (Chen et al., 2015). The mapping results of reads to the MTPT regions were visualized using Tablet software (Milne et al., 2012) and examined manually.

We used 18 mitochondrial genomes for phylogenetic analysis and selected the two Lamiaceae species *Scutellaria barbata* (NC_065025.1) and *Scutellaria franchetiana* (NC_065026.1) as outgroups. All mitochondrial genomes were downloaded from the RefSeq database with the following accession numbers: *C. arabica* (OL789880.1, OL789881.1), *N. cadamba* (MT320890.1-MT364442.1), *P. serpens* (NC_069806.1), *P. viridis* (NC_066984.1), *S. hydrophyllacea* (NC_057654.1), *D. indicus* (MZ285075.1), *O. corymbosa* (OX459128.1), *Gentiana crassicaulis* (OM320814.1), *Gentiana straminea* (OM328072.1), *Hoya lithophytica* (MW719051.1), *Rhazya stricta* (NC_024293.1), *Asclepias syriaca* (NC_022796.1), *Trachelospermum jasminoides* (OR333986.1), *Cynanchum wilfordii* (MH931257.1-MH931259.1) *Cynanchum auriculatum* (MH410146.1-MH410148.1), *Gelsemium elegans* (MN388837.1). The CDS sequences were extracted from Genbank format files using the PhyloSuite software (Zhang et al.,

2020) and aligned with MAFFT software (Katoh and Standley, 2013). The aligned sequences were used to construct the phylogenetic tree using IQTREE2 (Hoang et al., 2017) with the maximum-likelihood method. Subsequently, the bootstrap analysis was evaluated using UFBoot with 1,000 replicates (Hoang et al., 2017). Finally, the phylogenetic tree was visualized using the iTOL website (Letunic and Bork, 2019).

2.6 Collinearity and gene content analysis

We conducted a collinearity analysis on the mitochondrial genomes of closely related species of *C. arabica* within the Rubiaceae family, as identified in the phylogenetic analysis. Initially, we downloaded the reference genomes from NCBI for this analysis. Subsequently, we employed the online version of MAFFT at <https://mafft.cbrc.jp/alignment/server/> for sequence alignment and Dotplot visualization (Katoh and Standley, 2013). Considering the potential annotation errors in the record, we re-annotated all the genome files using IPMGA webserver to initiating the gene statistical analysis. Following this, we utilized a custom script to tally the annotation results.

2.7 RNA editing sites analysis

We used the Deepred-mt software to predict RNA editing events using the convolutional neural network (CNN) model (Edera et al., 2021). We retained predictions with probability values greater than 0.9. Subsequently, we randomly validated the prediction results of RNA editing events with PCR amplification and sanger sequencing. The PCGs were amplified with cDNA (complementary DNA) and gDNA (genomic DNA) to identify those RNA editing sites. The protein-coding genes (PCGs) were extracted from Genbank format files using the PhyloSuite software (Zhang et al., 2020). The IDT web server (Owczarzy et al., 2008) was used for designing the PCR primer of all PCGs. The primers used to validate the RNA editing sites are shown in Table S2. The PCR experiment conditions were the same as those described in subsection 2.4. The PCR products were sequenced using the Sanger method and mapped to the protein-coding sequences to validate the RNA editing sites.

3 Results

3.1 Mitochondrial genome assembly and annotation

We generated 11.1 Gb Nanopore sequencing data (SRA Accession Number: SRR17345007) and 6 Gb Illumina sequencing data (SRA Accession Number: SRR17345023) in total. The mitochondrial genome reads extracted from Illumina data were assembled using SPAdes (Bankevich et al., 2012; Jin et al., 2020) into a unitig graph. The unitig graph contained 13 double-bifurcating

structures (DBS) (Figure S2). We mapped the long Nanopore reads to each DBS to resolve these DBS using the Unicycler software (Figure S2). We defined the major conformations as the two reverse-complementing conformations mapped with more long reads. After retaining only the major conformation, the final assembly contained two mitochondrial genome chromosomes (MC): MC1 and MC2, which were 867,678 bp and 153,529 bp, respectively (Figure 1). The total GC content of the *C. arabica* mitochondrial genome MC1 and MC2 was 44.6% and 44.7%, respectively. To compare the mitochondrial genomes of *C. arabica* with those of other Rutaceae species, we analyzed the basic information of 14 released mitochondrial genomes in the NCBI database of Rutaceae, which included *C. arabica* (Table 1). Among them, *Neolamarckia cadamba*, which has two chromosomes like *C. arabica*, has a much smaller mitochondrial genome (Total length: 414,980 bp). The *C. arabica* mitochondrial genome is the largest among the 13 species, 104,888 bp larger than the second largest, *Psychotria serpens* (NC_069806.1/MT528155.1). The significant difference in mitochondrial genome size could be attributed to extensive rearrangements and sequence migration during the evolution of the Rutaceae mitochondrial genome (Kitazaki and Kubo, 2010).

All the 24 core PCGs of the plant mitochondrial genome have been found in the mitochondrial genome (Figure 1; Table 2). These included five ATP synthase genes (*atp1*, *atp4*, *atp6*, *atp8*, and *atp9*); nine NADH dehydrogenase genes (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, and *nad9*), four cytochrome c biogenesis genes (*ccmB*, *ccmC*, *ccmFc*, *ccmFn*), three cytochrome C oxidase genes (*cox1*,

cox2, and *cox3*), one protein transport subunit (*mttB*), one maturase (*matR*) and one cytochrome oxidase (*cob*). The core genes *atp1*, *atp6*, and *atp9* had two copies in the mitochondrial genome. Besides, there were 12 variable genes in the mitochondrial genome, including three genes encoding the ribosomal protein large subunit genes (*rpl5*, *rpl10*, and *rpl16*), seven genes encoding the ribosomal protein small subunit genes (*rps1*, *rps3*, *rps4*, *rps7*, *rps10*, *rps12*, and *rps13*), and one succinate dehydrogenase gene (*sdh3*, *sdh4*). The variable genes *rps1* and *rps12* had two copies in the mitochondrial genome. In addition, we identified three rRNA genes and 26 tRNA genes, corresponding to 18 unique genes, in the *C. arabica* mitochondrial genome.

3.2 Tandem repeat elements analysis

Simple sequence repeats (SSR), or microsatellite sequences, were tandem repeats with shorter repeat units than six bp (Bhattacharai et al., 2021). In this study, we used the Misa web server (Beier et al., 2017) to detect the SSRs in the *C. arabica* mitochondrial genome. We identified 270 SSRs in total (Table S3). MC1 and MC2 contained 206 and 64 SSRs, respectively. The most abundant type of SSRs were the tetramer SSRs, accounting for 37.41% of the total SSRs. The pentamer and hexameric SSRs were only found in MC1 (Table S3).

For the long tandem repeat sequences, twenty-eight were identified by TRF (Benson, 1999) in MC1, and six were found in MC2 (Table S4). The tandem repeat with a length of 9 bp from 553,613 to 553,656 had the largest copy numbers (4.9 times).

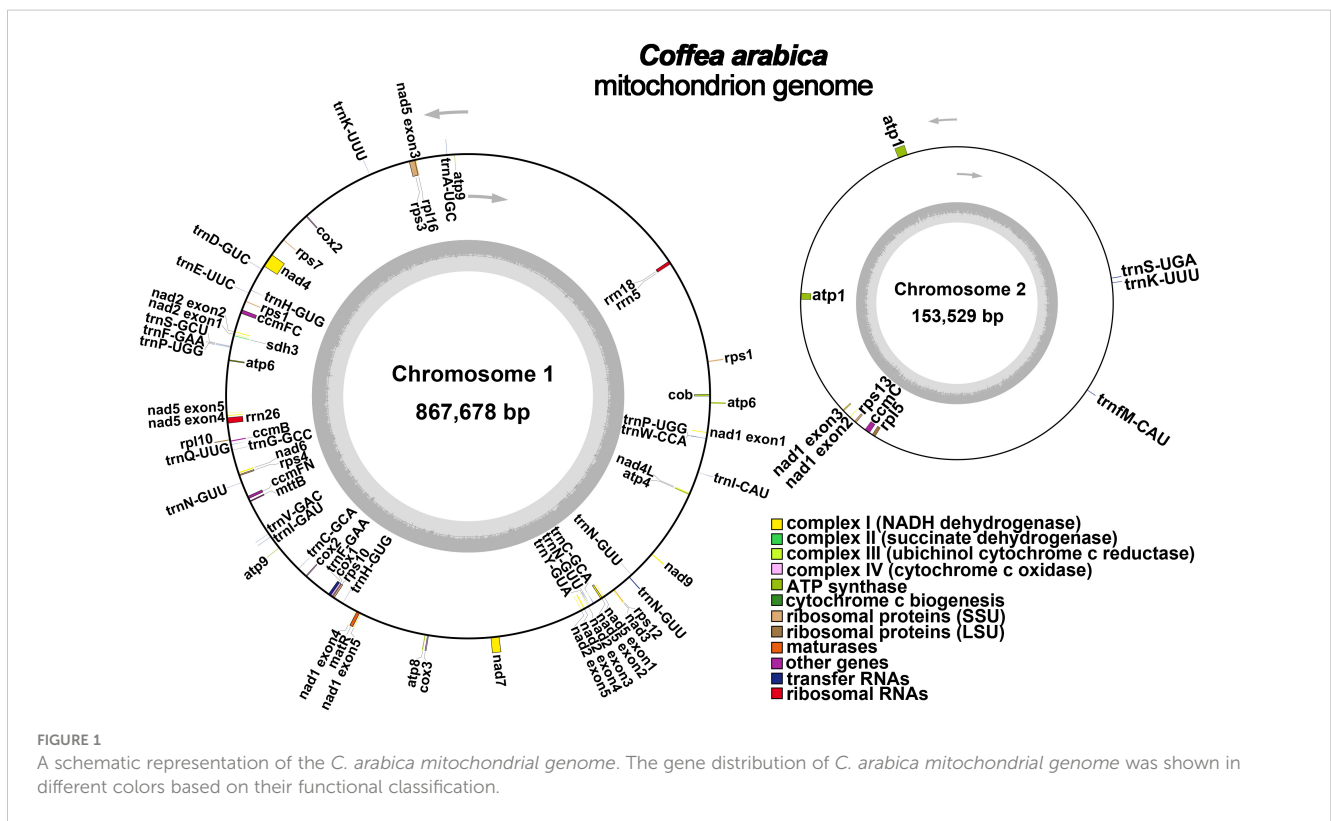


TABLE 1 Basic information on mitochondrial genomes of Rubiaceae.

Genome Number	Species	Chromosome Number	Accession Number	Length	Shape	GC content
1	<i>Coffea arabica</i>	chromosome 1	OL789880.1	867,678 bp	Circular	44.6%
	<i>Coffea arabica</i>	chromosome 2	OL789881.1	153,529 bp	Circular	44.7%
2	<i>Damnacanthus indicus</i> var. <i>indicus</i>	NA	MZ285071.1	417,815 bp	Circular	44.5%
3	<i>Damnacanthus indicus</i> var. <i>indicus</i>	NA	MZ285072.1	419,010 bp	Circular	44.5%
4	<i>Damnacanthus indicus</i> var. <i>indicus</i>	NA	MZ285073.1	419,435 bp	Circular	44.5%
5	<i>Damnacanthus indicus</i> var. <i>indicus</i>	NA	MZ285074.1	419,429 bp	Circular	44.5%
6	<i>Damnacanthus indicus</i> var. <i>indicus</i>	NA	MZ285075.1	417,661 bp	Circular	44.5%
7	<i>Damnacanthus indicus</i> var. <i>indicus</i>	NA	MZ285076.1	417,816 bp	Circular	44.5%
8	<i>Neolamarckia cadamba</i>	chromosome 1	MT320890.1	109,836 bp	Circular	45.5%
	<i>Neolamarckia cadamba</i>	chromosome 2	MT364442.1	305,144 bp	Linear	45%
9	<i>Psychotria serpens</i>	NA	NC_069806.1	916,319 bp	Circular	44.2%
11	<i>Psychotria viridis</i>	NA	NC_066984.1	615,370 bp	Circular	44.4%
12	<i>Psychotria viridis</i>	NA	ON064100.1	570,344 bp	Circular	44.4%
13	<i>Scyphiphora hydrophyllacea</i>	NA	NC_057654.1	354,155 bp	Circular	44.4%
14	<i>Oldenlandia corymbosa</i>	NA	OX459128.1	258,274 bp	Circular	44%

3.3 Dispersed repeat analysis and repeat mediated recombination analysis

A previous study has shown that dispersed repeats can mediate homologous recombination (Li et al., 2021). These dispersed repeats can vary significantly in terms of the percentage of sequence identity. As a result, we used BLASTn to compare the mitochondrial genome sequences to themselves. The resulting similar sequences were called High-scoring sequence pairs (HSP). They are equivalent to the dispersed repeats and were named “R,” followed by their HSP numbers.

We found 515 HSPs in MC1 (OL789880) and MC2 (OL789881) (Table S5). We then compared the DBS sequences with these HSP sequences. We found all thirteen DBS sequences identical to some HSP sequences (Table S5). We then mapped the long reads to the four conformations of these 13 DBS/HSP. However, the mapping results supported the recombined conformation for only eight HSP (Table S6). Among them, six HSPs: R1 (DBS06), R22, R30 (DBS09), R175, R278, and R285, were found in MC1 of *C. arabica*, and the

R406, R462 were found shared between the MC1 and MC2 (Table S6).

We defined the recombination frequency (RF) as the percentage of reads mapped to the minor conformations divided by those mapped to all four conformations. The RFs for all HSPs were less than 0.22, except for R1, which was 0.54. The HSPs could have contributed to the complex structure of plant mitochondrial genomes and increased the diversity of plant mitochondrial genomes (André et al., 1992).

In addition to bioinformatics analysis methods, we used PCR amplification and Sanger sequencing methods to verify the presence of the recombination products corresponding to the minor conformations. We obtained PCR products for those associated with repeats R1, R30, and R406. It should be pointed out that the R1 and R30 corresponded to the bs01 and bs09 found in the unitig graph. The schematic representation of the location of the primers used to amplify the fragments associated with R1 and R30 is shown in Figures 2A, B. The electrophoretic results of these recombination products are shown in Figure 2C.

TABLE 2 Genome composition in the *C. arabica* mitochondrial genome.

Group of genes	Name of genes
ATP Synthase	<i>atp1</i> (x2), <i>atp4</i> , <i>atp6</i> (x2), <i>atp8</i> , <i>atp9</i> (x2)
Cytochrome c Biogenesis	<i>ccmB</i> , <i>ccmC</i> , <i>ccmFc</i> , <i>ccmFn</i>
Cytochrome b	<i>cob</i>
Cytochrome Oxidase	<i>cox1</i> , <i>cox2</i> , <i>cox3</i>
Maturase	<i>matR</i>
Protein Transport Subunit	<i>mttB</i>
NADH Dehydrogenase	<i>nad1</i> , <i>nad2</i> , <i>nad3</i> , <i>nad4</i> , <i>nad4L</i> , <i>nad5</i> , <i>nad6</i> , <i>nad7</i> , <i>nad9</i>
Ribosomal Protein Large Subunit	<i>rpl5</i> , <i>rpl10</i> , <i>rpl16</i>
Ribosomal Protein Small Subunit	<i>rps1</i> (x2), <i>rps3</i> , <i>rps4</i> , <i>rps7</i> , <i>rps10</i> , <i>rps12</i> (x2), <i>rps13</i>
Succinate Dehydrogenase	<i>sdh3</i> , <i>sdh4</i>
Ribosomal RNA	<i>rrn5</i> , <i>rrn18</i> , <i>rrn26</i>
Transfer RNA	<i>trnC</i> -GCA (x2), <i>trnD</i> -GUC, <i>trnE</i> -UUC, <i>trnF</i> -GAA (x2), <i>trnM</i> -CAU, <i>trnG</i> -GCC, <i>trnH</i> -GUG (x2), <i>trnI</i> -GAU, <i>trnI</i> -CAU, <i>trnK</i> -UUU (x2), <i>trnN</i> -GUU (x4), <i>trnP</i> -UGG (x2), <i>trnQ</i> -UUG, <i>trnS</i> -GCU, <i>trnS</i> -UGA, <i>trnV</i> -GAC, <i>trnW</i> -CCA, <i>trnY</i> -GUA

In Figure 2C, we obtained the PCR products for R1, R30, and R406. The PCR products were sequenced with the Sanger methods. And the results are shown in Supplementary Files 1–3. Three pairs of repetitive sequences divided the genome into six contigs (contigs 1–6). Recombination mediated by R1 could rearrange contigs 1 and 2 to form the minor conformation Mic01 (Figure 3; Supplementary File 1). Similarly, the recombination of R30 could cause the rearrangement of contigs 2 and 3 to form the minor conformation Mic02 (Figure 3; Supplementary File 2). In contrast, MC1 and MC2 were combined by the recombination-mediated with the direct repeat R406 to form the minor conformation Mic03 (Figure 3; Supplementary File 3).

3.4 Identification of foreign DNAs in the mitochondrial genome

The plant mitochondrial genome could incorporate chloroplast and nuclear DNA during its evolution (Kang et al., 2021). There were 32 and 10 homologous DNA fragments in MC1 and MC2 (Figure 4; Table S7). The total length was 26,096 bp, 2.55% of the

whole mitochondrial genome. The longest fragment was 4,117 bp in MC1, and the shortest was only 31 bp in MC2. We annotated those DNA fragments, and ten complete genes (*rps7*, *psaA*, *trnN*-GUU, *trnP*-UGG, *trnW*-CCA, *trnD*-GUC, *trnH*-GUG, *trnN*-GUU, *trnI*-CAU) were found in those DNA fragments (Table S7). All the MTPTs were visualized and checked manually. The presence of long reads spanning the MTPTs supports that these sequences were indeed in the mitochondrial genome (Figures S3–44).

3.5 Phylogenetic analysis

We conducted the phylogenetic analysis of the mitochondrial genomes of the Rubiaceae, Gentianaceae, Apocynaceae, Gelsemiaceae and Lamiaceae species. thirty-eight common PCGs were identified from these genomes, namely, *atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFc*, *ccmFn*, *cob*, *cox2*, *cox3*, *matR*, *mttB*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*, *rpl2*, *rpl5*, *rpl10*, *rpl16*, *rps1*, *rps2*, *rps3*, *rps4*, *rps7*, *rps10*, *rps12*, *rps13*, *rps14*, *rps19* and *sdh3*. The best model was GTR+F+R2, according to the BIC. The phylogenetic tree showed that *C. arabica* and *S. hydrophyllacea* were clustered together with 100% bootstrap values in the context of the currently available mitochondrial genome (Figure 5).

3.6 Comparative mitogenomic analysis in the Rubiaceae family

By analyzing the protein-coding gene count of the Rubiaceae family, we observed that eight mitochondrial genomes all possess 24 core genes (Table S8). Yet, when it comes to variable genes, the protein-coding genes are not conserved. For instance, *C. arabica*, *S. hydrophyllacea*, and *N. cadamba* have one copy of the *sdh3* gene, with *S. hydrophyllacea* containing two copies, while other species lack the *sdh3* gene. *P. viridis* has lost the *rps1* gene, and both *P. viridis* and *P. serpens* are missing the *rps4* gene. Hence, in the Rubiaceae family, although core genes remain relatively conserved, there is a marked disparity in variable genes across different species.

Utilizing collinearity analyses, we observed a non-conserved mitochondrial genome structure within the Rubiaceae family, as evidenced by the limited number of conserved blocks presented in Figure S45. Intra-genomic comparisons of *C. arabica* highlighted the presence of a 7804 bp reverse repeat block. Contrastingly, when juxtaposed with other species, with the sole exception of *O. corymbosa*, *C. arabica* exhibited a reverse repeat block spanning an approximate length of 6000–7000 bp. Notably, *O. corymbosa* mitogome uniquely presented a collinear block around 3500 bp. Such observations underscored the substantial structural diversifications the mitochondrial genomes within the Rubiaceae family have encountered throughout evolutionary processes. It is plausible that these genomic rearrangements are tethered to distinct environmental acclimations, intricate biological interplays, or speciation events.

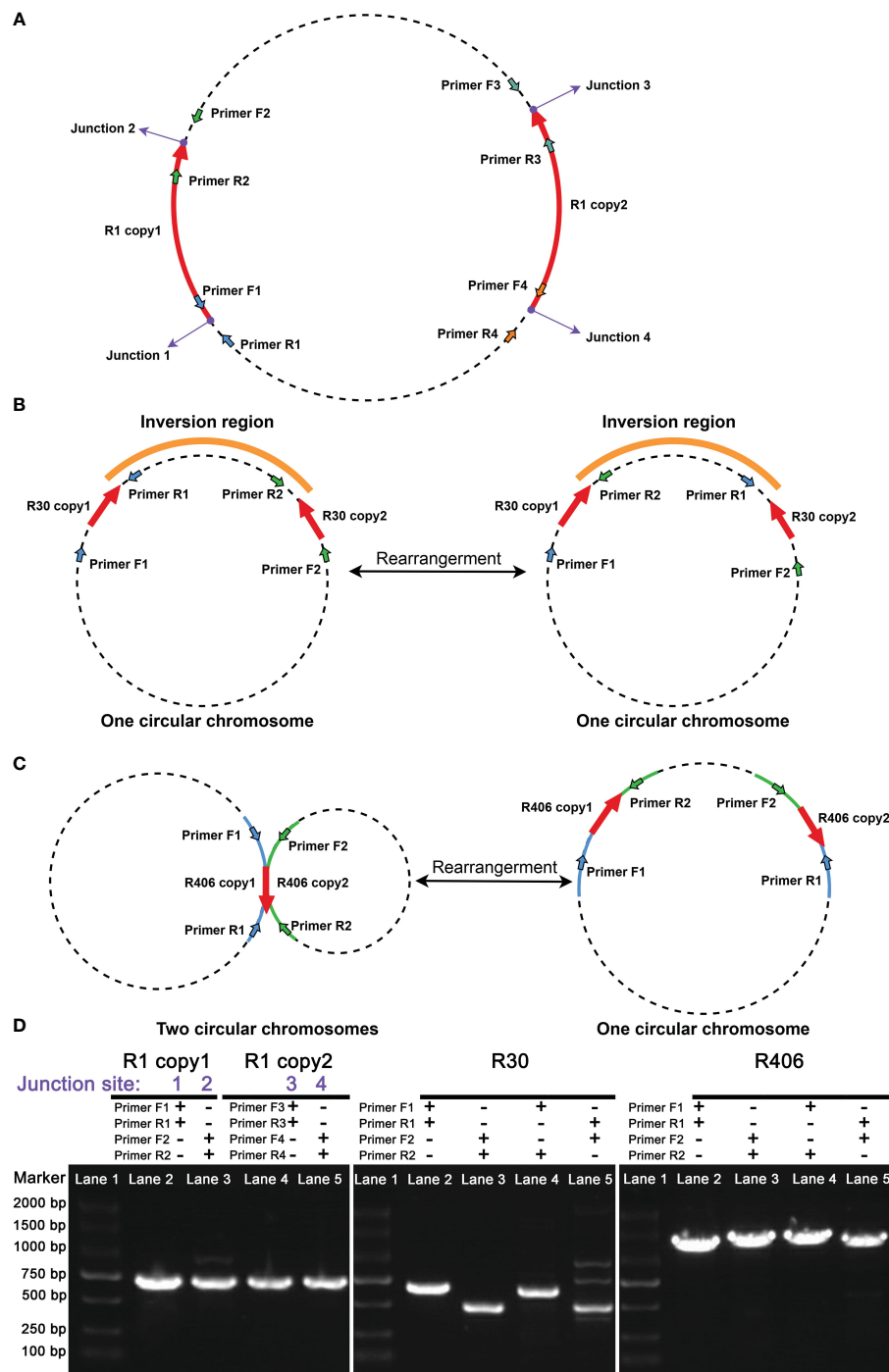


FIGURE 2 Validation of the repeat mediated recombination products in *C. arabica* mitochondrial genome. The purple dots represent junction sites. The red arrows represent the intervals and directions of the repeated sequences. Arrows of other colors represent primer positions. The orange line segment represents the interval that can be inverted. **(A)** The primers were designed for the junction site validation of the R1. Since the repeat sequence, R1, is longer than 5000 bp, it was difficult to obtain PCR products spanning the entire repetitive sequence. Therefore, we only performed PCR to verify the boundaries of R1. **(B)** The primers were designed to validate the major and minor conformations of R30. **(C)** The primers were designed for the major and minor conformations of R406. **(D)** The electrophoresis results of the PCR products for the fragments associated with R1, R30, and R406. The lane numbers and PCR primer names are shown above each panel.

3.7 RNA editing site identification

Using the Deepred-mt program, 472 RNA editing events were predicted (Table S9). Of these 472 predicted RNA editing events,

the *nad4* gene, *ccmB* gene and *ccmC* gene were edited most frequently, with 36, 34 and 33 editing site respectively. The *sdh3* gene was edited the least for only once (Figure S46). To further explore the amino acid changes before and after editing, we counted

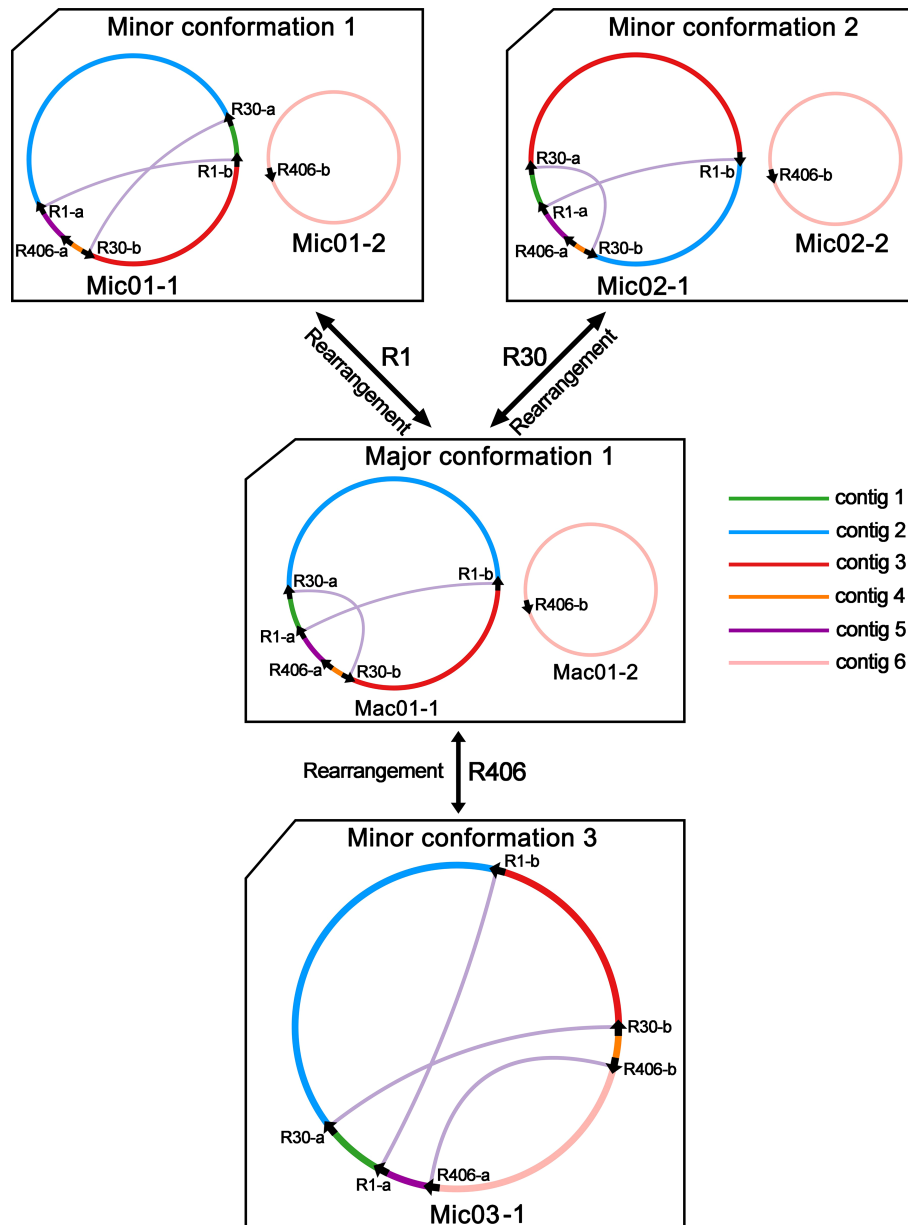


FIGURE 3

Hypothetical products generated by recombination mediated by repeats R1, R30, and R406 on the chromosomes of *C. arabica* mitochondrial genome. The repeats mediate the structure of the hypothetical recombination products. The circles represent the mitochondrial genome, and the different colored lines represent the regions between the repeated sequences. The black arrows represent the repeat sequence. The original electropherogram of the PCR amplification can be found in [Supplementary File 5](#).

the amino acid composition before and after RNA editing. Of these 472 RNA editing events, mainly non-synonymous substitutions occurred for 447 times, and synonymous substitutions occurred for a total of 25 times. The most abundant amino acid conversion was Ser to Leu, with a total of 110 times. In addition to this, we found that the *atp6* (CAA to UAA) and the *atp9* genes (CGA to UGA) gained a stop codon through the RNA editing event (Figure S47).

To validate that these predicted RNA editing sites were real, we used PCR to amplify all the PCGs of the coffee mitochondrial genome from genomic DNA and cDNA. The products were then sequenced using the Sanger method. A comparison of the results helped us to identify the RNA editing sites. The primers used for the

RNA editing site validation can be found in [Table S2](#). Finally, we obtained 103 sanger sequencing results of short DNA fragments for the 21 PCGs. We identified 90 RNA-edited events in 14 PCGs. All RNA editing events were of “C to U” and “G to A” types (Figure 6; Figures S48, 49). The “G to A” type is a symmetric type of the “C to U” type, as discussed before (Wu et al., 2017).

Among the 90 RNA editing events, 87 were non-synonymous substitutions, and only three (*ccmFC-378*, *nad2-252*, and *nad7-531*) were synonymous substitutions. Among the synonymous substitution, the *ccmFC-378* and *nad2-252* edited the codon from UUC (F) to UUU (F), and *nad7-531* edited the codon from UCC (S) to UCU (S). Most of the resulting amino acid changes were from Ser

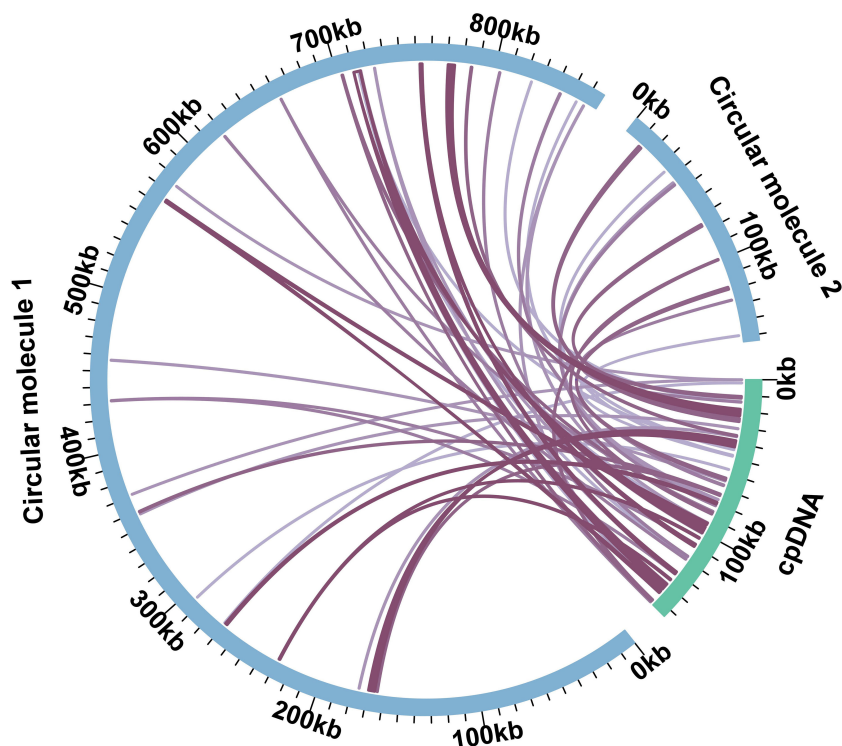


FIGURE 4

The DNA transfer between the *C. arabica* mitochondrial genome and plastome. The blue arcs represent the mitochondrial genome, and the green arcs represent the chloroplast genome. The purple link in the middle connects the homologous regions between the mitochondrial genome and the plastome.

to Leu, which occurred 25 times. The *mttB* and *nad2* genes had the most edited sites, including fourteen (Table S10). All the RNA editing sites of *ccmFc* were in the first exon. The *cox2* gene had only one RNA editing site: *cox2*-278. All the Sanger sequencing results of gDNA and cDNA can be found in the Supplementary File 4. The chromatograms of the 90 RNA-editing sites are shown in Figure 6 and Figures S48, 49. Interestingly, all these 87 validated sites were successfully predicted by the Deep-mt program. The three sites with synonymous substitutions were not successfully predicted by the Deep-mt program.

4 Discussion

4.1 Mitochondrial genome assembly is the current bottleneck in plant mitochondrial genome study

Studies on plant mitochondrial genomes lags behind that of cp genomes due to the relatively complex structures of plant mitochondrial genomes (Gualberto et al., 2014; Skippington et al., 2015; Skippington et al., 2017). It has been shown that the plant mitochondrial genomes might contain multiple-circular, linear, and branched chromosomes (Kozik et al., 2019; Li et al., 2021; Yang et al., 2022). These intricate structures contribute to challenges in assembling plant genomes.

Two assembly strategies have been proposed for mitochondrial genomes assembly. The first one is similar to the nuclear genome

assembly strategy, which uses long reads for *de novo* assembly and short reads for polishing the assembled results (Yu et al., 2011). However, this approach has two major limitations: (1) it cannot restore the complex structure of the mitochondrial genome and often yields a simplified assembly result, such as a single circular molecule (Liu J. et al., 2023); (2) it can introduce false positive results due to the presence of homologous DNA fragments shared between the mitochondrial genomes and the nuclear genomes and the plastomes, such as NUMTs and MTPTs. For example, the reads derived from plastid DNA fragments homologous to MTPT sequences might be used to polish the MTPT sequences, mistakenly changing the MTPT sequences (Yang H. et al., 2022; Jiang et al., 2023; Yang et al., 2023).

A second strategy, which is called the hybrid strategy, has been developed to overcome the above limitations. This strategy assembled the short reads into a unitig graph with different kmer lengths (Jin et al., 2020). Then, the long reads were used to resolve the double bifurcating structures in the unitig graph caused by repetitive sequences in the genome (Park et al., 2018; Wang et al., 2018; Yang et al., 2018; Kim et al., 2020). In our study, we obtained the short mitochondrial genome reads by employing the seed reads extension method through the GetOrganelle software. We utilized this hybrid assembly approach to assemble the *C. arabica* mitochondrial genome, which we discovered contains two chromosomes. We advocate using hybrid strategies to unravel the major conformations of complex plant mitochondrial genome structures.

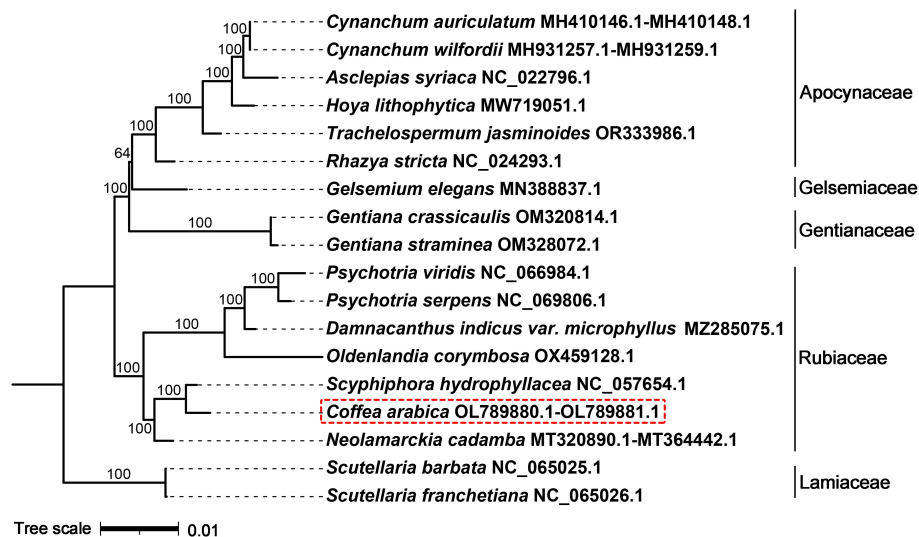


FIGURE 5

Phylogenetic relationships of *C. arabica* and other 17 species. The black branches represent the phylogenetic relationships constructed by the shared mitochondrial genes. The number on each branch node represents the bootstrap number. The mitochondrial genome NCBI accession number for each species follows the Latin name. The families of these species are shown on the right side.

4.2 The complex structure and representation of the plant mitochondrial genomes

Previous studies have shown that the mitochondrial genomes might have circular, linear and multi-branch structures (Gualberto et al., 2014; Chevigny et al., 2020). We found that the coffee mitochondrial genomes can be represented as a unitig graph and two independent chromosomes after splitting the unitig graph based on the mapping results of long reads. However, representing the complex structures of the mitochondrial genomes is a challenging problem. There are two primary methods for addressing this issue.

The first method uses a graph to represent the intricate structure of plant mitochondrial genomes (graph-based method). After assembling the long reads into a graph, the mitochondrial genome structure is represented as multiple connected contigs (Li et al., 2022). This approach has been employed in numerous studies. For example, the mitochondrial genomes of the three species, *Selaginella nipponica*, *Abelmoschus esculentus*, and *Picea sitchensis*, all use this format to represent their mitochondrial genomes and annotate the mitochondrial genomes according to the corresponding mitochondrial contigs (Jackman et al., 2020; Kang et al., 2020; Li et al., 2022).

The second method follows the convention of using circular or linear conformations to represent the mitochondrial genomes (conformation-based method). This approach is based on the understanding that a genome may consist of multiple conformations. Long reads can resolve the unitig graph, resulting in conformations that can undergo homologous recombination mediated by repetitive sequences (André et al., 1992; Alverson et al., 2011). This process can lead to various alternative genomes conformations, typically of low-frequency and referred to as minor

conformations. For example, in the *Prunus salicina* mitochondrial genome (Fang et al., 2021), nine repeat sequences were involved in homologous recombination, resulting in two low-frequency chromosomes. Similar findings were observed in the mitochondrial genomes of *Ipomoea batatas* (Yang Z. et al., 2022) and *Scutellaria tsinyunensis* (Li et al., 2021), where direct repeats led to the division of the mitochondrial genome into smaller chromosomes corresponding to minor conformations. In the *Cannabissativa* mitochondrial genome (Liu J. et al., 2023), the major conformation is a circular molecule. However, it can form multiple minor conformations through recombinations mediated by 15 repeat sequences. *Salvia officinalis* has three repeat sequences in the genus *Salvia* (Yang et al., 2023), which can mediate homologous recombination, while there are nine pairs of such repeat sequences in *S. multiorrhiza* (Yang H. et al., 2022). *Taraxacum mongolicum* has five pairs of repeat sequences confirmed to mediate homologous recombination, producing multiple minor conformations (Jiang et al., 2023).

Both methods have some limitations. Firstly, there are several issues with the graph-based representation of the mitochondrial genome. Until now, most analytic methods and tools for genomes were developed based on the assumption that the genome is a simple circular or linear molecule. Very few methods and tools can be used to analyze a genome represented by a graph (Kang et al., 2020). Secondly, the graph representations of the plant mitochondrial genomes are mostly generated from bioinformatics analysis (Kang et al., 2020). Although the repeat sequences involved could be validated by PCR amplification and Sanger sequencing (Lai et al., 2022), the complete graphic structures of the plant mitochondrial genomes are difficult to validate. Lastly, there is a lack of comparative tools to compare genomes represented as graph. As a result, it would be difficult to conduct comparative genomic studies if the genomes were represented with graph.

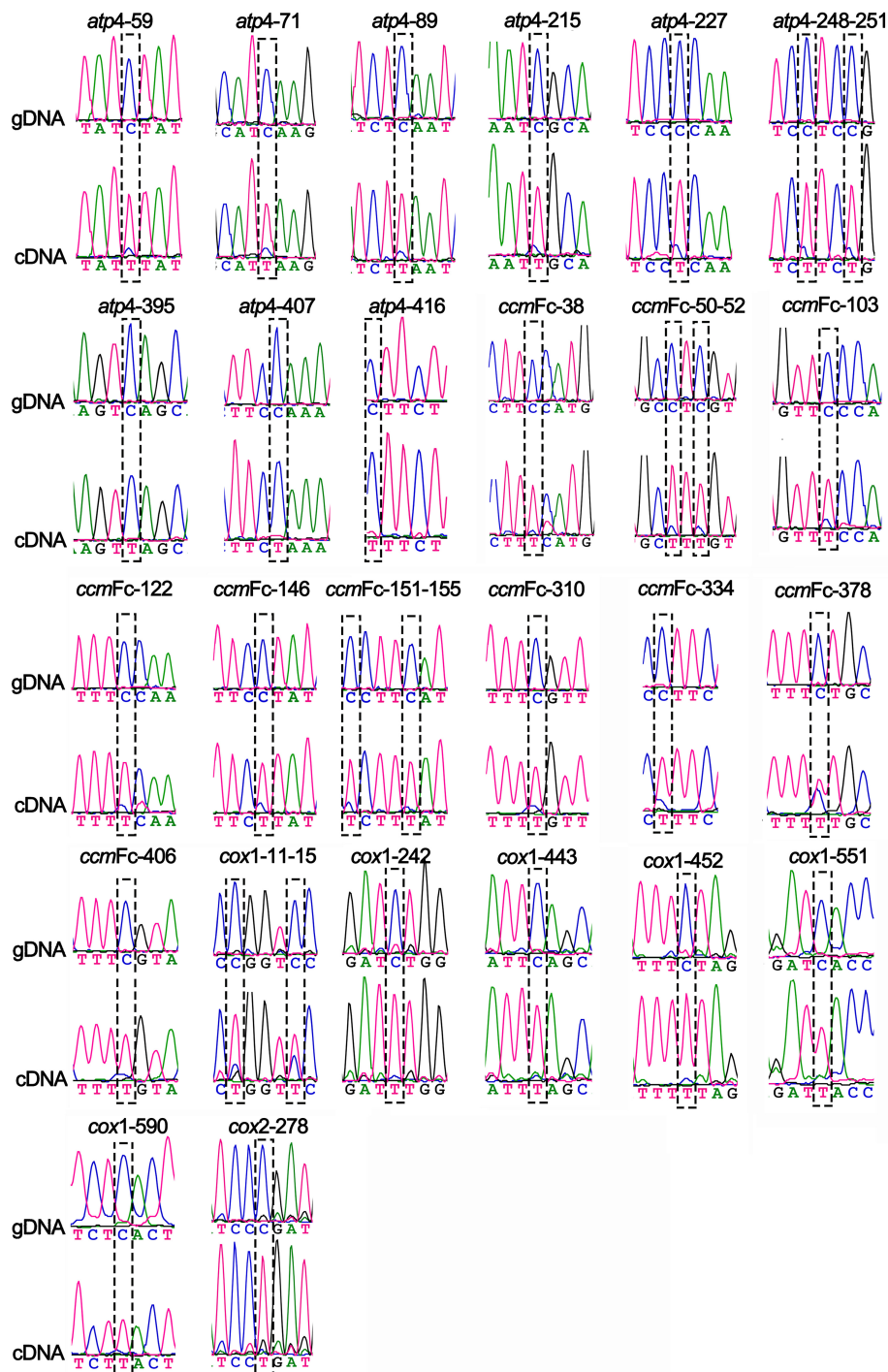


FIGURE 6
Validation of RNA editing sites in the *atp4*, *ccmFc*, *cox1*, and *cox2* genes of *C. arabica* mitochondrial genome. Chromatograms showing the sequences before and after editing at the hypothetical RNA-editing sites. For each RNA editing site, the name is shown on the top. The results from the genomic DNA (gDNA) and complementary DNA (cDNA) are shown in the middle and bottom. Black dashed rectangles framed the RNA editing sites.

Secondly, there are several issues with the conformation-based representation of the mitochondrial genome. Firstly, when resolving the unitig graph, the most abundant conformation was selected at each DBS point. However, whether or not the most abundant conformation forms a chromosome remains to be confirmed.

Secondly, the dependence among different repetitive sequences for recombination is not clear. At present time, the minor conformations were generating based on the assumption that the repeat-mediated recombinations are independent of each other. As a result, the exact set of conformation cannot be determined. As a

result, the representation of the complex mitochondrial genome structures requires further investigation (de Abreu et al., 2023).

4.3 The RNA editing sites “C” to “U.”

RNA editing is a widely observed phenomenon in plant mitochondria (Varré et al., 2019). This process involves modifying the information in transcripts of nearly all angiosperm mitochondrial protein-coding genes (Mower and Palmer, 2006). In our study, we first identified the RNA editing sites using a convoluted neural network-based method: deepred-mt. we validated 90 RNA editing events in the coffee mitochondrial genome, which involved the conversion of cytidine (C) to uridine (U). We had not been able to obtain unambiguous sequencing results for the other nine unigenes after multiple attempts. The likely reasons are that the amplification of the sequences is interfered with highly homologous sequences such as MTPT or NUMT and etc.

Studies on the model plant *Arabidopsisthaliana* have reported more than 400 RNA editing events involving the substitution of cytidines with uridines in the mitochondrial genome (Bentolila et al., 2013). In *T. mongolicum*, a total of 278 RNA-editing sites were predicted and 213 were validated (Jiang et al., 2023). In *C. sativa* mitochondrion, RNA editing was found to be tissue specific (Liu J. et al., 2023). Using PCR amplification and Sanger sequencing methods, 113 of the 126 RNA editing sites from 11 PCGs were validated in the *S. officinalis*. In the *S. miltiorrhiza*, 225 “C to U” sites in the protein coding regions were discovered (Wu et al., 2017). In the present study, we predicted 472 RNA editing sites, all of which were edited from C to U, which is consistent with the number and type of RNA editing times previously reported.

During plant mitochondrial RNA editing, non-synonymous and synonymous substitutions have different biological significance and impact (Lu et al., 1998). Non-synonymous substitutions involve changes in the amino acid sequence and, therefore, may affect the structure and function of the protein. These changes may result in enhanced, diminished, or complete loss of protein function (Yates and Sternberg, 2013). Of the 472 predicted sites, 447 were non-synonymous. Of the validated 90 sites, 87 were non-synonymous. Only *nad7-531*, *ccmFC-378*, and *nad2-252* were synonymous. Notably, we found that the stop codons of *atp6* and *atp9* genes were created by RNA editing events from the prediction results of Deepred-mt program. Previous study reported the *atp6* gene premature termination expression in maize, sorghum and *Oenothera* (Kumar and Levings, 1993). The transcripts of *atp9* gene have a stop codon create by the RNA editing events in rapeseed (Handa, 1993). There may be the same possibility of that the same type of (non-synonymous substitution) editing occurred in the coffee. The function of these non-synonymous substitutions, if there are any, still needs to be further investigated.

5 Conclusions

In this study, we sequenced and analyzed the *C. arabica* mitochondrial genome. We identified 40 PCGs, 3 rRNA genes, 26

tRNA genes, 270 SSRs, and 34 tandem repeats. In particular, we found three repeats mediating recombination, 54 fragments originating from the chloroplast genome, and 90 RNA-editing sites. We showed that the *C. arabica* mitochondrial genome had a complex structure caused by a plethora of molecular mechanisms.

Data availability statement

The organelle sequences supporting the conclusions of this article are available in GenBank (<https://www.ncbi.nlm.nih.gov/>) with accession numbers: OL789880-OL789881 (mitochondrial genome), OL789882 (cpgenome), respectively. The sample was deposited in the Institute of Medicinal Plant Development (Beijing, China) with accession numbers 1KMG_CF01. The raw data has been released through GenBank with the following accession numbers: (1) Nanopore DNA reads BioProject PRJNA792668, BioSample SAMN24449604, SRA database SRR17345007. (2) Illumina DNA reads: BioProject PRJNA792714, BioSample SAMN24451482, SRA database SRR17345023.

Ethics statement

We collected fresh leaf materials of *Coffea arabica* for this study. The Institute of Medicinal Plants identified the plant sample with Yang Ni. We prepared and deposited the voucher specimens in the Institute of Medicinal Plant Development (Beijing, China) with accession numbers 1KMG_CF01. The study, including plant sample collection, complies with relevant institutional, national, and international guidelines and legislation. No specific permits are required for plant collection.

Author contributions

CL: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Visualization, Writing – review & editing. YN: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. XZ: Formal Analysis, Resources, Software, Validation, Writing – review & editing. JL: Formal Analysis, Investigation, Methodology, Software, Writing – review & editing. QL: Formal Analysis, Methodology, Validation, Visualization, Writing – review & editing. HC: Data curation, Formal Analysis, Supervision, Validation, Writing – review & editing. BM: Formal Analysis, Validation, Writing – review & editing.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. The study

was supported by the National Natural Science Foundation of China [81872966], CAMS Innovation Fund for Medical Sciences (CIFMS) (2021-I2M-1-022), and the National Science & Technology Fundamental Resources Investigation Program of China [2018FY100705]. The funders were not involved in the study design, data collection, analysis, publication decision, or manuscript preparation.

Conflict of interest

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

References

- Alverson, A. J., Zhuo, S., Rice, D. W., Sloan, D. B., and Palmer, J. D. (2011). The mitochondrial genome of the legume *Vigna radiata* and the analysis of recombination across short mitochondrial repeats. *PLoS One* 6 (1), e16404. doi: 10.1371/journal.pone.0016404
- André, C., Levy, A., and Walbot, V. (1992). Small repeated sequences and the structure of plant mitochondrial genomes. *Trends Genet.* 8 (4), 128–132. doi: 10.1016/0168-9525(92)90370-J
- Araya, A., Begu, D., Crabier, F., Mouras, A., Litvak, S., and Blanc, V. (1995). RNA editing in wheat mitochondria. *Biochimie* 77 (1–2), 87–91. doi: 10.1016/0300-9084(96)88109-0
- Backert, S., Nielsen, B. L., and Börner, T. (1997). The mystery of the rings: structure and replication of mitochondrial genomes from higher plants. *Trends Plant Sci.* 2 (97), 477–483. doi: 10.1016/S1360-1385(97)01148-5
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19 (5), 455–477. doi: 10.1089/cmb.2012.0021
- Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33 (16), 2583–2585. doi: 10.1093/bioinformatics/btx198
- Bendich, A. J. (1996). Structural analysis of mitochondrial DNA molecules from fungi and plants using moving pictures and pulsed-field gel electrophoresis. *J. Mol. Biol.* 255 (4), 564–588. doi: 10.1006/jmbi.1996.0048
- Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27 (2), 573–580. doi: 10.1093/nar/27.2.573
- Bentolila, S., Oh, J., Hanson, M. R., and Bukowski, R. (2013). Comprehensive high-resolution analysis of the role of an Arabidopsis gene family in RNA editing. *PLoS Genet.* 9 (6), e1003584. doi: 10.1371/journal.pgen.1003584
- Bhattarai, G., Shi, A., Kandel, D. R., Solis-Gracia, N., da Silva, J. A., and Avila, C. A. (2021). Genome-wide simple sequence repeats (SSR) markers discovered from whole-genome sequence comparisons of multiple spinach accessions. *Sci. Rep.* 11 (1), 9999–9999. doi: 10.1038/s41598-021-89473-0
- Bhupathiraju, S. N., Pan, A., Manson, J. E., Willett, W. C., van Dam, R. M., and Hu, F. B. (2014). Changes in coffee intake and subsequent risk of type 2 diabetes: three large cohorts of US men and women. *Diabetologia* 57 (7), 1346–1354. doi: 10.1007/s00125-014-3235-7
- Brenner, W. G., Mader, M., Müller, N. A., Hoenicka, H., Schroeder, H., Zorn, I., et al. (2019). High level of conservation of mitochondrial RNA editing sites among four populus species. *G3 (Bethesda)* 9 (3), 709–717. doi: 10.1534/g3.118.200763
- Cano-Marquina, A., Tarin, J. J., and Cano, A. (2013). The impact of coffee on health. *Maturitas* 75 (1), 7–21. doi: 10.1016/j.maturitas.2013.02.002
- Chen, C., Chen, H., Zhang, Y., Thomas, H. R., and Xia, R. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13 (8), 1194–1202. doi: 10.1016/j.molp.2020.06.009
- Chen, Y., Wei, X., Lin, G., Liu, Y., and Zhang, Y. (2020). The complete mitochondrial genome of an endangered mangrove plant: *Scyphiphora hydrophyllacea*. *Mitochondrial DNA Part B* 5 (3), 2779–2780. doi: 10.1080/23802359.2020.1788460
- Chen, Y., Ye, W., Zhang, Y., and Xu, Y. (2015). High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic Acids Res.* 43 (16), 7762–7768. doi: 10.1093/nar/gkv784
- Cheng, Y., He, X., Priyadarshani, S., Wang, Y., Ye, L., Shi, C., et al. (2021). Assembly and comparative analysis of the complete mitochondrial genome of *Suaeda glauca*. *BMC Genomics* 22 (1), 167. doi: 10.1186/s12864-021-07490-9
- Chevigny, N., Schatz-Daas, D., Lotfi, F., and Gualberto, J. M. (2020). DNA repair and the stability of the plant mitochondrial genome. *Int. J. Mol. Sci.* 21 (1), 328. doi: 10.3390/ijms21010328
- Covello, P. S., and Gray, M. W. (1989). RNA editing in plant mitochondria. *Nature* 341 (6243), 662–666. doi: 10.1038/341662a0
- de Abreu, V. A., Alves, R. M., Silva, S. R., Ferro, J. A., Domingues, D. S., Miranda, V. F., et al. (2023). Comparative analyses of *Theobroma cacao* and *T. grandiflorum* mitogenomes reveal conserved gene content embedded within complex and plastic structures. *Gene* 849, 146904. doi: 10.1016/j.gene.2022.146904
- Denoeud, F., Carretero-Paulet, L., Dereeper, A., Droc, G., Guyot, R., Pietrella, M., et al. (2014). The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* 345, 1181–1184. doi: 10.1126/science.1255274
- Edera, A. A., Gandini, C. L., and Sanchez-Puerta, M. V. (2018). Towards a comprehensive picture of C-to-U RNA editing sites in angiosperm mitochondria. *Plant Mol. Biol.* 97, 215–231. doi: 10.1007/s11103-018-0734-9
- Edera, A. A., Small, I., Milone, D. H., and Sanchez-Puerta, M. V. (2021). Deepred-Mt: Deep representation learning for predicting C-to-U RNA editing in plant mitochondria. *Comput. Biol. Med.* 136, 104682. doi: 10.1016/j.combiomed.2021.104682
- Fang, B., Li, J., Zhao, Q., Liang, Y., and Yu, J. (2021). Assembly of the complete mitochondrial genome of chinese plum (*Prunus salicina*): characterization of genome recombination and RNA editing sites. *Genes (Basel)* 12 (12), 1970. doi: 10.3390/genes12121970
- Ferreira, T., Shuler, J., Guimaraes, R., and Farah, A. (2019). "Introduction to coffee plant and genetics," in *Coffee: Production, Quality and Chemistry*. Eds. A. Farah and A. Farah. (The Royal Society of Chemistry). doi: 10.1039/9781782622437-00001
- Gott, J. M., and Emeson, R. B. (2000). Functions and mechanisms of RNA editing. *Annu. Rev. Genet.* 34 (1), 499–531. doi: 10.1146/annurev.genet.34.1.499
- Grewe, F., Edger, P. P., Keren, I., Sultan, L., Pires, J. C., Ostersetzter-Biran, O., et al. (2014). Comparative analysis of 11 Brassicales mitochondrial genomes and the mitochondrial transcriptome of Brassica oleracea. *Mitochondrion* 19 Pt B, 135–143. doi: 10.1016/j.mito.2014.05.008
- Grimes, B. T., Sisay, A. K., Carroll, H. D., and Cahoon, A. B. (2014). Deep sequencing of the tobacco mitochondrial transcriptome reveals expressed ORFs and numerous editing sites outside coding regions. *BMC Genomics* 15, 31. doi: 10.1186/1471-2164-15-31
- Gualberto, J. M., Lamattina, L., Bonnard, G., Weil, J.-H., and Grienenberger, J.-M. (1989). RNA editing in wheat mitochondria results in the conservation of protein sequences. *Nature* 341 (6243), 660–662. doi: 10.1038/341660a0
- Gualberto, J. M., Mileshina, D., Wallet, C., Niazi, A. K., Weber-Lotfi, F., and Dietrich, A. (2014). The plant mitochondrial genome: dynamics and maintenance. *Biochimie* 100, 107–120. doi: 10.1016/j.biochi.2013.09.016
- Gualberto, J. M., and Newton, K. J. (2017). Plant mitochondrial genomes: dynamics and mechanisms of mutation. *Annu. Rev. Plant Biol.* 68, 225–252. doi: 10.1146/annurev-arplant-043015-112232
- Han, E.-K., Cho, W.-B., Tamaki, I., Choi, I.-S., and Lee, J.-H. (2021). Comparative mitogenomic analysis reveals gene and intron dynamics in Rubiaceae and intra-specific diversification in *damnacanthus indicus*. *Int. J. Mol. Sci.* 22 (13), 7237. doi: 10.3390/ijms22137237
- Handa, H. (1993). RNA editing of rapeseed mitochondrial atp9 transcripts: RNA editing changes four amino acids, but termination codon is already encoded by genomic sequence. *Japan. J. Genet.* 68 (1), 47–54. doi: 10.1266/jjg.68.47

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1261012/full#supplementary-material>

- Handa, H. (2003). The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. *Nucleic Acids Res.* 31 (20), 5907–5916. doi: 10.1093/nar/gkg795
- Hazkani-Covo, E., Zeller, R. M., and Martin, W. (2010). Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet.* 6 (2), e1000834. doi: 10.1371/journal.pgen.1000834
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2017). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35 (2), 518–522. doi: 10.1101/153916
- Hoch, B., Maier, R. M., Appel, K., Igloi, G. L., and Kössel, H. (1991). Editing of a chloroplast mRNA by creation of an initiation codon. *Nature* 353 (6340), 178–180. doi: 10.1038/353178a0
- Jackman, S. D., Coombe, L., Warren, R. L., Kirk, H., Trinh, E., MacLeod, T., et al. (2020). Complete mitochondrial genome of a gymnosperm, Sitka spruce (*Picea sitchensis*), indicates a complex physical structure. *Genome Biol. Evol.* 12 (7), 1174–1179. doi: 10.1093/gbe/evaa108
- Jang, W., Lee, H. O., Lee, J. W., Kwon, N., Kim, D. H., Bang, K. H., et al. (2021). The complete mitochondrial genome of Panax ginseng (Apiales, Araliaceae): an important medicinal plant. *Mitochondrial. DNA B Resour* 6 (10), 3080–3081. doi: 10.1080/23802359.2021.1981167
- Jiang, M., Ni, Y., Li, J., and Liu, C. (2023). Characterisation of the complete mitochondrial genome of *Taraxacum mongolicum* revealed five repeat-mediated recombinations. *Plant Cell Rep.* 42 (4), 775–789. doi: 10.1007/s00299-023-02994-y
- Jin, J.-J., Yu, W.-B., Yang, J.-B., Song, Y., dePamphilis, C. W., Yi, T.-S., et al. (2020). GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *Genome Biol.* 21 (1), 241. doi: 10.1186/s13059-020-02154-5
- Julca, I., Mutwil-Anderwald, D., Manoj, V., Khan, Z., Lai, S. K., Yang, L. K., et al. (2023). Genomic, transcriptomic, and metabolomic analysis of *Oldenlandia corymbosa* reveals the biosynthesis and mode of action of anti-cancer metabolites. *J. Integr. Plant Biol.* 65 (6), 1442–1466. doi: 10.1111/jipb.13469
- Kang, B.-C., Bae, S.-J., Lee, S., Lee, J. S., Kim, A., Lee, H., et al. (2021). Chloroplast and mitochondrial DNA editing in plants. *Nat. Plants* 7 (7), 899–905. doi: 10.1038/s41477-021-00943-9
- Kang, J. S., Zhang, H. R., Wang, Y. R., Liang, S. Q., Mao, Z. Y., Zhang, X. C., et al. (2020). Distinctive evolutionary pattern of organelle genomes linked to the nuclear genome in Selaginellaceae. *Plant J.* 104 (6), 1657–1672. doi: 10.1111/tpj.15028
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30 (4), 772–780. doi: 10.1093/molbev/mst010
- Kim, C.-K., Jin, M.-W., and Kim, Y.-K. (2020). The complete mitochondrial genome sequences of *Bupleurum falcatum* (Apiales: Apiaceae). *Mitochondrial. DNA Part B Resour.* 5, 2576–2577. doi: 10.1080/23802359.2020.1781566
- Kitazaki, K., and Kubo, T. (2010). Cost of having the largest mitochondrial genome: evolutionary mechanism of plant mitochondrial genome. *J. Bot.* 2010, 620137. doi: 10.1155/2010/620137
- Kolahdouzan, M., and Hamadeh, M. J. (2017). The neuroprotective effects of caffeine in neurodegenerative diseases. *CNS Neurosci. Ther.* 23 (4), 272–290. doi: 10.1111/cns.12684
- Kozik, A., Rowan, B. A., Lavelle, D., Berke, L., Schranz, M. E., Michelmore, R. W., et al. (2019). The alternative reality of plant mitochondrial DNA: One ring does not rule them all. *PLoS Genet.* 15 (8), e1008373. doi: 10.1371/journal.pgen.1008373
- Kumar, R., and Levings, C. S. (1993). RNA editing of a chimeric maize mitochondrial gene transcript is sequence specific. *Curr. Genet.* 23 (2), 154–159. doi: 10.1007/BF00352015
- Lai, C., Wang, J., Kan, S., Zhang, S., Li, P., Reeve, W. G., et al. (2022). Comparative analysis of mitochondrial genomes of *Broussonetia* spp. (Moraceae) reveals heterogeneity in structure, synteny, intercellular gene transfer, and RNA editing. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1052151
- Lenz, H., Hein, A., and Knoop, V. (2018a). Plant organelle RNA editing and its specificity factors: enhancements of analyses and new database features in PREPACT 3.0. *BMC Bioinf.* 19 (1), 1–18. doi: 10.1186/s12859-018-2244-9
- Lenz, H., Hein, A., and Knoop, V. (2018b). Plant organelle RNA editing and its specificity factors: enhancements of analyses and new database features in PREPACT 3.0. *BMC Bioinf.* 19 (1), 255. doi: 10.1186/s12859-018-2244-9
- Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47 (W1), W256–W259. doi: 10.1093/nar/gkz239
- Lewis, S. E., Searle, S., Harris, N., Gibson, M., Iyer, V., Richter, J., et al. (2002). Apollo: a sequence annotation editor. *Genome Biol.* 3, research0082.1. doi: 10.1186/gb-2002-3-12-research0082
- Li, J., Li, J., Ma, Y., Kou, L., Wei, J., and Wang, W. (2022). The complete mitochondrial genome of okra (*Abelmoschus esculentus*): Using nanopore long reads to investigate gene transfer from chloroplast genomes and rearrangements of mitochondrial DNA molecules. *BMC Genomics* 23 (1), 481. doi: 10.1186/s12864-022-08706-2
- Li, J., Xu, Y., Shan, Y., Pei, X., Yong, S., Liu, C., et al. (2021). Assembly of the complete mitochondrial genome of an endemic plant, *Scutellaria tsinyunensis*, revealed the existence of two conformations generated by a repeat-mediated recombination. *Planta* 254 (2), 36. doi: 10.1007/s00425-021-03684-3
- Liu, S., Ni, Y., Li, J., Zhang, X., Yang, H., Chen, H., et al. (2023). CPGView: A package for visualizing detailed chloroplast genome structures. *Mol. Ecol. Resour.* 23 (3), 694–704. doi: 10.1111/1755-0998.13729
- Liu, J., Ni, Y., and Liu, C. (2023). Polymeric Structure of the *Cannabis Sativa* L. mitochondrial genome identified with an assembly graph model. *Gene* 853, 147081. doi: 10.1016/j.gene.2022.147081
- Lu, M.-Z., Szmidt, A. E., and Wang, X.-R. (1998). RNA editing in gymnosperms and its impact on the evolution of the mitochondrial *coxI* gene. *Plant Mol. Biol.* 37, 225–234. doi: 10.1023/A:1005972513322
- Margulis, L., and Bermudes, D. (1985). Symbiosis as a mechanism of evolution: status of cell symbiosis theory. *Symbiosis* 1, 101–124.
- Mekbib, Y., Tesfaye, K., Dong, X., Saina, J. K., Hu, G. W., and Wang, Q. F. (2022). Whole-genome resequencing of *Coffea arabica* L. (Rubiaceae) genotypes identify SNP and unravels distinct groups showing a strong geographical pattern. *BMC Plant Biol.* 22 (1), 69. doi: 10.1186/s12870-022-03449-4
- Michael, T., Pascal, L., Tommaso, P., Ulbricht-Jones, E. S., Axel, F., Ralph, B., et al. (2017). GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45 (W1), W6–W11. doi: 10.1093/nar/gkx391
- Milne, I., Stephen, G., Bayer, M., Cock, P. J. A., Pritchard, L., Cardle, L., et al. (2012). Using Tablet for visual exploration of second-generation sequencing data. *Briefings Bioinf.* 14 (2), 193–202. doi: 10.1093/bib/bbs012
- Min, J., Kim, Y., Xi, H., Heo, K. I., and Park, J. (2019). The complete chloroplast genome of coffee tree, *Coffea arabica* L. 'Typica' (Rubiaceae). *Mitochondrial. DNA B Resour* 4 (2), 2240–2241. doi: 10.1080/23802359.2019.1624213
- Mower, J. P. (2005). PREP-Mt: predictive RNA editor for plant mitochondrial genes. *BMC Bioinf.* 6, 96. doi: 10.1186/1471-2105-6-96
- Mower, J. P., and Palmer, J. D. (2006). Patterns of partial RNA editing in mitochondrial genes of *Beta vulgaris*. *Mol. Genet. Genomics* 276 (3), 285–293. doi: 10.1007/s00438-006-0139-3
- Ni, Y., Li, J., Chen, H., Yue, J., Chen, P., and Liu, C. (2022). Comparative analysis of the chloroplast and mitochondrial genomes of *Saposhnikovia divaricata* revealed the possible transfer of plastome repeat regions into the mitogenome. *BMC Genomics* 23 (1), 570. doi: 10.1186/s12864-022-08821-0
- Noir, S., Patheyron, S., Combes, M. C., Lashermes, P., and Chalhou, B. (2004). Construction and characterisation of a BAC library for genome analysis of the allotetraploid coffee species (*Coffea arabica* L.). *Theor. Appl. Genet.* 109 (1), 225–230. doi: 10.1007/s00122-004-1604-1
- Owczarzy, R., Tataurov, A. V., Wu, Y., Manthey, J. A., McQuisten, K. A., Almabrazi, H. G., et al. (2008). IDT SciTools: a suite for analysis and design of nucleic acid oligomers. *Nucleic Acids Res.* 36 (suppl_2), W163–W169. doi: 10.1093/nar/gkn198
- Park, J., Kim, Y., Xi, H., and Heo, K. I. (2019). The complete chloroplast genome of coffee tree, *Coffea arabica* L. 'Blue Mountain' (Rubiaceae). *Mitochondrial. DNA B Resour* 4 (2), 2436–2437. doi: 10.1080/23802359.2019.1636729
- Park, I., Yang, S., Kim, W. J., Noh, P., Lee, H. O., and Moon, B. C. (2018). The complete chloroplast genome of *Cuscuta pentagona* Engelm. *Mitochondrial. DNA Part B* 3 (2), 523–524. doi: 10.1080/23802359.2018.1467229
- Picardi, E., Horner, D. S., Chiara, M., Schiavon, R., Valle, G., and Pesole, G. (2010). Large-scale detection and analysis of RNA editing in grape mtDNA by RNA deep-sequencing. *Nucleic Acids Res.* 38 (14), 4755–4767. doi: 10.1093/nar/gkq202
- Poole, A. M., and Penny, D. (2007). Evaluating hypotheses for the origin of eukaryotes. *Bioessays* 29 (1), 74–84. doi: 10.1002/bies.20516
- Scott, I., and Logan, D. C. (2007). "Mitochondrial Dynamics: The Control of Mitochondrial Shape, Size, Number, Motility, and Cellular Inheritance." in *Annual Plant Reviews Volume 31: Plant Mitochondria*. 2007, 1–35.
- Shi, L., Chen, H., Jiang, M., Wang, L., Wu, X., Huang, L., et al. (2019). CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* 47 (W1), W65–W73. doi: 10.1093/nar/gkz345
- Shtolz, N., and Dan, M. (2019). The mitochondrial genome—on selective constraints and signatures at the organism, cell, and single mitochondrion levels. *Front. Ecol. Evol.* 7. doi: 10.3389/fevo.2019.00342
- Skippington, E., Barkman, T. J., Rice, D. W., and Palmer, J. D. (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), E3515–E3524. doi: 10.1073/pnas.1504491112
- Skippington, E., Barkman, T. J., Rice, D. W., and Palmer, J. D. (2017). Comparative mitogenomics indicates respiratory competence in parasitic *Viscum* despite loss of complex I and extreme sequence divergence, and reveals horizontal gene transfer and remarkable variation in genome size. *BMC Plant Biol.* 17 (1), 1–12. doi: 10.1186/s12870-017-0992-8
- Sloan, D. B. (2017). Nuclear and mitochondrial RNA editing systems have opposite effects on protein diversity. *Biol. Lett.* 13 (8), 20170314. doi: 10.1098/rsbl.2017.0314
- Sloan, D. B., Alverson, A. J., Chuckalovcak, J. P., Wu, M., McCauley, D. E., Palmer, J. D., et al. (2012). Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biol.* 10 (1), e1001241. doi: 10.1371/journal.pbio.1001241

- Stephan, G., Pascal, L., and Ralph, B. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucl. Acids Res.* W1, W59–W64. doi: 10.1093/nar/gkz238
- Timmis, J. N., Ayliffe, M. A., Huang, C. Y., and Martin, W. (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* 5 (2), 123–135. doi: 10.1038/nrg1271
- Valero, T. (2014). Mitochondrial biogenesis: pharmacological approaches. *Curr. Pharm. Des.* 20 (35), 5507–5509. doi: 10.2174/138161282035140911142118
- Varani, A. M., Silva, S. R., Lopes, S., Barbosa, J. B. F., Oliveira, D., Corrêa, M. A., et al. (2022). The complete organellar genomes of the entheogenic plant *Psychotria viridis* (Rubiaceae), a main component of the ayahuasca brew. *PeerJ* 10, e14114. doi: 10.7717/peerj.14114
- Varré, J.-S., D'Agostino, N., Touzet, P., Gallina, S., Tamburino, R., Cantarella, C., et al. (2019). Complete sequence, multichromosomal architecture and transcriptome analysis of the solanum tuberosum mitochondrial genome. *Int. J. Mol. Sci.* 20 (19), 4788. doi: 10.3390/ijms20194788
- Vidal, R. O., Mondego, J. M., Pot, D., Ambrósio, A. B., Andrade, A. C., Pereira, L. F., et al. (2010). A high-throughput data mining of single nucleotide polymorphisms in *Coffea* species expressed sequence tags suggests differential homeologous gene expression in the allotetraploid *Coffea arabica*. *Plant Physiol.* 154 (3), 1053–1066. doi: 10.1104/pp.110.162438
- Wang, X., Li, L.-L., Xiao, Y., Chen, X.-Y., Chen, J.-H., and Hu, X.-S. (2021). A complete sequence of mitochondrial genome of *Neolamarckia cadamba* and its use for systematic analysis. *Sci. Rep.* 11 (1), 21452. doi: 10.1038/s41598-021-01040-9
- Wang, C., Wang, Y., Liang, Q., Yang, T. J., Li, Z., and Zhang, Y. (2018). The complete chloroplast genome of *Plagiorhegma dubia* Maxim., a traditional Chinese medicinal herb. *Mitochondrial. DNA Part B* 3 (1), 112–114. doi: 10.1080/23802359.2018.1424586
- Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* 13 (6), e1005595. doi: 10.1371/journal.pcbi.1005595
- Wu, B., Chen, H., Shao, J., Zhang, H., Wu, K., and Liu, C. (2017). Identification of symmetrical RNA editing events in the mitochondria of *Salvia miltiorrhiza* by strand-specific RNA sequencing. *Sci. Rep.* 7 (1), 1–11. doi: 10.1038/srep42250
- Yang, H., Chen, H., Ni, Y., Li, J., Cai, Y., Ma, B., et al. (2022). *De novo* hybrid assembly of the *Salvia miltiorrhiza* mitochondrial genome provides the first evidence of the multi-chromosomal mitochondrial DNA structure of *Salvia* species. *Int. J. Mol. Sci.* 23 (22), 14267. doi: 10.3390/ijms232214267
- Yang, H., Chen, H., Ni, Y., Li, J., Cai, Y., Wang, J., et al. (2023). Mitochondrial Genome Sequence of *Salvia officinalis* (Lamiales: Lamiaceae) Suggests Diverse Genome Structures in Cogeneric Species and Finds the Stop Gain of Genes through RNA Editing Events. *Int. J. Mol. Sci.* 24 (6), 5372. doi: 10.3390/ijms24065372
- Yang, Z., Ni, Y., Lin, Z., Yang, L., Chen, G., Nijjati, N., et al. (2022). *De novo* assembly of the complete mitochondrial genome of sweet potato (*Ipomoea batatas* [L.] Lam) revealed the existence of homologous conformations generated by the repeat-mediated recombination. *BMC Plant Biol.* 22 (1), 285. doi: 10.1186/s12870-022-03665-y
- Yang, Z., Peng, Z., Zhang, H., Joongku, L., Liu, X., and Fu, C. (2018). The complete chloroplast genome of *Gymnospermium kiangnanense* (Berberidaceae): an endangered species endemic to Eastern China. *Mitochondrial. DNA Part B* 3 (2), 713–714. doi: 10.1080/23802359.2018.1483760
- Yates, C. M., and Sternberg, M. J. (2013). The effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on protein–protein interactions. *J. Mol. Biol.* 425 (21), 3949–3963. doi: 10.1016/j.jmb.2013.07.012
- Yu, Q., Guyot, R., de Kochko, A., Byers, A., Navajas-Pérez, R., Langston, B. J., et al. (2011). Micro-collinearity and genome evolution in the vicinity of an ethylene receptor gene of cultivated diploid and allotetraploid coffee species (*Coffea*). *Plant J.* 67 (2), 305–317. doi: 10.1111/j.1365-313X.2011.04590.x
- Zamzami, N., Hirsch, T., Dallaporta, B., Petit, P. X., and Kroemer, G. (1997). Mitochondrial implication in accidental and programmed cell death: apoptosis and necrosis. *J. Bioenerg. Biomembr.* 29 (2), 185–193. doi: 10.1023/A:1022694131572
- Zhang, D., Gao, F., Jakovlić, I., Zou, H., Zhang, J., Li, W. X., et al. (2020). PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.* 20 (1), 348–355. doi: 10.1111/1755-0998.13096
- Zhang, H., Meltzer, P., and Davis, S. (2013). RCircos: an R package for Circos 2D track plots. *BMC Bioinf.* 14 (1), 1–5. doi: 10.1186/1471-2105-14-244