ultivar-resolved assembly and characterization of the mitochondrial genome of Medicago sativa cv. AH reveal repeat-mediated alternative junctions and extensive RNA editing

Yingtong Mu
Inner Mongolia Agricultural University
Jingshi Lu
Inner Mongolia Agricultural University
Shushuan Wang
Inner Mongolia Agricultural University
Kefan Cao
caokefan1003@163.com

Inner Mongolia Agricultural University
Fengling Shi
Inner Mongolia Agricultural University

Research Article

Keywords: Medicago sativa, mitochondrial genome, PacBio HiFi, repeat-mediated recombination, RNA editing, plastid-to-mitochondrion DNA transfer

Posted Date: July 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10150042/v1

Additional Declarations: No competing interests reported.

Abstract

Background: Plant mitochondrial genomes are characterized by extensive structural dynamics driven by repeat-mediated recombination, widespread RNA editing, and plastid-to-mitochondrion DNA transfer (MTPT). However, cultivar-resolved mitochondrial genome assemblies that integrate structural validation, RNA editing landscapes, and phylogenetic context remain limited in *Medicago sativa*, an important legume forage crop.

Results: We assembled the complete mitochondrial genome of *Medicago sativa* cv. AH using PacBio HiFi long reads and a graph-resolving workflow. The circular mitogenome was 300,797 bp in length with a GC content of 45.36%, encoding 33 unique protein-coding genes, 16 tRNA genes, and three rRNA genes. Repeat analysis identified 59 simple sequence repeats, nine tandem repeats, and 112 pairs of dispersed repeats. A 4,354 bp long direct repeat was confirmed by junction PCR and Sanger sequencing to mediate homologous recombination, generating both master-circle and alternative subgenomic configurations. Six MTPT fragments totaling 725 bp were detected, harboring five intact plastid-derived tRNA genes. Maximum-likelihood phylogenetic analysis based on 23 conserved protein-coding genes placed cv. AH within the *M. sativa* clade with strong support (bootstrap = 96), yet synteny analysis revealed substantial rearrangements in gene-block order and orientation among intraspecific mitogenome assemblies. A total of 468 high-confidence C-to-U RNA editing sites were predicted across all 33 protein-coding genes, with nonsynonymous edits accounting for 95.73% of total events. The gene *nad4* harbored the most editing sites (45), and the most frequent amino acid conversions were Ser→Leu (107) and Pro→Leu (104). Representative editing sites were experimentally validated by RT-PCR and Sanger sequencing, including the restoration of a canonical AUG start codon in *nad4L*through ACG-to-AUG editing.

Conclusions: This study provides a cultivar-resolved mitochondrial genomic resource for *M. sativa*cv. AH that integrates structural isoform validation, MTPT characterization, phylogenetic placement, synteny-level rearrangement analysis, and an experimentally supported RNA editing landscape within a single framework. These findings advance our understanding of mitochondrial genome dynamics in alfalfa and provide a foundation for future investigations into cytoplasmic diversity and breeding applications.

1. Introduction

Plant mitochondrial genomes encode key components of oxidative phosphorylation, yet their genome architectures are unusually dynamic, showing extensive rearrangements and structural plasticity across plant lineages [1]. Recombination across repeated sequences is widely recognized as a major driver of plant organelle genome structural variation, generating alternative junctions and configurations [2].

Consistently, physical representations of plant mitochondrial DNA as a single “master circle” can be misleading because multiple isoforms may coexist in vivo [3]. Beyond structural variation, plant mitochondrial transcripts are extensively modified by RNA editing, predominantly C-to-U conversions, which can change coding potential and translation signals [4]. In parallel, plastid-to-mitochondrion DNA

transfer (MTPT) provides an additional layer of organelle genome mosaicism and can persist as plastid-derived fragments in angiosperm mitogenomes [5].

In alfalfa, cultivar-level mitogenome assemblies have revealed substantial structural reorganization and repeat-associated recombination signals, indicating that intraspecific mitochondrial structural variation is a relevant feature within *Medicago sativa* [6]. At the genus scale, comparative mitogenomics across multiple Medicago species further reported that genome conformations and repeat landscapes can vary markedly, and that *M. sativa* may exhibit complex multipartite configurations [7]. Outside Medicago, repeat-mediated recombination has been experimentally linked to alternative mitochondrial genome conformations using long-read evidence and PCR/Sanger validation, underscoring the value of orthogonal validation for structural isoforms [8].

Similarly, RNA editing has been shown to generate biologically meaningful codon changes in plant mitochondria, including restoration of canonical start or stop codons in specific transcripts [9]. MTPT fragments can also contribute to mitochondrial genome content and, in some cases, plastid-derived sequences have been implicated in mitochondrial gene structures, highlighting potential functional consequences beyond passive insertions [10]. Against this background, the extent to which specific long direct repeats in Medicago cultivars generate experimentally verifiable alternative junctions—and how such junctions relate to coexisting genome configurations—remains insufficiently resolved at the cultivar scale.

Although available Medicago assemblies support intraspecific rearrangements, an integrated framework that jointly considers phylogenetic placement, synteny-level rearrangements, and candidate recombinogenic repeats is still needed to interpret cultivar-to-cultivar differences. For RNA editing, broad mechanistic understanding points to nuclear-encoded factors (notably PPR proteins) as key determinants of editing specificity, yet cultivar-resolved editing landscapes and their functional interpretation in the context of dynamic mitogenome structures are seldom treated together [11]. Moreover, because MTPT accumulation and decay histories differ among lineages, plastid-derived fragments—especially intact tRNA genes—may provide informative cytoplasmic signatures, but their spectrum and potential utility in Medicago cultivar mitochondria remain to be clarified [5].

From an applied perspective, mitochondrial genome variation is a recognized source of cytoplasmic traits such as cytoplasmic male sterility (CMS) and therefore has direct relevance to breeding strategies that rely on cytoplasmic diversity and nuclear restoration [12, 13].

Here, we generated a cultivar-resolved mitochondrial genome assembly for *Medicago sativa* cv. AH using PacBio HiFi long reads and a graph-resolving workflow, followed by systematic characterization of repeats, MTPTs, phylogenetic placement, and synteny-level rearrangements within Medicago. To connect structural and transcript-level features in a single experimental framework, we combined junction PCR for repeat-mediated alternative junctions with RT-PCR/Sanger validation of representative RNA editing sites.

Overall, this study provides direct experimental evidence for a long-repeat–mediated alternative junction in an alfalfa cultivar, places cv. AH within the Medicago mitochondrial phylogenetic context while revealing pronounced intraspecific rearrangements, and delivers a cultivar-specific mitochondrial RNA-editing landscape supported by targeted validation.

2. Materials and Methods

2.1 Plant materials and nucleic acid preparation

Fresh leaves of *Medicago sativa* L. cv. AH were collected in Hohhot, Inner Mongolia, China (40.57°N, 111.93°E). Samples are deposited at Inner Mongolia Agricultural University (Hohhot, China). Genomic DNA was extracted from fresh leaves using a Plant DNA Extraction Kit following the manufacturer's protocol. For RNA-based validation, total RNA was extracted from fresh leaf tissue and treated with RNase-free DNase I to remove residual genomic DNA. First-strand cDNA was synthesized using random primers (or oligo(dT) where applicable) following the manufacturer's protocol.

2.2 PacBio HiFi sequencing

PacBio SMRTbell libraries were prepared from high-quality genomic DNA and sequenced to generate circular consensus (HiFi/CCS) reads. The HiFi strategy provides high per-read accuracy through circular consensus of multiple passes [14].

2.3 Mitochondrial genome assembly and graph resolution

HiFi reads were assembled using Oatk, a de novo assembly tool for complex plant organelle genomes [15]. Assembly was initiated with the syncasm module using parameters "-k 1001 -c 30", producing a GFA assembly graph.

To identify mitochondrial contigs, a local BLAST database was built from assembled contigs, and mitochondrial gene sequences from a close reference (e.g., *Medicago sativa NC_068105.1*) were used as queries in BLASTn searches using BLAST+ [16]. Parameters were: -evalue 1e-5 -outfmt 6 -max_hsps 10 -word_size 7 -task blastn-short.

Assembly graphs were visualized and inspected using Bandage [17]. Branches were resolved by selecting read-supported connections through the graph to obtain a circular mitochondrial genome sequence. Graph-aware concepts such as repeat graphs provide a general framework for understanding repeat-induced ambiguity in long-read assemblies [18].

2.4 Genome annotation and visualization

Protein-coding genes and intron/exon boundaries were annotated using PMGA (Plant Mitochondrial Genome Annotator[19]). tRNA genes were annotated using tRNAscan-SE 2.0 [20], and rRNA genes were identified using BLASTn [16]. Annotations were manually inspected and edited using Apollo [21]. Circular genome maps were generated using OGDRAW [22].

2.5 Codon usage analysis

Relative synonymous codon usage (RSCU) was calculated for unique mitochondrial protein-coding genes based on annotated coding sequences. Protein-coding sequences were extracted using PhyloSuite [23], and codon usage metrics were computed in MEGA12 [24]. Codons with RSCU > 1 were treated as preferentially used in summary plots.

2.6 Repeat identification and visualization

Simple sequence repeats (SSRs; 1–6 bp motifs) were identified using MISA-web [25]. Tandem repeats were detected using Tandem Repeats Finder [26]. Dispersed repeats were identified using ROUSFinder (https://github.com/flydoc2000/ROUSfinder; accessed 2026-01-30). Repeat distributions and links were summarized using Circos-style visualizations with Circos [27].

2.7 Repeat-mediated recombination validation by long-range PCR and Sanger sequencing

To validate alternative mitochondrial genome conformations mediated by a pair of long direct repeats, four primers were designed in the unique flanking regions immediately adjacent to each repeat copy based on the assembled mitochondrial genome sequence.

Four primer combinations were used to distinguish the master-circle junctions from the predicted recombined junctions: (i) F1/R1 and (ii) F2/R2 for the master-circle configuration; and (iii) F1/R2 and (iv) F2/R1 for the predicted recombined configurations.

Long-range PCR was performed using a high-fidelity long-amplicon DNA polymerase following the manufacturer's recommendations. Each 25 μL reaction typically contained 1× long-range buffer, 200 μM of each dNTP, 0.2–0.4 μM of each primer, 20–50 ng total genomic DNA, and 0.5–1 U polymerase. Cycling conditions were: 95–98°C for 2 min; 30–32 cycles of 95–98°C for 10 s, 58–62°C for 15–20 s, and 68–72°C for 4.5–5.5 min; followed by a final extension at 68–72°C for 10 min.

PCR products were separated on 0.8–1.0% agarose gels in 1×TAE buffer and visualized by nucleic acid staining. Bands of the expected sizes were excised, gel-purified, and subjected to Sanger sequencing using the corresponding amplification primers to confirm junction sequences spanning the repeat–flank boundaries.

2.8 Identification of plastid-to-mitochondrion DNA transfer (MTPT) fragments

To identify plastid-derived segments in the mitochondrial genome, the assembled mitochondrial sequence was compared to the corresponding chloroplast genome using BLASTn [16]. Alignments with identity > 70%, alignment length > 40 bp, and e-value < 1e-5 were retained as candidate MTPTs.

2.9 Phylogenetic inference

To infer the phylogenetic placement of cv. AH, a set of conserved mitochondrial protein-coding genes shared across sampled taxa was extracted, aligned, and concatenated. Multiple sequence alignments were performed using MAFFT [28]. Maximum-likelihood phylogenetic inference was performed using IQ-TREE 2 [29] under the best-fit model selected in the workflow (reported as GTR + F+R3), with branch support assessed using ultrafast bootstrap (1,000 replicates) and SH-aLRT (1,000 replicates). Trees were visualized and annotated using iTOL v6 [30].

2.10 Synteny analysis

To compare mitogenome organization among Medicago assemblies, homologous regions were detected using BLASTn [16] with parameters: -evalue 1e-5 -word_size 9 -gapopen 5 -gapextend 2 -reward 2 -penalty – 3. Collinear blocks were identified using MCScanX [31]. Only collinear blocks longer than 500 bp were retained for downstream visualization. Synteny plots were generated using TBtools [20].

2.11 RNA editing prediction and validation by RT-PCR and Sanger sequencing

RNA editing sites in mitochondrial protein-coding genes were predicted using Deepred-Mt with a probability cutoff of 0.9 [32].

For selected RNA editing sites, primers were designed to amplify fragments spanning the predicted editing positions based on the mitochondrial genome sequence and gene annotation. PCR was conducted using both genomic DNA (gDNA) and cDNA as templates, and a minus–reverse transcriptase control (RT–) was included to confirm the absence of gDNA contamination in the cDNA preparation. PCR products were checked by agarose gel electrophoresis, purified, and Sanger sequenced. RNA editing was confirmed by comparing chromatograms from gDNA and cDNA at the target nucleotide position, where a C peak in gDNA corresponds to a T peak (representing U in RNA) in cDNA.

3. Results

3.1 Assembly and general features of the *Medicago sativa* cv. AH mitochondrial genome

As shown in Fig. 1A, the long-read-based mitochondrial assembly graph of *Medicago sativa* cv. AH was branched and comprised three connected contigs (ctg1, ctg2 and ctg3). Notably, ctg3 exhibited markedly higher read depth than ctg1 and ctg2, consistent with its presence as a duplicated segment in the final resolved structure. By mapping long reads to the assembly graph and prioritizing the read-supported connections, we simplified the graph into a single dominant path (ctg1–ctg3_copy–ctg2–ctg3) and obtained a circular sequence representing the complete mitochondrial genome (Fig. 1B).

The finalized mitochondrial genome of *M. sativa* cv. AH was 300,797 bp in length with a GC content of 45.36% (Fig. 1C). Genome annotation identified 33 unique protein-coding genes (*PCGs*), including 24

core mitochondrial genes involved in oxidative phosphorylation and cytochrome c biogenesis, and nine non-core genes mainly encoding ribosomal proteins and sdh components (Fig. 1D). In addition, 16 tRNA genes and three rRNA genes were annotated, with *atp1*, *trnF-GAA* and *trnQ-UUG* present in two copies.

Codon usage bias was further evaluated using the mitochondrial PCGs (Fig. 1E). Codons with RSCU values > 1 were considered preferentially used. As expected, methionine (AUG) and tryptophan (UGG), which are each encoded by a single codon, both showed RSCU values of 1.0. Among synonymous codons, alanine preferentially used GCU (RSCU = 1.57), while arginine and histidine showed higher usage preferences for AGA (RSCU = 1.44) and CAU (RSCU = 1.50), respectively. Notably, the stop codon UAA displayed the highest usage preference (RSCU = 1.70) across mitochondrial PCGs.

3.2 Characteristics of repeat sequences in the *Medicago sativa* cv. AH mitogenome

Microsatellite or simple sequence repeats (SSRs) are a special type of tandem repeat with repeat units typically ranging from 1 to 6 nucleotides. In the mitochondrial genome of *Medicago sativa* cv. AH, a total of 59 SSRs were identified (Fig. 2A). Among these SSRs, tetranucleotide repeats represented the largest category, accounting for 37.29% (22/59) of all SSRs, followed by dimeric (14), monomeric (12) and trimeric (11) repeats (Fig. 2A). Notably, no pentameric or hexameric SSRs were detected in this mitogenome (Fig. 2A). Within mononucleotide SSRs, T-monomer repeats were predominant, representing 66.67% (8/12) of monomeric SSRs.

Tandem repeats (also referred to as satellite DNA) consist of adjacent, multiple copies of a repeat unit. Using TRF, we detected nine tandem repeats in the AH mitogenome, with percent matches ≥ 86% and repeat unit/consensus lengths ranging from 10 to 51 bp (Fig. 2B). The detailed coordinates and repeat parameters of each tandem repeat are summarized in Fig. 2B.

In addition, dispersed repeats were further surveyed across the mitochondrial genome. A total of 112 pairs of dispersed repeats (≥ 30 bp) were identified, including 63 palindromic repeats and 49 forward repeats. The longest dispersed repeat unit was R1, with a length of 4,355 bp. The genomic distribution of SSRs, tandem repeats, and dispersed repeats is summarized in a Circos-style plot (Fig. 2C), in which palindromic and forward dispersed repeats are shown as blue and yellow links, respectively.

3.3 Repeat sequence mediated homologous recombination in the mitochondrial genome of *Medicago sativa* cv. AH

A growing body of evidence indicates that repetitive sequences play important roles in homologous recombination (HR) and structural dynamics of plant mitochondrial genomes. In the mitochondrial genome of *Medicago sativa* cv. AH, a long direct repeat (R1/ctg3, 4,354 bp) was identified and supported by the assembly graph and read-depth characteristics (Fig. 3A), suggesting that this repeat could mediate alternative genome configurations via repeat-associated HR.

To further verify whether the 4.354-kb direct repeat could mediate HR, junction PCR amplification was performed using primer pairs designed in the unique flanking regions of the two repeat copies (primer strategy shown in Fig. S1). As shown in Fig. 3B, two primer combinations targeting the master-circle junctions generated single and clear amplicons of ~ 5,043 bp (F1/R1) and ~ 4,910 bp (F2/R2). Importantly, the cross combinations also yielded distinct products of ~ 4,957 bp (F1/R2) and ~ 4,996 bp (F2/R1), which are diagnostic of alternative junctions consistent with repeat-mediated recombination. The sizes of all four amplicons approximate the repeat length (4,354 bp) plus short flanking sequences (~ 0.56–0.69 kb), consistent with primers spanning the repeat–flank junctions. Notably, the amplicons corresponding to the master-circle junctions appeared stronger than those for the recombined junctions, suggesting that the master configuration may represent the predominant isoform while the recombined configurations may be present at lower relative abundance.

Based on the assembly graph and the junction PCR results, we propose a model in which the 4.354-kb direct repeat mediates homologous recombination to generate alternative mitochondrial genome configurations, including a master circle and two repeat-associated subgenomic molecules (Fig. 3C).

3.4 Identification of *MTPTs*

Mitochondrial plastid DNAs (MTPTs) are plastid-derived DNA fragments that have been transferred into the mitochondrial genome. Based on sequence similarity analysis between the *Medicago sativa* 'AH' mitochondrial genome and chloroplast genome using BLASTN, a total of six shared homologous fragments were identified, spanning 725 bp and accounting for 0.24% of the mitogenome (Fig. 4).

Among these *MTPTs*, *MTPT6* was the longest fragment (294 bp). Annotation of the six homologous fragments further revealed five complete genes, all of which were tRNA genes: *trnD-GUC*, *trnH-GUG*, *trnM-CAU*, *trnN-GUU*, and *trnW-CCA*.

3.5 Phylogenetic and synteny analysis

To clarify the evolutionary placement of the *Medicago sativa* cv. AH mitogenome, we constructed a maximum-likelihood phylogenetic tree using concatenated amino acid sequences of 23 conserved mitochondrial protein-coding genes (*PCGs*). In the resulting phylogeny, the mitogenome generated in this study clustered within the *Medicago sativa* lineage and formed a well-supported clade with two previously reported *M. sativa* mitogenomes, *M. sativa* subsp. falcata (*PV916043.1*) and *M. sativa* (*NC_068105.1*) (node support value = 96; Fig. 5A). This topology indicates a close evolutionary relationship among available *M. sativa* mitogenome assemblies while retaining detectable intraspecific variation.

To further evaluate structural conservation and genome rearrangements among Medicago mitogenomes, homologous regions were identified and visualized as syntenic blocks connected by ribbons. As shown in Fig. 5B, extensive collinear blocks were shared among *M. sativa* subsp. falcata (*PV916043.1*), *M. sativa* cv. AH (this study), and *M. sativa* (*NC_068105.1*). However, these blocks were arranged in different orders and orientations, resulting in multiple crossing ribbons, indicating substantial

structural rearrangements among *M. sativa* mitogenomes. Notably, despite the close phylogenetic relationship, the mitochondrial genome organization of cv. AH is not fully conserved relative to previously published *M. sativa* assemblies, consistent with the highly dynamic nature of plant mitochondrial genomes. Compared with the three *M. sativa* mitogenomes, collinearity with the more distant relative *Medicago polymorpha* (*MW971561.1*) appeared more fragmented, in line with increased evolutionary divergence (Fig. 5B).

3.6 Analysis of RNA editing events

RNA editing sites were predicted for mitochondrial protein-coding genes (PCGs) of *Medicago sativa* using a probability cutoff of 0.9. A total of 468 high-confidence RNA editing sites were identified across 33 PCGs, and all predicted events were cytidine-to-uridine (C-to-U) conversions. The gene nad4 harbored the largest number of predicted editing sites (45), followed by ccmB (34) and nad7 (33) (Fig. 6D).

Most editing events were nonsynonymous (95.73%, 448/468), leading to 21 categories of amino-acid outcomes (Fig. 6C). The most frequent conversions were Ser→Leu (107) and Pro→Leu (104), followed by Ser→Phe (62) and Arg→Cys (36). Notably, several sites were predicted to generate canonical start codons through RNA editing (e.g., ACG→AUG), and a small number of edits introduced stop codons, suggesting that RNA editing may potentially contribute to restoring conserved translation signals in mitochondrial transcripts.

To experimentally validate representative high-confidence sites, we amplified three loci containing predicted editing events from both genomic DNA (gDNA) and cDNA. RT-PCR produced amplicons of identical sizes for *nad4L* (282 bp), *atp4* (564 bp) and *ccmFN* (577 bp) (Fig. 6A). Sanger sequencing confirmed the expected C-to-U editing signatures, including *nad4L-2* (ACG→AUG; Thr→Met), *atp4-35* (CCA→CUA; Pro→Leu), and two sites in *ccmFN* (251: UCA→UUA; Ser→Leu; 259: CGG→UGG; Arg→Trp), as indicated by C peaks in gDNA and corresponding T peaks in cDNA chromatograms (Fig. 6B).

4. Discussion

We resolved the *Medicago sativa* cv. AH mitochondrial genome into a single circular sequence by simplifying a branched assembly graph supported by long-read evidence, and we obtained a complete gene annotation and codon-usage summary (Fig. 1). High-accuracy PacBio HiFi/CCS reads are well suited to organelle genome assembly because they reduce error propagation and improve repeat disambiguation compared with error-prone long reads [14]. Even with accurate long reads, plant mitochondrial genomes often assemble into branched graphs because repeats and recombination-related alternative paths are embedded in the underlying DNA population, motivating graph-aware resolution and explicit reporting of alternative structures when supported [1, 3].

We characterized the repeat landscape of cv. AH and identified abundant SSRs, tandem repeats, and dispersed repeats, including a long dispersed repeat candidate (Fig. 2). Across plants, unusually large repeats are common features of mitochondrial genomes and are strongly associated with genome

structural dynamics and size variation, providing a substrate for recombination and rearrangements [33]. SSRs within organelle genomes are frequently considered as potential cytoplasmic markers, but their utility in breeding contexts depends on population-scale validation of polymorphism and inheritance patterns, which remains to be established for cv. AH [34].

A key finding of cv. AH is that a 4.354 kb direct repeat shows diagnostic master and recombined junctions by junction PCR, supporting repeat-mediated homologous recombination and coexistence of alternative configurations (Fig. 3). Repeat-associated homologous recombination is a central mechanism shaping plant organelle genome architecture, producing alternative junctions and structural variants; our junction-level PCR evidence provides direct support for this mechanism in an alfalfa cultivar [2]. Similar validation logic—using long-read assembly evidence coupled with PCR/Sanger confirmation of alternative junctions—has been applied in other plant mitogenomes, strengthening the interpretation that alternative amplicons reflect bona fide isoforms rather than assembly artifacts [8].

The stronger master-junction bands in our PCR gels suggest a predominant configuration, but without quantitative read-mapping or long-range validation this should be treated as an inference rather than a measured stoichiometry [3]. Our configuration model summarizes a master-circle and repeat-associated subgenomic molecules that can interconvert through recombination at the long direct repeat (Fig. 3C). The master-circle concept is useful as a representation but does not imply equimolarity or exclusivity of a single physical molecule, as demonstrated by analyses showing substoichiometric shifting and multiple alternative arrangements in plant mitochondrial genomes [35]. Therefore, interpreting our cv. AH structure as a dominant configuration plus lower-abundance alternatives is consistent with broader evidence, but the physical form and relative abundance of each isoform should be addressed with long-read spanning counts or orthogonal data in future work (hypothesis-driven follow-up) [3].

Despite clustering with published *Medicago sativa* mitogenomes in a conserved-PCG phylogeny, cv. AH displays substantial synteny-level rearrangements relative to other *M. sativa* assemblies (Fig. 5). In alfalfa, He et al. assembled and compared two cultivated varieties (Zhongmu No.1 and Zhongmu No.4) and used Hi-C and long-read data to validate isoforms, highlighting heteroplasmy and repeat-mediated rearrangements at the cultivar level [6]. At a broader Medicago scale, Yang et al. reported chromosomal mitochondrial genome assemblies across multiple Medicago species and performed standardized repeat and comparative analyses, emphasizing the diversity of repeat landscapes and genome conformations within the genus [7].

Relative to these studies, our work adds (i) junction-PCR confirmation of a specific long-repeat recombination event, (ii) a cultivar-resolved synteny comparison aligned to phylogenetic placement, and (iii) an integrated RNA-editing landscape with targeted Sanger validation, thereby linking structural and transcript-level evidence within a single cultivar dataset [6, 7].

We detected six plastid-to-mitochondrion DNA transfer fragments (MTPTs) totaling 725 bp in cv. AH, and these MTPTs include five intact plastid-derived tRNA genes (Fig. 4). MTPT insertions are widespread in angiosperm mitochondrial genomes and are expected to accumulate and decay under weak deletion

pressure and mutational bias, so the small MTPT fraction in cv. AH is consistent with variable insertion/retention histories among lineages [5]. Plant Cell Reports studies of newly assembled mitogenomes have repeatedly noted that MTPT regions often harbor plastid-derived tRNAs, supporting the interpretation that tRNAs can be preferentially retained within transferred plastid fragments [36].

Although plastid-derived segments can sometimes contribute to mitochondrial gene structures in some plants, our data provide sequence-level evidence of transfer and tRNA retention but do not by themselves demonstrate functional contribution [10]. More broadly, MTPTs represent one outcome of intracellular DNA transfer between organelles and the nucleus, a pervasive evolutionary process in plants [37].

Using a conserved-PCG phylogeny and synteny visualization, we show that cv. AH is phylogenetically close to available *M. sativa* mitogenomes but differs markedly in gene-block order and orientation (Fig. 5). Such a pattern—conserved gene content and slow sequence divergence but extensive rearrangement of intergenic architecture—is characteristic of plant mitochondrial genome evolution, where recombination and repair processes can reshuffle genome structure without proportionally accelerating coding-sequence changes [1, 2]. Comparative analyses in multiple taxa similarly report extensive rearrangements mediated by repeats, reinforcing that structural collinearity can diverge rapidly even among close relatives [7, 8].

We predicted 468 high-confidence C-to-U RNA editing sites across 33 mitochondrial protein-coding genes in cv. AH, with a strong predominance of nonsynonymous edits and uneven gene-wise distributions (Fig. 6C–D). Plant mitochondrial RNA editing is predominantly C-to-U and can substantially alter coding sequences; review syntheses emphasize its broad distribution and functional potential across organellar genes [4, 38]. Empirically, editing-site counts in angiosperm mitogenomes are often reported in the hundreds, and published Plant Cell Reports mitogenome studies have documented similar magnitudes and gene-level skews, supporting the plausibility of the cv. AH editing load [36].

Because our editing calls depend on model-based prediction, their completeness and false-positive rates depend on the chosen predictor and thresholds, motivating future RNA-seq-based confirmation for genome-wide site lists [32]. We experimentally validated representative editing sites by RT-PCR and Sanger sequencing, including a case where editing converts a non-canonical ACG to a canonical AUG start codon in nad4L, along with multiple nonsynonymous conversions (Fig. 6A–B). Editing events that restore translation signals or increase amino-acid conservation are widely discussed as key functional consequences of organellar RNA editing, and start/stop restoration is repeatedly reported in plant mitogenome studies [4].

Mechanistically, editing specificity is largely governed by nuclear-encoded factors, particularly PPR proteins, providing a framework for interpreting cultivar-to-cultivar differences as potential outcomes of nuclear background variation [11]. A testable hypothesis emerging from our validated sites is that highly efficient editing at translation-critical positions (e.g., start restoration) will show stronger conservation

across Medicago accessions than edits at less constrained sites, which can be examined by population-scale transcriptome data (hypothesis) [38].

Together, our results highlight that mitochondrial structural isoforms, MTPT content, and RNA editing features coexist within a single alfalfa cultivar and may contribute to cytoplasmic diversity relevant to applied breeding. Mitochondrial genome rearrangements are central to the emergence of cytoplasmic male sterility (CMS) in multiple crops, and therefore cultivar-resolved mitogenome resources provide a necessary foundation for identifying candidate cytoplasmic variants and guiding cytotype-aware breeding strategies (application context; not claiming CMS in cv. AH) [12, 13]. However, linking any specific repeat-mediated configuration or editing profile to agronomic phenotypes requires dedicated phenotyping and genetic tests, including screening for novel chimeric ORFs and validating their expression and effects [12].

This study is limited by the analysis of a single cultivar and by the lack of genome-wide quantitative validation of isoform stoichiometry and RNA editing efficiency. Future work should expand sampling across alfalfa cultivars and related Medicago taxa, and should combine long-read spanning counts, targeted quantitative PCR, and/or long-range data (e.g., Hi-C) to quantify isoform abundance and stability across tissues and environments, as demonstrated in recent alfalfa cultivar studies [6]. Genome-wide confirmation of predicted RNA editing sites by RNA-seq, together with analysis of nuclear-encoded editing factors, will be essential to connect editing variation to nuclear background and potential stress or developmental responsiveness [11].

5. Conclusion

In this study, we assembled and characterized the complete mitochondrial genome of *Medicago sativa* cv. AH using PacBio HiFi long reads and a graph-resolving workflow. The finalized mitogenome was 300,797 bp in length with a GC content of 45.36%, encoding 33 unique protein-coding genes, 16 tRNA genes, and three rRNA genes. Repeat analysis revealed 59 SSRs, nine tandem repeats, and 112 pairs of dispersed repeats, among which a 4,354 bp long direct repeat was identified as a candidate recombinogenic element. Junction PCR and Sanger sequencing confirmed that this repeat mediates homologous recombination, generating both master-circle and alternative subgenomic configurations, providing direct experimental evidence for repeat-mediated structural heteroplasmy in an alfalfa cultivar.

Six plastid-to-mitochondrion DNA transfer fragments totaling 725 bp were detected, harboring five intact plastid-derived tRNA genes, consistent with the widespread but lineage-variable pattern of MTPT accumulation in angiosperm mitochondria. Phylogenetic analysis placed cv. AH within the *M. sativa* clade with strong support, yet synteny comparisons revealed extensive rearrangements in gene-block order and orientation among intraspecific mitogenome assemblies, underscoring the highly dynamic nature of plant mitochondrial genome architecture even among closely related cultivars.

A total of 468 high-confidence C-to-U RNA editing sites were predicted across all 33 protein-coding genes, with nonsynonymous edits predominating. Representative sites were experimentally validated by

RT-PCR and Sanger sequencing, including the restoration of a canonical AUG start codon in *nad4L* through ACG-to-AUG editing, demonstrating the functional relevance of RNA editing in maintaining translational fidelity.

Together, this work provides a cultivar-resolved mitochondrial genomic resource for *Medicago sativa* that integrates structural isoform validation, MTPT characterization, phylogenetic and synteny analyses, and an experimentally supported RNA editing landscape within a single framework. These findings contribute to a deeper understanding of mitochondrial genome dynamics in alfalfa and offer a foundation for future investigations into cytoplasmic diversity, cytoplasmic male sterility, and cytotype-informed breeding strategies.

Abbreviations

BLAST
Basic Local Alignment Search Tool
CCS
Circular Consensus Sequencing
CMS
Cytoplasmic Male Sterility
C-to-U
Cytidine-to-Uridine
gDNA
Genomic DNA
GC
Guanine-Cytosine
GFA
Graphical Fragment Assembly
HiFi
High-Fidelity
HR
Homologous Recombination
MTPT
Mitochondrial Plastid DNA (plastid-to-mitochondrion DNA transfer)
ORF
Open Reading Frame
PCG
Protein-Coding Gene
PCR
Polymerase Chain Reaction
PPR
Pentatricopeptide Repeat

RSCU
Relative Synonymous Codon Usage
rRNA
Ribosomal RNA
RT-PCR
Reverse Transcription Polymerase Chain Reaction
SH-aLRT
Shimodaira–Hasegawa Approximate Likelihood Ratio Test
SMRTbell
Single Molecule Real-Time Bell
SSR
Simple Sequence Repeat
tRNA
Transfer RNA
TRF
Tandem Repeats Finder

Declarations

Acknowledgements

The authors gratefully acknowledge the technical support provided by the College of Grassland Science and the Key Laboratory of Grassland Resources, Ministry of Education, Inner Mongolia Agricultural University.

Funding

This work was supported by the First-Class Disciplines Scientific Research Project of Inner Mongolia Agricultural University (YLXKZX-NND-003). The authors also acknowledge the technical assistance provided by the College of Grassland Science and the Key Laboratory of Grassland Resources, Ministry of Education, Inner Mongolia Agricultural University.

Availability of data and materials

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

This study does not involve any ethical issues, as no human participants or animals were used in the research. Therefore, ethics approval and consent to participate are not applicable.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

The authors give their consent for the publication of this research in its current form.

Authors' contributions

All authors contributed to the study conception and design. Fengling Shi (SFL) conceived and designed the project. Material preparation, data collection, and analysis were performed by Yingtong Mu (MYT), Jingshi Lu (LJS), Shushuan Wang (WSS), and Kefan Cao (CKF). The first draft of the manuscript was written by Yingtong Mu (MYT). Fengling Shi (SFL) supervised the study and revised the manuscript. All authors read and approved the final manuscript.

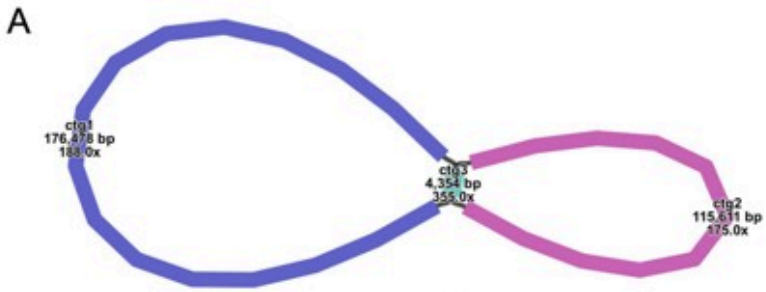
References

1. Gualberto JM, Newton KJ. Plant Mitochondrial Genomes: Dynamics and Mechanisms of Mutation. Annu Rev Plant Biol. 2017;68:225–52. https://doi.org/10.1146/annurev-arplant-043015-112232.
2. Maréchal A, Brisson N. Recombination and the maintenance of plant organelle genome stability. New Phytol. 2010;186:299–317. https://doi.org/10.1111/j.1469-8137.2010.03195.x.
3. 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. https://doi.org/10.1371/journal.pgen.1008373.
4. Knoop V, C-to-U. U-to-C: RNA editing in plant organelles and beyond. J Exp Bot. 2023;74:2273–94. https://doi.org/10.1093/jxb/erac488.
5. 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. https://doi.org/10.1093/gbe/evu253.
6. He X, Zhang X, Deng Y, Yang R, Yu L-X, Jia S, et al. Structural Reorganization in Two Alfalfa Mitochondrial Genome Assemblies and Mitochondrial Evolution in Medicago Species. IJMS. 2023;24:17334. https://doi.org/10.3390/ijms242417334.
7. Yang R, Wang M, Wang M, Li J, Li J, Huang C-H. Assembly and comparative analysis of chromosomal mitochondrial genomes in multiple Medicago species. BMC Plant Biol. 2025;25:1667. https://doi.org/10.1186/s12870-025-07650-z.
8. Hao Z, Zhang Z, Jiang J, Pan L, Zhang J, Cui X, et al. Complete mitochondrial genome of Melia azedarach L., reveals two conformations generated by the repeat sequence mediated recombination. BMC Plant Biol. 2024;24:645. https://doi.org/10.1186/s12870-024-05319-7.
9. Guo N, Wang S, Wang T, Duan M, Zong M, Miao L, et al. A graph-based pan-genome of Brassica oleracea provides new insights into its domestication and morphotype diversification. Plant Commun. 2024;5:100791. https://doi.org/10.1016/j.xplc.2023.100791.

10. Wang D, Rousseau-Gueutin M, Timmis JN. Plastid Sequences Contribute to Some Plant Mitochondrial Genes. Mol Biol Evol. 2012;29:1707–11. https://doi.org/10.1093/molbev/mss016.
11. Barkan A, Small I. Pentatricopeptide Repeat Proteins in Plants. Annu Rev Plant Biol. 2014;65:415–42. https://doi.org/10.1146/annurev-arplant-050213-040159.
12. Chen Z, Zhao N, Li S, Grover CE, Nie H, Wendel JF, et al. Plant Mitochondrial Genome Evolution and Cytoplasmic Male Sterility. CRC Crit Rev Plant Sci. 2017;36:55–69. https://doi.org/10.1080/07352689.2017.1327762.
13. Schnable P. The molecular basis of cytoplasmic male sterility and fertility restoration. Trends Plant Sci. 1998;3:175–80. https://doi.org/10.1016/S1360-1385(98)01235-7.
14. Wenger AM, Peluso P, Rowell WJ, Chang P-C, Hall RJ, Concepcion GT, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat Biotechnol. 2019;37:1155–62. https://doi.org/10.1038/s41587-019-0217-9.
15. Zhou C, Brown M, Blaxter M, Darwin Tree of Life Project Consortium, McCarthy SA, Durbin R. Oatk: a de novo assembly tool for complex plant organelle genomes. Genome Biol. 2025;26:235. https://doi.org/10.1186/s13059-025-03676-6.
16. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421. https://doi.org/10.1186/1471-2105-10-421.
17. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of *de novo* genome assemblies. Bioinformatics. 2015;31:3350–2. https://doi.org/10.1093/bioinformatics/btv383.
18. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37:540–6. https://doi.org/10.1038/s41587-019-0072-8.
19. Li J, Ni Y, Lu Q, Chen H, Liu C. PMGA: A plant mitochondrial genome annotator. Plant Comm. 2025;6. https://doi.org/10.1016/j.xplc.2024.101191.
20. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol Plant. 2020;13:1194–202. https://doi.org/10.1016/j.molp.2020.06.009.
21. Lewis S, Searle S, Harris N, Gibson M, Iyer V, Richter J, et al. Apollo: a sequence annotation editor. Genome Biol. 2002;3. https://doi.org/10.1186/gb-2002-3-12-research0082. research0082.1.
22. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47:W59–64. https://doi.org/10.1093/nar/gkz238.
23. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, et al. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 2020;20:348–55. https://doi.org/10.1111/1755-0998.13096.
24. Kumar S, Stecher G, Suleski M, Sanderford M, Sharma S, Tamura K. Mol Biol Evol. 2024;41:msae263. https://doi.org/10.1093/molbev/msae263. MEGA12: Molecular Evolutionary Genetic Analysis Version 12 for Adaptive and Green Computing.

25. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33:2583–5. https://doi.org/10.1093/bioinformatics/btx198.
26. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27:573–80. https://doi.org/10.1093/nar/27.2.573.
27. Krzywinski M, Schein J, Birol İ, Connors J, Gascoyne R, Horsman D, et al. Circos: An information aesthetic for comparative genomics. Genome Res. 2009;19:1639–45. https://doi.org/10.1101/gr.092759.109.
28. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013;30:772–80. https://doi.org/10.1093/molbev/mst010.
29. 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. https://doi.org/10.1093/molbev/msaa015.
30. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024;52:W78–82. https://doi.org/10.1093/nar/gkae268.
31. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40:e49–49. https://doi.org/10.1093/nar/gkr1293.
32. Edera AA, Small I, Milone DH, Sanchez-Puerta MV. Deepred-Mt: Deep representation learning for predicting C-to-U RNA editing in plant mitochondria. Comput Biol Med. 2021;136:104682. https://doi.org/10.1016/j.compbiomed.2021.104682.
33. Wynn EL, Christensen AC. Repeats of Unusual Size in Plant Mitochondrial Genomes: Identification, Incidence and Evolution. G3 Genes|Genomes|Genetics. 2019;9:549–59. https://doi.org/10.1534/g3.118.200948
34. Guo H, Liu Q, Chen Y, Niu H, Zhao Q, Song H, et al. Comprehensive assembly and comparative examination of the full mitochondrial genome in *Castanea mollissima* Blume. Genomics. 2023;115:110740. https://doi.org/10.1016/j.ygeno.2023.110740.
35. Mower JP, Case AL, Floro ER, Willis JH. Evidence against Equimolarity of Large Repeat Arrangements and a Predominant Master Circle Structure of the Mitochondrial Genome from a Monkeyflower (Mimulus guttatus) Lineage with Cryptic CMS. Genome Biol Evol. 2012;4:670–86. https://doi.org/10.1093/gbe/evs042.
36. Guo S, Li Z, Li C, Liu Y, Liang X, Qin Y. Assembly and characterization of the complete mitochondrial genome of Ventilago leiocarpa. Plant Cell Rep. 2024;43:77. https://doi.org/10.1007/s00299-023-03126-2.
37. Timmis JN, Ayliffe MA, Huang CY, Martin W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet. 2004;5:123–35. https://doi.org/10.1038/nrg1271.
38. Chateigner-Boutin A-L, Small I, Plant RNA. editing. RNA Biology. 2010;7:213–9. https://doi.org/10.4161/rna.7.2.11343

Figures



B

Contig	Length (bp)	Depth (×)
ctg1	176,478	188
ctg2	115,611	175
ctg3	4,354	355

D

Group of genes	Name of genes
ATP synthase	*atp*1（×2）, *atp*4, *atp*6, *atp*8, *atp*9
NADH dehydrogenase	*nad*1, *nad*2, *nad*3, *nad*4, *nad*4L, *nad*5, *nad*6, *nad*7, *nad*9
Cytochrome *b*	*cob*
Cytochrome *c* biogenesis	*ccm*B, *ccm*C, *ccm*FC, *ccm*FN
Cytochrome *c* oxidase	*cox*1, *cox*2, *cox*3
Maturases	*mat*R
Protein transport subunit	*mtt*B
Ribosomal protein large subunit	*rpl*5, *rpl*16
Ribosomal protein small subunit	*rps*3, *rps*4, *rps*7, *rps*10, *rps*12, *rps*14
Succinate dehydrogenase	*sdh*4
Ribosome RNA	*rrn*5, *rrn*18, *rrn*26
Transfer RNA	*trn*C-GCA, *trn*D-GUC, *trn*E-UUC, *trn*F-GAA（×2）, *trnf*M-CAU, *trn*G-GCC, *trn*H-GUG, *trn*I-CAU, *trn*K-UUU, *trn*M-CAU, *trn*N-GUU, *trn*P-UGG, *trn*Q-UUG（×2）, *trn*S-GGA, *trn*W-CCA, *trn*Y-GUA

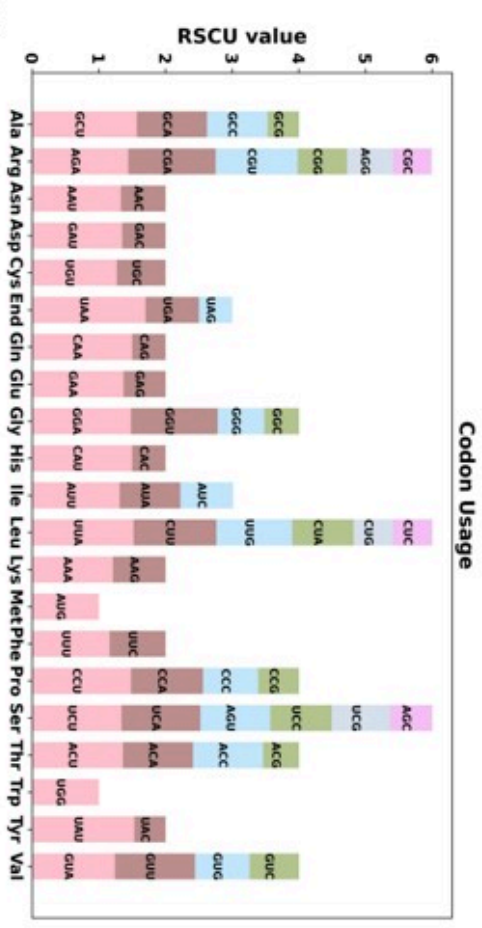
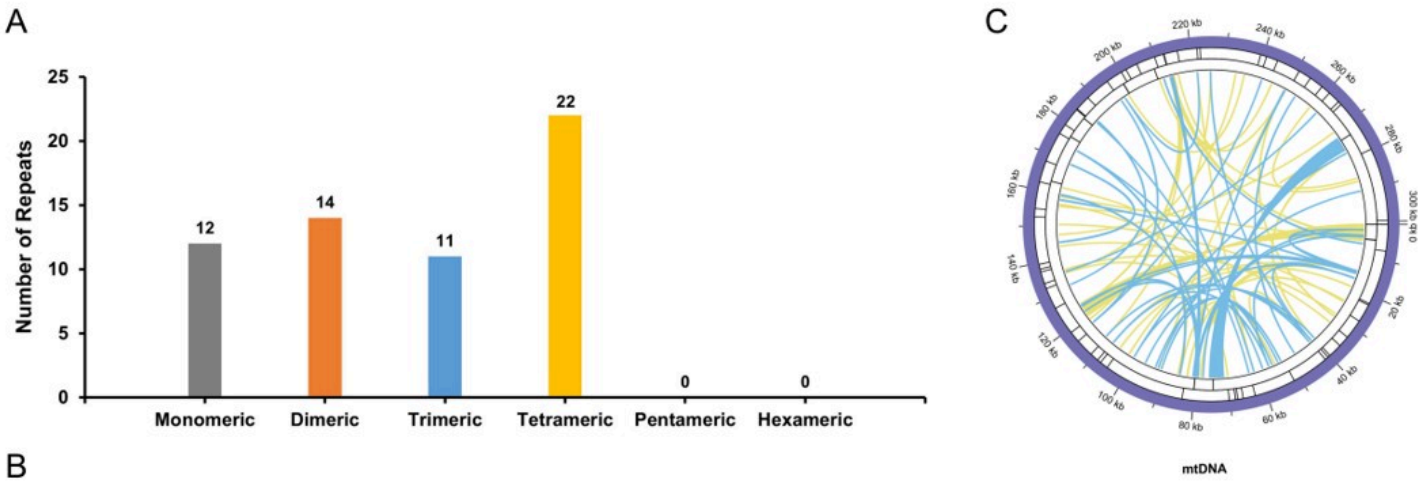


Figure 1

Assembly, annotation and codon usage of the *Medicago sativa* cv. AH mitochondrial genome. A Bandage visualization of the long-read assembly graph showing three connected contigs (ctg1, ctg2 and ctg3). Numbers on nodes indicate contig length (bp) and sequencing depth (×). B Schematic representation of the resolved circular mitochondrial genome after read-supported graph simplification (ctg1–ctg3_copy–ctg2–ctg3).

C Circular gene map of the *M. sativa* cv. AH mitochondrial genome generated using OGDRAW. Genes on the outer and inner circles are transcribed clockwise and counterclockwise, respectively; functional categories are color-coded. D Summary of mitochondrial gene content and functional categories,

including 33 unique PCGs (24 core and nine non-core genes), 16 tRNA genes and three rRNA genes. Multi-copy genes (atp1, trnF-GAA and trnQ-UUG) are indicated. E Relative synonymous codon usage (RSCU) of mitochondrial PCGs. Codon families are shown on the x-axis and corresponding codons are labeled within bars; RSCU > 1 indicates preferential usage. Gene names are italicized.



B

Mitochondrial genome	Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score	A	C	G	T	Entropy (0-2)
mtDNA	4458--4561	51	2	51	98	0	199	21	28	35	14	1.92
mtDNA	74573--74597	13	1.9	13	100	0	50	64	8	8	20	1.46
mtDNA	80211--80274	27	2.4	27	97	2	121	17	31	3	48	1.62
mtDNA	163073--163108	19	1.9	19	88	5	56	13	27	27	30	1.94
mtDNA	177167--177204	10	3.8	10	86	10	51	65	2	2	28	1.19
mtDNA	183714--183753	19	2.1	19	100	0	80	25	15	10	50	1.74
mtDNA	198894--198965	30	2.4	30	97	0	135	19	38	36	5	1.75
mtDNA	208833--208858	12	2.2	12	100	0	52	34	7	7	50	1.6
mtDNA	273362--273386	13	1.9	13	100	0	50	20	8	8	64	1.46

Figure 2

Repeat sequence landscape of the *Medicago sativa* cv. AH mitochondrial genome. A Type and number of simple sequence repeats (SSRs) identified in the AH mitogenome. Bars indicate the counts of SSRs classified by repeat unit length (monomeric to hexameric). B Tandem repeats detected in the AH mitogenome. For each tandem repeat, the table reports its genomic coordinates (Indices), repeat period/consensus size, estimated copy number, similarity statistics (Percent Matches and Percent Indels), and additional sequence composition metrics as output by TRF. C Integrated Circos plot of repeats in the AH mitogenome. The inner links connect dispersed repeat pairs; blue links represent palindromic repeats and yellow links represent forward repeats. The black blocks in the inner track indicate tandem repeat positions, and the black blocks in the outer track indicate SSR positions. The outermost ring shows the coordinate scale along the mitochondrial genome.

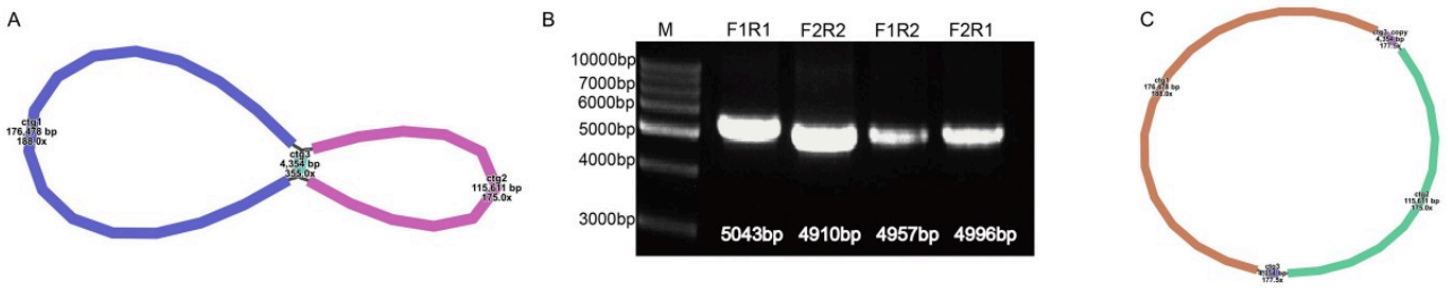


Figure 3

A 4.354-kb direct repeat mediates homologous recombination and generates alternative mitochondrial genome configurations in *Medicago sativa* cv. AH. A Schematic representation of the long direct repeat (R1/ctg3, 4,354 bp) and its two genomic copies linking the major unique regions (ctg1 and ctg2). Homologous recombination between the two repeat copies is expected to generate alternative junctions and configurations. B Junction PCR verification of repeat-mediated alternative junctions. Lane M: 1 kb DNA ladder (the ~5 kb region is used as the reference). Lane 1: F1/R1, master-circle junction across repeat copy 1, 5,043 bp. Lane 2: F2/R2, master-circle junction across repeat copy 2,4,910 bp. Lane 3: F1/R2, diagnostic product for recombined junction, 4,957 bp. Lane 4: F2/R1, diagnostic product for recombined junction, 4,996 bp. The amplicon sizes are close to the repeat length (4,354 bp) plus short flanking sequences, consistent with primer placement spanning repeat–flank junctions. C Proposed alternative mitochondrial genome configurations generated by homologous recombination between the two copies of the 4.354-kb direct repeat. The model summarizes the coexistence of master-circle junctions and recombined junctions supported by junction PCR (panel B).

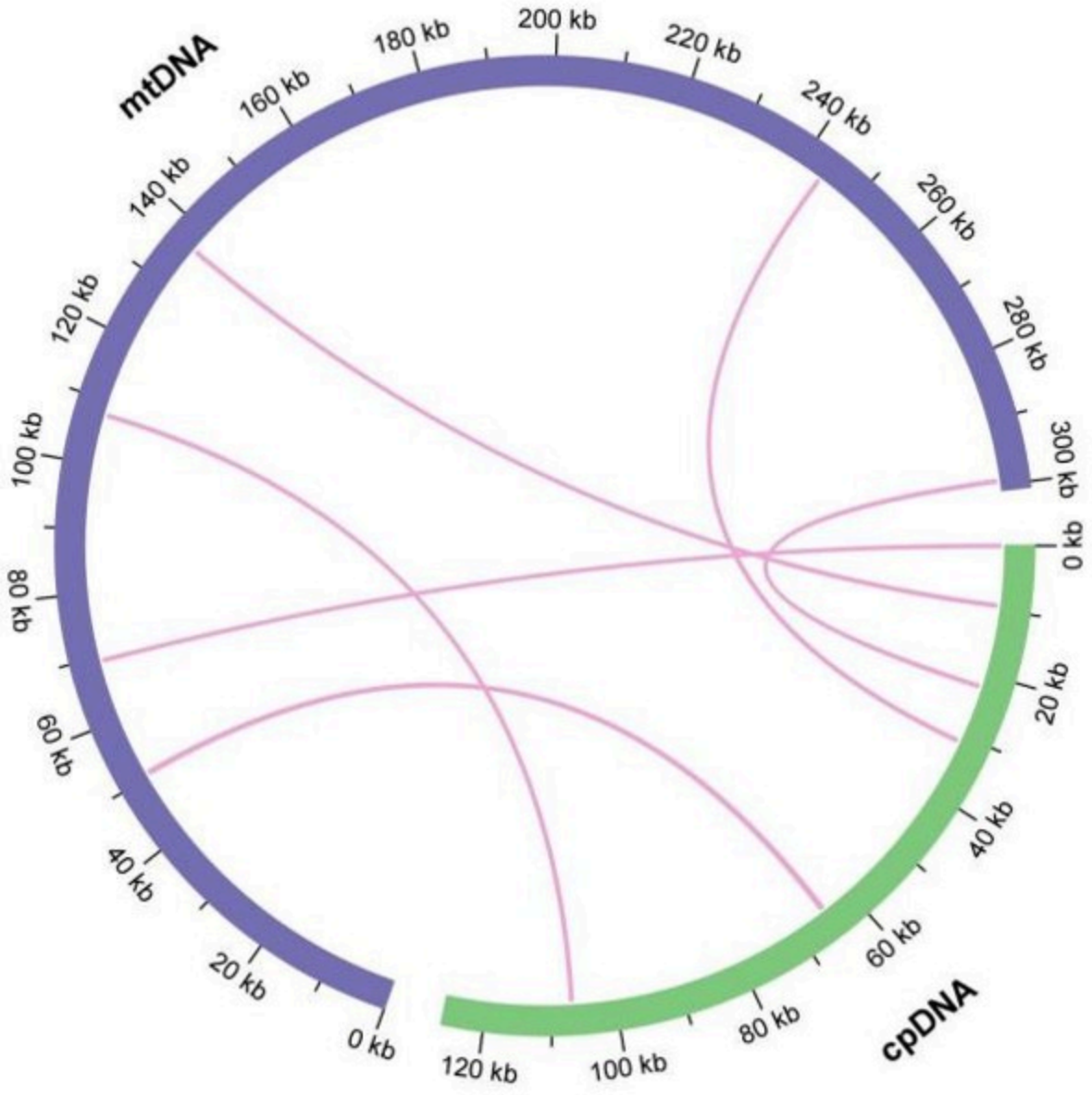


Figure 4

Identification of mitochondrial plastid DNA sequences (MTPTs) in *Medicago sativa* 'AH'. The purple arc represents the mitochondrial genome, and the green arc represents the chloroplast genome. Pink connecting ribbons indicate homologous fragments (*MTPTs*) detected between the two organelle genomes based on BLASTN similarity. A total of six homologous fragments were identified with a combined length of 725 bp (0.24% of the mitogenome), and the longest fragment was *MTPT6* (294 bp). Annotation of the *MTPT* regions identified five complete *tRNA*genes: *trnD-GUC*, *trnH-GUG*, *trnM-CAU*, *trnN-GUU*, and *trnW-CCA*.

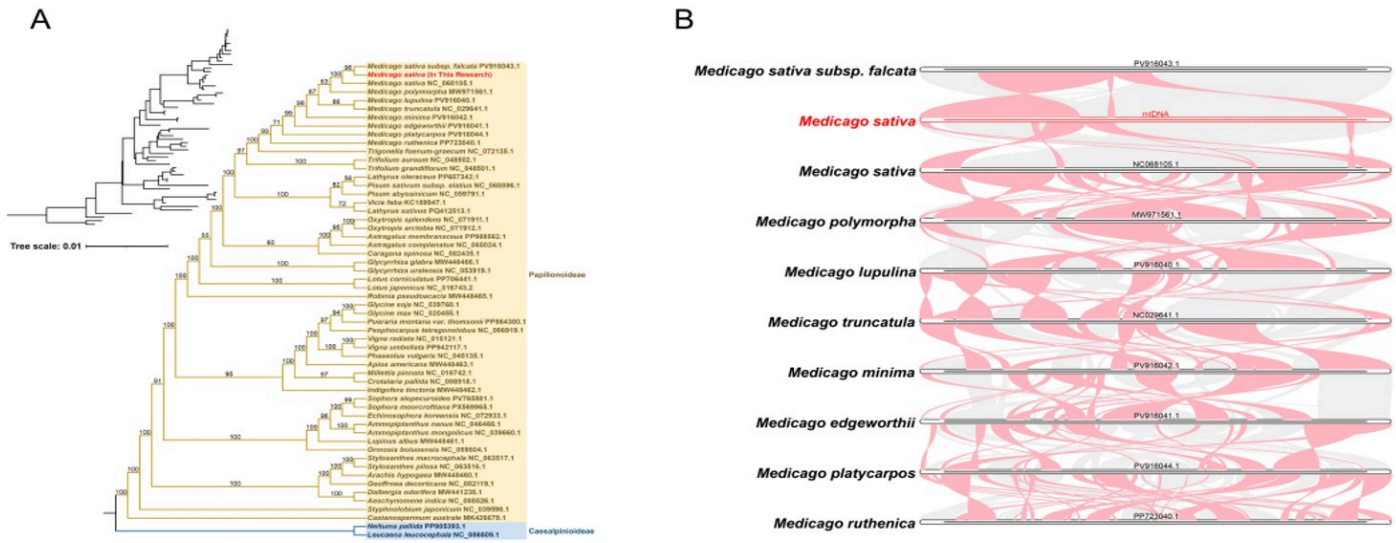


Figure 5

Phylogenetic placement and synteny comparison of the *Medicago sativa* cv. AH mitogenome. A Maximum-likelihood phylogenetic tree inferred from concatenated amino acid sequences of 23 conserved mitochondrial PCGs: *atp1*, *atp4*, *atp6*, *atp8*, *ccmB*, *ccmC*, *ccmFC*, *ccmFN*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad4*, *nad5*, *nad6*, *nad7*, *nad9*, *rpl5*, *rpl16*, *rps3*, *rps14*, and *sdh4* (see Table Sx for taxa and gene list). The mitogenome obtained in this study is highlighted in red. Numbers at nodes indicate branch support values. B Synteny comparison among Medicago mitogenomes visualized as collinear blocks connected by ribbons. Horizontal bars represent mitogenome assemblies of *M. sativa* subsp. falcata (*PV916043.1*), *M. sativa* cv. AH (this study), *M. sativa* (*NC_068105.1*), and *Medicago polymorpha*(*MW971561.1*). Ribbons indicate homologous regions identified by sequence similarity; crossing ribbons and changes in block order/orientation indicate rearrangements among mitogenomes.

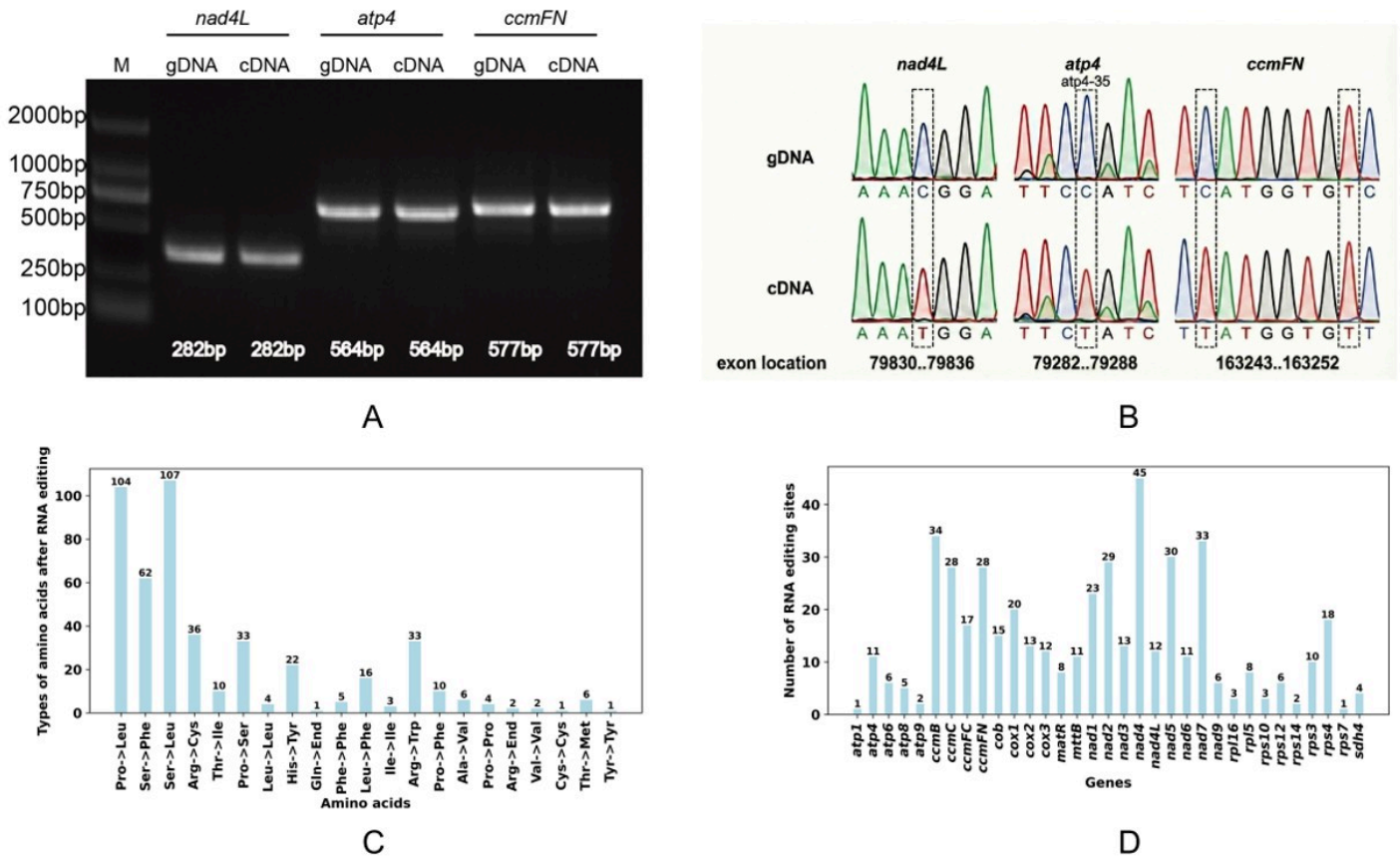


Figure 6

Landscape and experimental validation of RNA editing events in the mitochondrial PCGs of *Medicago sativa*. A RT-PCR amplification of three loci containing representative predicted RNA editing sites in *nad4L*, *atp4*and *ccmFN* using gDNA and cDNA templates. Lane M, DNA size marker; expected amplicon sizes are 282 bp (*nad4L*), 564 bp (*atp4*), and 577 bp (*ccmFN*). B Sanger sequencing chromatograms comparing gDNA and cDNA amplicons. Dashed boxes indicate edited cytidines detected as C peaks in gDNA and corresponding T peaks in cDNA, consistent with C-to-U RNA editing. Validated sites include *nad4L-2*, *atp4-35*, and two sites in *ccmFN*(base positions 251 and 259). Genomic coordinates are indicated below each panel (*nad4L*: 79833; *atp4*: 79285; *ccmFN*: 163244 and 163252; 1-based coordinates). C Distribution of amino-acid outcomes caused by predicted RNA editing sites (n = 468). Numbers above bars indicate counts of each amino-acid change type. D Number of predicted RNA editing sites in each mitochondrial PCG. All predicted sites and their annotations (gene, base position, codon change, amino-acid change, and probability) are provided in.