BMC Genomics

RESEARCH

Open Access

The complete mitochondrial genome assembly of *Capsicum pubescens* reveals key evolutionary characteristics of mitochondrial genes of two *Capsicum* subspecies

Lin Li[1,2], Huizhen Fu[1,2], Muhammad Ahsan Altaf[1,2], Zhiwei Wang[1,2] and Xu Lu[1,2*]

Abstract

Background Pepper (*Capsicum pubescens*), one of five domesticated pepper species, has unique characteristics, such as numerous hairs on the epidermis of its leaves and stems, black seeds, and vibrant purple flowers. To date, no studies have reported on the complete assembly of the mitochondrial genome (mitogenome) of *C. pubescens*. Understanding the mitogenome is crucial for further research on *C. pubescens*.

Results In our study, we successfully assembled the first mitogenome of *C. pubescens*, which was assigned the GenBank accession number OP957066. This mitogenome has a length of 454,165 bp and exhibits the typical circular structure observed in most mitogenomes. We annotated a total of 70 genes, including 35 protein-coding genes (PCGs), 30 tRNA genes, 3 rRNA genes, and 2 pseudogenes. Compared to the other three pepper mitogenomes (KJ865409, KJ865410, and MN196478), *C. pubescens* OP957066 exhibited four unique PCGs (*atp4*, *atp8*, *mttB*, and *rps1*), while two PCGs (*rpl10* and *rps3*) were absent. Notably, each of the three pepper mitogenomes from *C. annuum* (KJ865409, KJ865410, and MN196478) experienced the loss of four PCGs (*atp4*, *atp8*, *mttB*, and *rps1*). To further explore the evolutionary relationships, we reconstructed a phylogenetic tree using the mitogenomes of *C. pubescens* and fourteen other species. Structural comparison and synteny analysis of the above four pepper mitogenomes revealed that *C. pubescens* shares high sequence similarity with KJ865409 and that *C. pubescens* has rearranged with the other three pepper mitogenomes. Interestingly, we observed 72 similar sequences between the mitochondrial and chloroplast genomes, which accounted for 12.60% of the mitogenome, with a total length of 57,207 bp. These sequences encompassed 12 tRNA genes and the rRNA gene (*rrn18*). Remarkably, selective pressure analysis suggested that the *nad5* gene underwent obvious positive selection. Furthermore, a single-base mutation in three genes (*nad1*, *nad2*, and *nad4*) resulted in an amino acid change.

Conclusion This study provides a high-quality mitogenome of pepper, providing valuable molecular data for future investigations into the exchange of genetic information between pepper organelle genomes.

Keywords Pepper, *Capsicum pubescens*, Mitochondrial genome, Organelle genome, Phylogenetic, Homologous sequences, Positive selection

*Correspondence:
Xu Lu
luxu@hainanu.edu.cn
Full list of author information is available at the end of the article

Background

The *Capsicum* genus includes five domesticated cultivars: *C. pubescens*, *Capsicum annuum*, *Capsicum baccatum*, *Capsicum chinense*, and *Capsicum frutescens* [1]. Among them, *C. annuum*, *C. chinense*, and *C. frutescens* have been cultivated in many countries and regions worldwide [2, 3]. *C. pubescens* has abundant hairs on the epidermis of its leaves and stems and produces purple flowers. Additionally, it is distinctive from other pepper cultivars in the presence of black seeds [4]. *C. pubescens* has considerable economic value. It is widely planted in the Andes [3] and is often used as a condiment or a fruit for consumption [5]. Moreover, *C. pubescens* provides support and shade in orchards and has medicinal properties, including anti-inflammatory, antibacterial, and digestive benefits [6]. Interestingly, of the five domesticated pepper cultivars, *C. pubescens* is a hardy cultivar [7].

Mitochondria play a crucial role in converting the proton concentration gradient in living cells into ATP, serving as the "energy factories" of cells. However, it is essential to note that mitochondria have diverse functions beyond their primary role as the powerhouse of the cell. They are involved in maintaining calcium homeostasis, facilitating the biosynthesis of heme and ubiquinone, assembling iron-sulfur clusters, regulating fatty acid metabolism, and performing various other vital functions. In terms of inheritance, while most plants inherit nuclear genetic information from both parents, chloroplast and mitochondrial DNA are typically inherited maternally [8]. Mitochondrial DNA differs from chloroplast DNA and nuclear DNA, with a generally lower substitution rate in plant mitochondria than in chloroplast and nuclear DNA [9]. Mitochondrial DNA is commonly used to study evolutionary relationships, hybridization events, and nuclear interactions among species [10–12].

With the rapid development of second-generation sequencing and genome assembly technology, completing genome sequencing has become efficient and affordable [13]. As a result, there has been a substantial increase in the publication of information regarding organelle genomes [14, 15]. As of July 2023, the GenBank database contains 10,388 published complete organelle genomes of land plants, including 464 mitogenomes and 9,924 chloroplast (cp) genomes. The number of published cp genomes surpasses that of mitogenomes because assembling mitogenomes is more complex than assembling cp genomes [8, 16]. Obtaining and assembling complete mitogenome sequences from plants pose challenges due to the presence of large repetitive sequences [17–19]. Additionally, mitogenomes often exhibit extensive recombination and rearrangement, predominantly caused by the abundance of repetitive sequences [20]. Some repeats maintain a certain degree of specificity between species, so these specific repeats play an essential role as detailed genetic markers in studying the evolutionary connection between species [21].

Typically, the mitogenomes of plants are circular, except for *Oryza sativa*, which has linear mitogenomes [22]. The size of plant mitogenomes ranges from 66 kb in *Viscum scurruloideum* [23] to 11.3 Mb in *Silene conica* [24], with most ranging from 200 to 800 kb. The size variation is primarily due to the absorption of foreign DNA from other organisms by the mitogenome and the complex nature of repetitive sequences, leading to extensive changes in size [25, 26]. In other words, in addition to variations in the size of repeat regions, it is essential to consider another significant contributor to mitogenome diversity—plastid DNA (cpDNA)—which enters mitochondria through the process of endosymbiotic gene transfer (EGT). In addition to their size, plant mitogenomes also exhibit significant variation in gene length and content [8, 27].

Although various plant mitogenomes have been published, there have been no relevant reports on the sequencing of the *C. pubescens* mitogenome. In our study, we sequenced and assembled the *C. pubescens* mitogenome and obtained a circular structure with a size of 454,165 bp. We comprehensively describe the *C. pubescens* mitogenome, including its genomic features, codon usage bias, repetitive sequences, and RNA editing sites. Additionally, we combined other sequenced and published mitogenomes of higher plants to comprehensively analyze the structure and evolution of the mitogenome of *C. pubescens*.

Results

Genomic characterization of the mitogenome of *C. pubescens*

The *C. pubescens* mitogenome is 454,165 bp in length and has a typical circular structure (Fig. 1). The genome nucleotide composition was 28.04% A, 27.68% T, 22.29% G, and 22.00% C. Moreover, the GC content was 44.29% (Table 1). We annotated a total of 70 genes in the *C. pubescens* mitogenome, which included 35 PCGs, 30 tRNA genes, 3 rRNA genes, and 2 pseudogenes (Table S1). The coding genes had a combined length of 37,303 bp, accounting for 8.21% of the entire mitogenome. Among them, PCGs made up 6.50% of the mitogenome, while rRNA and tRNA genes accounted for 1.22% and 0.50%, respectively (Table 1). The mitogenome of *C. pubescens* encodes 35 proteins, which can be categorized into different groups (Table S2): ribosomal small subunit (SSU), NADH dehydrogenase, ATP synthase, cytochrome *c* biogenesis, cytochrome *c* oxidase, ribosomal large subunit (LSU), transport membrane protein, maturases, succinate dehydrogenase, ubiquinol cytochrome *c* reductase,

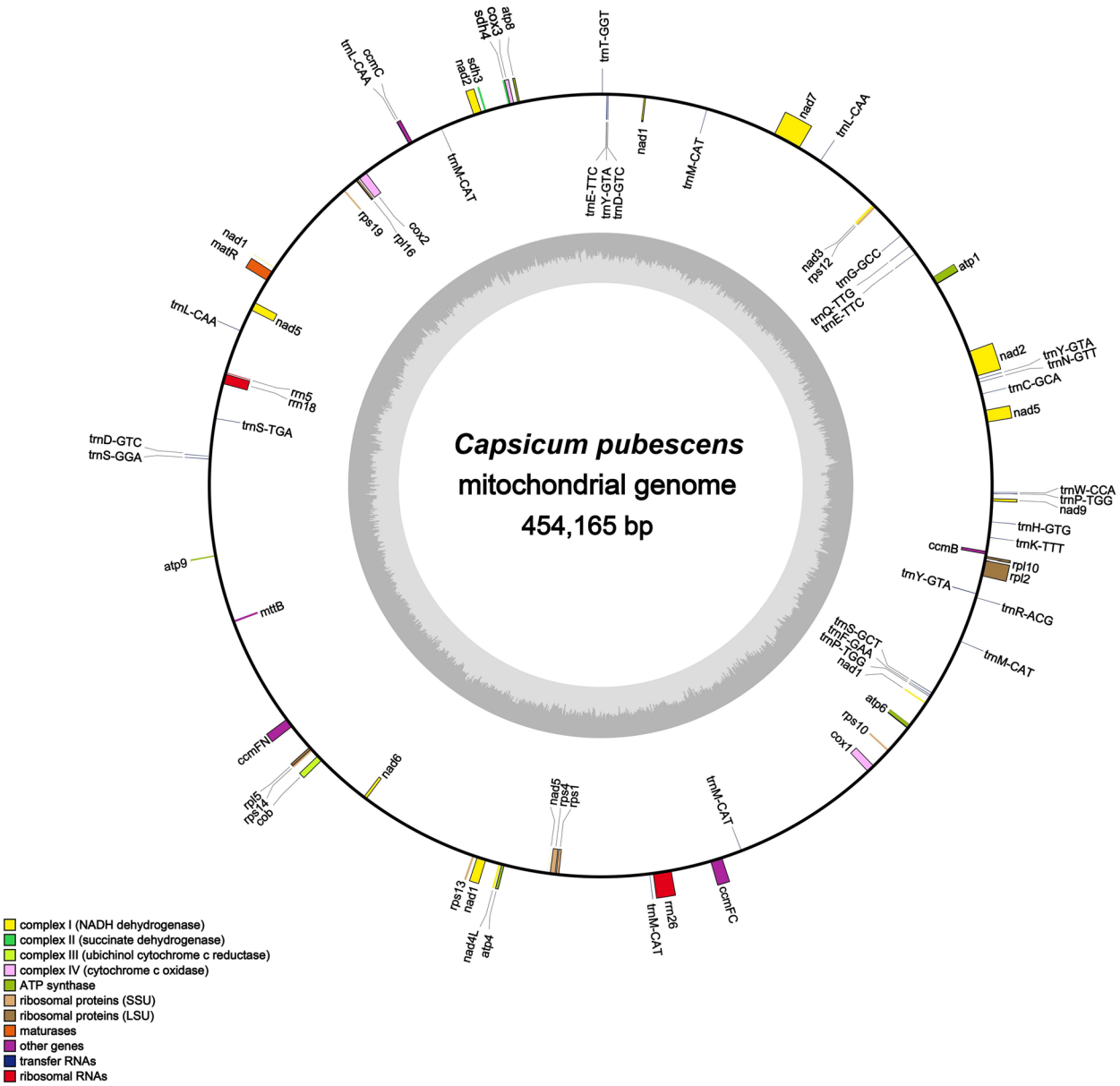

Fig. 1 Circular map of the C. pubescens mitogenome. The forward coding genes are placed outside the circle, while the reverse coding genes are nestled on the inside, and the inner gray circle represents the GC content. A circular map of the C. pubescens mitogenome was created using the OGDRAW online tool (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html)

Table 1 Genomic features of the C. pubescens mitogenome

Feature	Size (bp)	A (%)	T (%)	G (%)	C (%)	GC (%)	Proportion in Genome (%)
Mitogenome	454,165	28.04	27.68	22.29	22	44.29	100
PCGs	29,514	26.24	31.02	21.63	21.11	42.74	6.5
tRNAs	2,262	22.55	26.22	28.43	22.81	51.24	0.5
rRNAs	5,527	26.13	22.24	28.86	22.78	51.64	1.22

NADH dehydrogenase subunits, ATP synthase subunits, cytochrome *c* oxidase subunits, ubiquinol cytochrome *c* reductase subunits, and succinate dehydrogenase subunits. There are 10 intronic genes (*ccmFC*, *cox2*, *nad1*, *nad2*, *nad4*, *nad5*, *nad7*, *rpl2*, *rps1*, and *trnY-GTA*) in the mitogenome of *C. pubescens*, with a total of 24 introns. The *nad1*, *nad2*, *nad5*, and *nad7* genes each contain 4 introns, while the *nad4* gene contains 3 introns. Additionally, the *ccmFC*, *cox2*, *rpl2*, *rps1*, and *trnY-GTA* genes each had only one intron (Table S2).

Protein coding genes (PCGs) and codon usage analysis

The length of the 35 PCGs in the *C. pubescens* mitogenome was 29,514 bp, accounting for 79.12% of the total length of all coding genes (37,303 bp). All PCGs start with the ATG codon as the start codon. The following stop codons were used with the following frequencies: TAA (51.43%), TGA (34.29%), and TAG (14.29%) (Table S3). Analysis of the frequency of amino acid usage in the PCGs revealed that Leu was the most commonly used amino acid, followed by Ser, Ile, Gly, Arg, Phe, Ala, and Val (Table S4). We investigated the relative synonymous codon usage (RSCU) of the PCGs in the *C. pubescens* mitogenome and identified 30 optimal codons (RSCU > 1): GCU, UAU, CCU, UAA, CAA, GGA, CAU, ACU, UUA, AGA, GAU, UCU, AAU, GAA, AUU, CGA, GGU, CUU, GUU, CGU, AAA, UUG, GUA, UGU, UCA, UUU, CCA, AGU, ACA, and UGA. The most commonly used codons are UUU (Phe), AUU (Ile), and UUC (Phe). From Fig. 2, we could observe that the two amino acids, Met (ATG) and Trp (TGG), presented no preference due to having only one codon, while every other amino acid had its preferred codon (Table S4).

Analysis of repeat sequences in the *C. pubescens* mitogenome

We discovered 473 interspersed repeats (> 30 kb) in the *C. pubescens* mitogenome. Among these repeats, there were 246 forward repeats and 227 palindromic repeats but no complement or reverse repeats (Fig. 3a). The

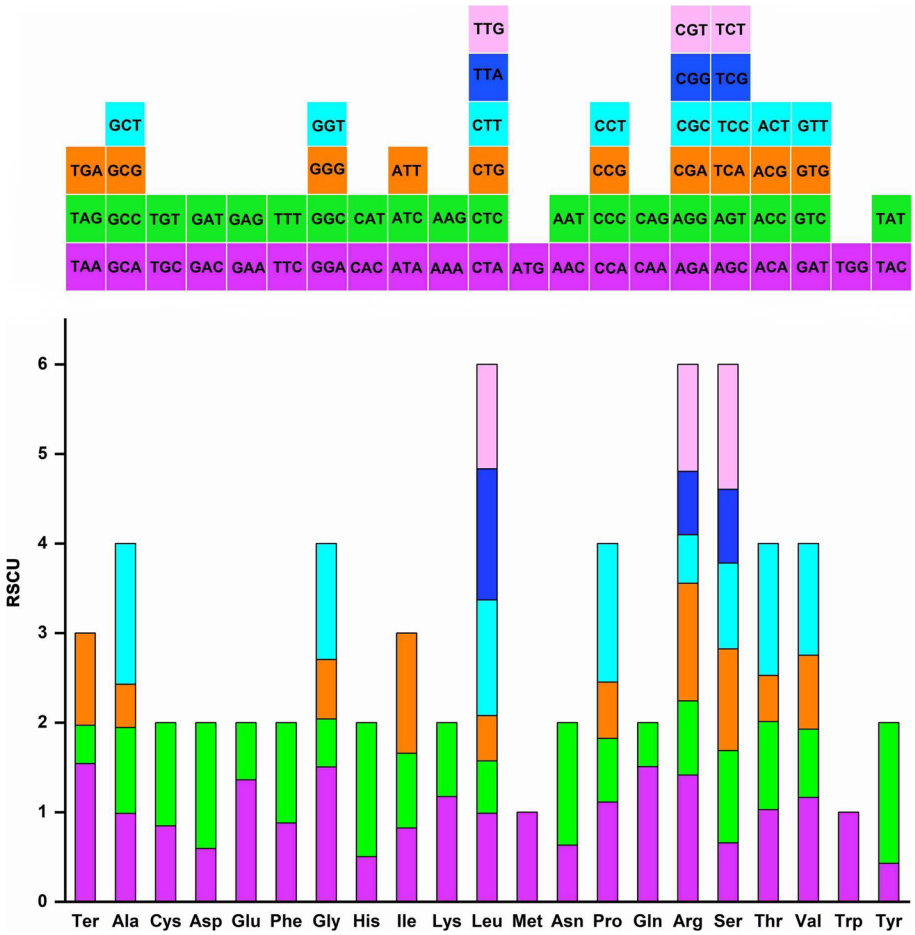

Fig. 2 Relative synonymous codon usage (RSCU) in the mitogenome of C. pubescens. The lower squares represent all codons encoding each amino acid, and the height of the upper columns represents the sum of all codon RSCU values

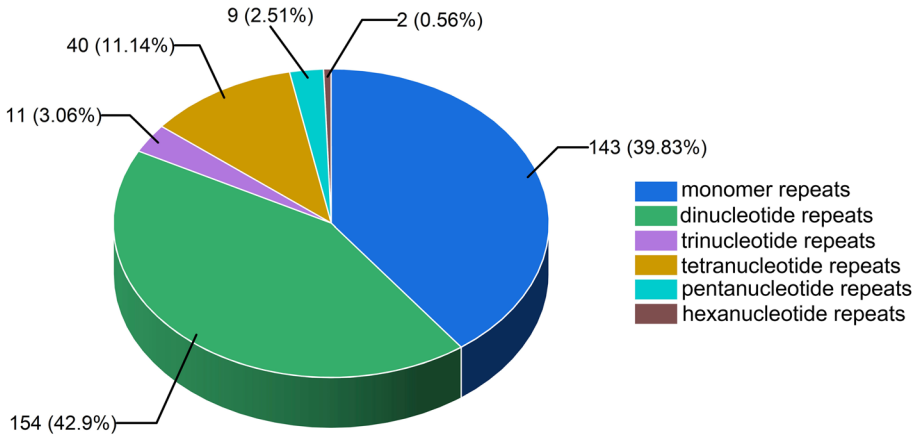

Fig. 3 Analysis of repeat sequences in the C. pubescens mitogenome. (a) Statistics of the number of repetitions of four types of interspersed repeat sequences in the genome. F, P, R, and C represent forward, palindromic, reverse, and complementary repeats, respectively; (b) Statistics of the types and numbers of simple sequence repeats (SSRs)

total length of these interspersed repeats was 38,393 bp, accounting for 8.45% of the mitogenome length (454,165 bp). Most of the repeats we found ranged from 30 to 39 bp in length, with a total of 285 repeats, making up approximately 60.25% of the total repeats. Additionally, we identified 5 repeats that exceeded 1,000 bp in length, ranging from 1,347 bp to 4,425 bp (Table S5). Genes that appeared in large repeats might generate multiple copies [8]. In the *C. pubescens* mitogenome, the gene with the most extensive repeat length of 4,425 bp was trnL-CAA, which currently has three copies (Table S2).

Microsatellites or simple sequence repeats (SSRs) are DNA fragments containing 1–6 base pairs, and they are widely used in genotyping [28, 29]. We discovered 359 SSRs in the *C. pubescens* mitogenome (Table S6). When we examined the different proportions of SSR sites, we found that dinucleotide repeats had the highest number, with 154, accounting for 42.90% of the total SSR count. Monomer repeats followed closely with 143 repeats, accounting for 39.83%. The number of hexanucleotide repeats was the lowest, at only 2, accounting for a mere 0.56% (Fig. 3b). Among the monomer SSRs, 86.71% are monomer repeats composed of A/T bases. Among dinucleotide SSRs, 64.29% are dinucleotide repeats composed of AG/CT bases.

Tandem repeats can be found in eukaryotic and some prokaryotic genomes [30]. They usually refer to repetitive sequences formed by taking 1–200 bases as the repeating unit and then connecting them in series, also known as satellite DNA. By setting the matching degree to greater than 95%, we identified 17 tandem repeats with lengths of 7—39 bp in the *C. pubescens* mitogenome (Table S7).

Prediction of RNA editing sites

An online website (http://cloud.genepioneer.com:9929/#/tool/alltool/detail/336) was used to predict RNA editing sites. Prediction of RNA editing sites for 35 PCGs in the mitogenome of *C. pubescens* revealed 448 RNA editing sites (Table S8). Among these RNA editing sites, 89 (19.87%) appeared in the first base position of the codon, 359 (80.13%) in the second base position, and no RNA editing occurred in the third base position. These PCGs have varying numbers of RNA editing sites, ranging from 2 to 35. The *rps10* gene and the *sdh3* gene have at least 2 RNA editing sites, while the *ccmB* gene has at most 35 RNA editing sites (Fig. 4). Table S9 shows that the hydrophilicity of 11.83% (53 sites) of the amino acids did not change after RNA editing. On the other hand, 47.32% (212 sites) of the sites are predicted to change from hydrophilic to hydrophobic. A total of 32.81% (147 sites) of the amino acids remained unchanged in hydrophobicity, while 7.37% (33 sites) were predicted to change from hydrophobic to hydrophilic. Additionally, 0.67% (3 sites) of the amino acids changed from hydrophilic to terminated. Furthermore, most amino acids tended to be converted from Ser to Leu (23.88%, 107 sites), Pro to Leu (21.88%, 98 sites), and Ser to Phe (14.51%, 65 sites). The remaining 178 RNA editing sites are distributed among other RNA editing types, including His to Tyr, Arg to Cys, Thr to Ile, Thr to Met, Arg to Trp, Ser to Leu, Ser to Phe, Pro to Ser, Pro to Leu, Pro to Phe, Leu to Phe, Ala to Val, and Gln to X (X = stop codon).

We aligned the RNA-seq data (NCBI project number PRJNA822667) with the mitogenome data and searched for potential RNA editing sites (Table S10). The results revealed 454 RNA editing sites in 35 PCGs, with a frequency of approximately 15.6% (71 out of 454) at the

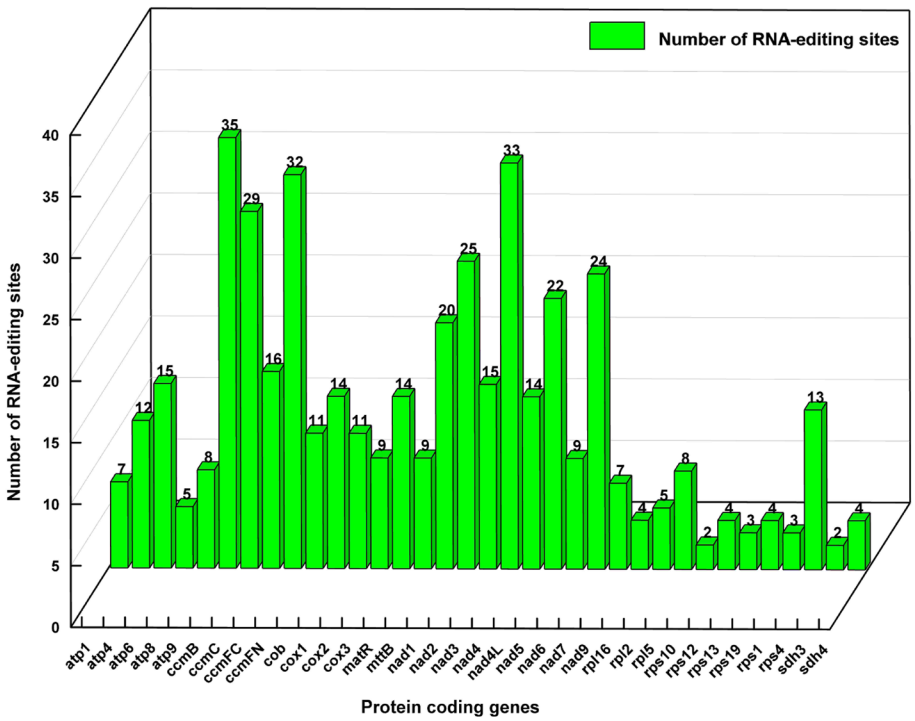

Fig. 4 Distribution of RNA editing sites in PCGs of the C. pubescens mitogenome. An online website (http://prep.unl.edu/) was used to predict RNA editing sites

third codon position. It is evident that this result is not solely achieved through software prediction, highlighting the necessity of verifying it using RNA-seq data.

Phylogenetic analysis of mitogenomes in higher plants

A phylogenetic tree was created for the *C. pubescens* mitogenome and the mitogenomes of 14 other species, including 9 eudicots, 3 monocots, and 2 gymnosperms, to explain the evolution of the *C. pubescens* mitogenome. Twenty-three PCGs from 15 species were used for analysis, including the *matR*, *cox2*, *cox3*, *cox1*, *atp9*, *nad4L*, *atp1*, *rpl16*, *rps4*, *rps12*, *nad4*, *nad3*, *nad2*, *nad1*, *ccmFN*, *ccmFftC*, *nad7*, *nad6*, *nad5*, *cob*, *ccmC*, *ccmB*, and *nad9* genes. The gene sequences of each species were concatenated head-to-tail, and then multiple sequence alignments were performed using MAFFT v7.490 software (default parameters). The aligned sequences were imported into MEGA 11 to predict the most appropriate amino acid substitution model (GTR + G). A maximum-likelihood phylogenetic tree was constructed using MEGA 11 (with the bootstrap value set to 1,000). As shown in Fig. 5, among the 12 nodes of the phylogenetic tree, only one node's bootstrap support value was less than or equal to 90, while the other 11 nodes' bootstrap support value was greater than 90. Phylogenetic trees strongly support the division of eudicots and monocots into two clades and the division of angiosperms and gymnosperms into two clades. The clustering results of this phylogenetic tree conform to the hierarchical relationships among species at different family and genus levels. Our clustering results of common PCGs based on the mitogenome are reliable. The scientific names and GenBank accession numbers are shown in Table S11.

Comparison of mitogenome size, GC content, and PCGs between *C. pubescens* and other species

Organellar genomes have two main parameters, size and GC content. The mitogenome size and GC content of *C. pubescens* were compared with those of 18 other green plants, including 9 eudicots, 3 monocots, 2 gymnosperms, 2 bryophytes, and 2 phycophyta. The species abbreviations of these plants and their mitogenome GenBank accession numbers are displayed in Table S11. Figure 6 show that the mitogenomes ranged in size from 62,477 bp (*Chlorella heliozoae*) to 680,603 bp (*Zea mays*). Compared with those of land plants, the mitogenomes of phycophyta and bryophytes were generally smaller, while the mitogenome of *C. pubescens* (474,330 bp) was was at an average level. Similarly, the GC content of the mitogenome varied, ranging from 32.24% in *C. heliozoae* to 50.36% in *Ginkgo biloba*. Generally, the GC content of angiosperms, including eudicots and monocots, is lower

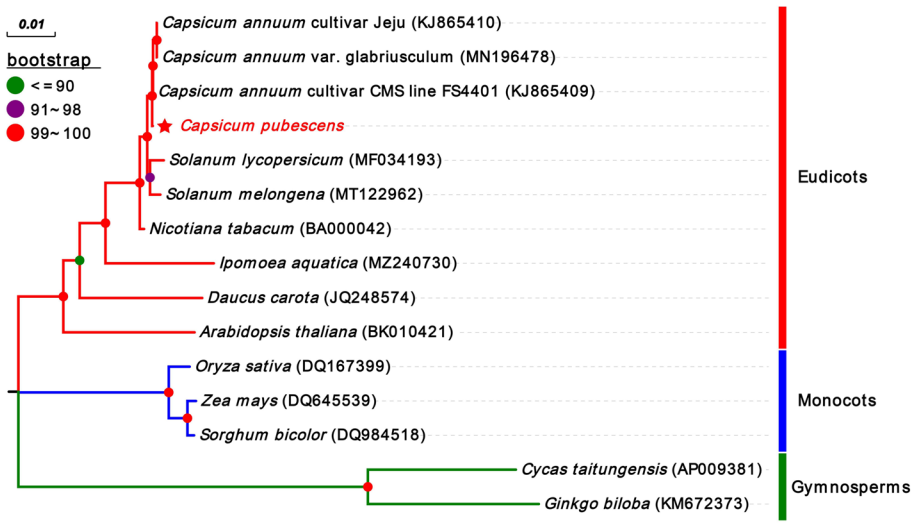

Fig. 5 Phylogenetic tree of 23 common conserved PCGs of the mitogenomes of C. pubescens and 14 other species (9 eudicots, 3 monocots, and 2 gymnosperms)

than that of gymnosperms but higher than that of phycophyta and bryophytes. Remarkably, *C. heliozoae* and *Nitella hyalina*, with GC contents of 32.24% and 41.00%, respectively, indicate that the GC contents of phycophyta fluctuate widely. In contrast, although mitogenome sizes in angiosperms vary widely, ranging from 281,132 bp (*Daucus carota*) to 680,603 bp (*Z. mays*), their GC content has been remarkably conserved throughout evolution and is usually stable at approximately 44%.

PCGs in plants are often lost during evolution [16]. We investigated the loss of PCGs in eudicots, monocots, gymnosperms, bryophytes, and phycophyta (Fig. 7). Compared to highly variable ribosomal proteins and genes encoding succinate dehydrogenase, most PCGs are conserved in different plant mitogenomes, especially in the following genes: ATP synthase, cytochrome *c* biogenesis, ubichinol cytochrome *c* reductase, cytochrome *c* oxidase, maturases, transport membrane protein, and NADH dehydrogenase [31]. To some extent, the conservation of these genes indicates that they play a crucial role in plant mitochondrial function. As indicated by the red dashed box, compared with the other three pepper mitogenomes, 4 PCGs (*atp4*, *atp8*, *mttB*, and *rps1*) were unique to *C. pubescens*, while 2 PCGs (*rpl10* and *rps3*) were missing (Fig. 7).

Non-synonymous (Ka) and synonymous (Ks) mutation rate analysis of PCGs

In this study, we calculated the Ka/Ks values of 30 PCGs shared by the *C. pubescens* mitogenome and the mitogenomes of three peppers (KJ865409, KJ865410, and MN196478) using TBtools software (Table S12). Following a previous study, we changed the Ka/Ks ratio of the genes that do not apply (NA) to zero [32]. The results indicate that the PCGs shared by *C. pubescens* and the other three peppers are close homologs because most PCGs have Ka and Ks values of zero. When comparing *C. pubescens* and KJ865409, only two genes (*sdh3* and *rps10*) had Ka/Ks values less than 0, indicating that they underwent negative selection. In contrast, the Ka/Ks values of the remaining genes were all equal to 0 or -Infinity (Ks = 0). In the other two comparison combinations, the Ka/Ks values of the *sdh3* and *rps10* genes were also less than 0. Furthermore, the only gene that underwent negative selection was *rpl16* in the comparison of *C. pubescens* and MN196478. The genes that underwent positive selection during the evolutionary process, with a Ka/Ks value greater than 1 (*nad5*), existed only in the combination of *C. pubescens* vs. KJ865410 and *C. pubescens* vs. MN196478.

To further elucidate the sequence differences among the 30 PCGs in the four pepper mitogenomes, we used MAFFT v7.490 software (default parameters) to carry out multiple sequence alignments for the 30 PCGs, respectively (Fig. 8). In general, 13 PCGs had sequence differences caused by at least one base change, although the remaining 17 PCGs had no sequence differences. PCGs with sequence differences included *matR*, *rps10*, *rpl2*, *nad1*, *nad2*, *nad4*, *nad5*, *rps4*, *cox2*, *ccmFC*, *ccmFN*, *rpl16*, and *sdh3*. Among them, *rps10*, *rpl16*, and *sdh3* had a wide range of base deletions (more than 10 bases) in the four peppers, but more deletions occurred in KJ865409, KJ865410, and MN196478. There are a few basic changes in the remaining PCGs. Some PCGs, such as *matR*, *rpl2*,

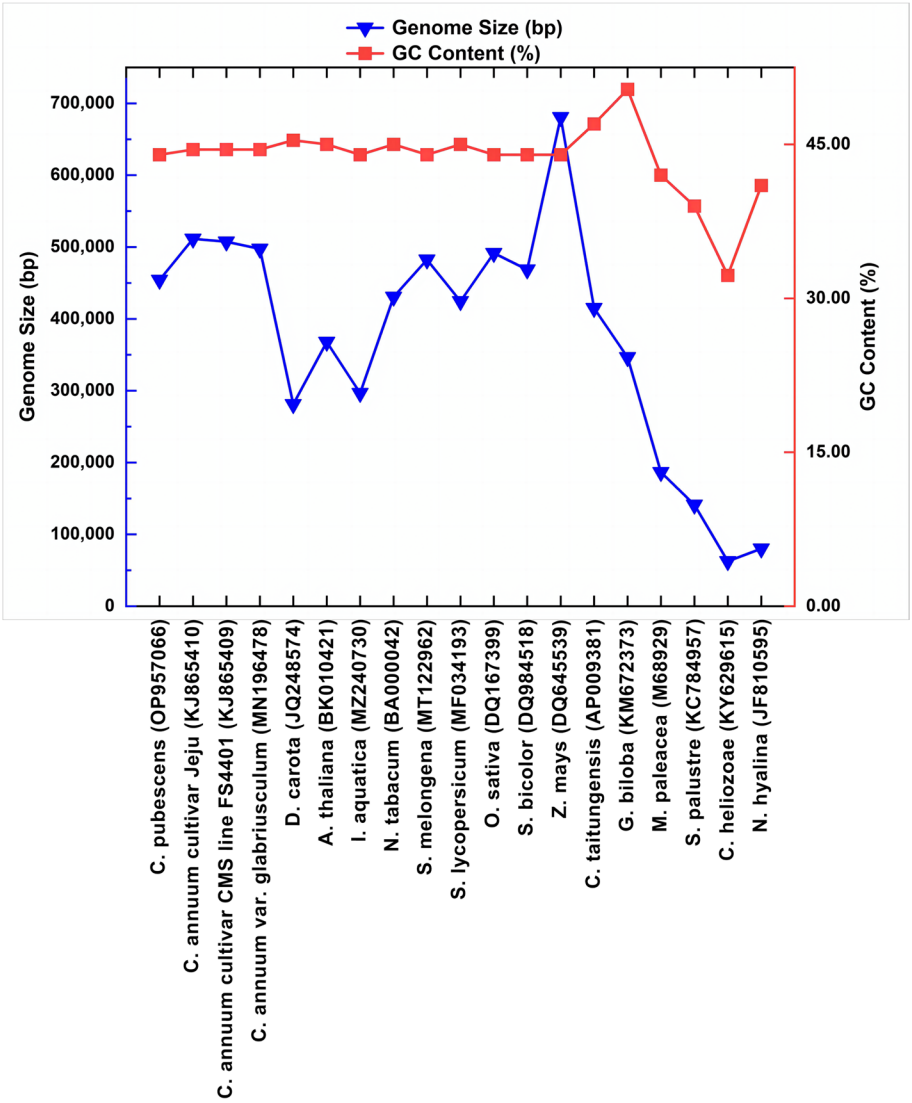

Fig. 6 Comparison of the mitogenome size and GC content of C. pubescens with those of other species

rps4, and *ccmFC*, are not affected in their amino acids as a result of base substitution. Some PCGs, including *nad1*, *nad2*, *ccmFN*, and *nad4*, exhibit amino acid changes as a result of base substitution.

Notably, the only gene that underwent positive selection was the *nad5* gene. With *C. pubescens* as a reference, the *nad5* gene of the other two peppers (KJ865410 and MN196478) had a substitution of the C/T base at the 1,863rd base (Fig. 8a). However, due to the base change, the amino acid remains unchanged, resulting in a synonymous mutation. Furthermore, between the 1,447th and 1,468th bases, KJ865410 and MN196478 have multiple base substitutions, leading to changes in amino acids, causing nonsynonymous mutations. However, the difference in the sequence of the *nad5* gene between *C. pubescens* and KJ865409 was not significant, with only a substitution of the C/T base at the 1,863rd position. Additionally, the amino acid remains unchanged due to the base change.

In summary, a significantly greater number of genes undergo negative selection than positive selection. Moreover, base substitutions in some PCGs may result in changes in their corresponding amino acids.

Structural comparison and synteny analysis of mitogenomes

The online tool Proksee (https://proksee.ca/) was used to perform comparative analyses of mitogenome structures across species. The first to sixth tracks, moving from the outside to the inside of the circles, indicate the results

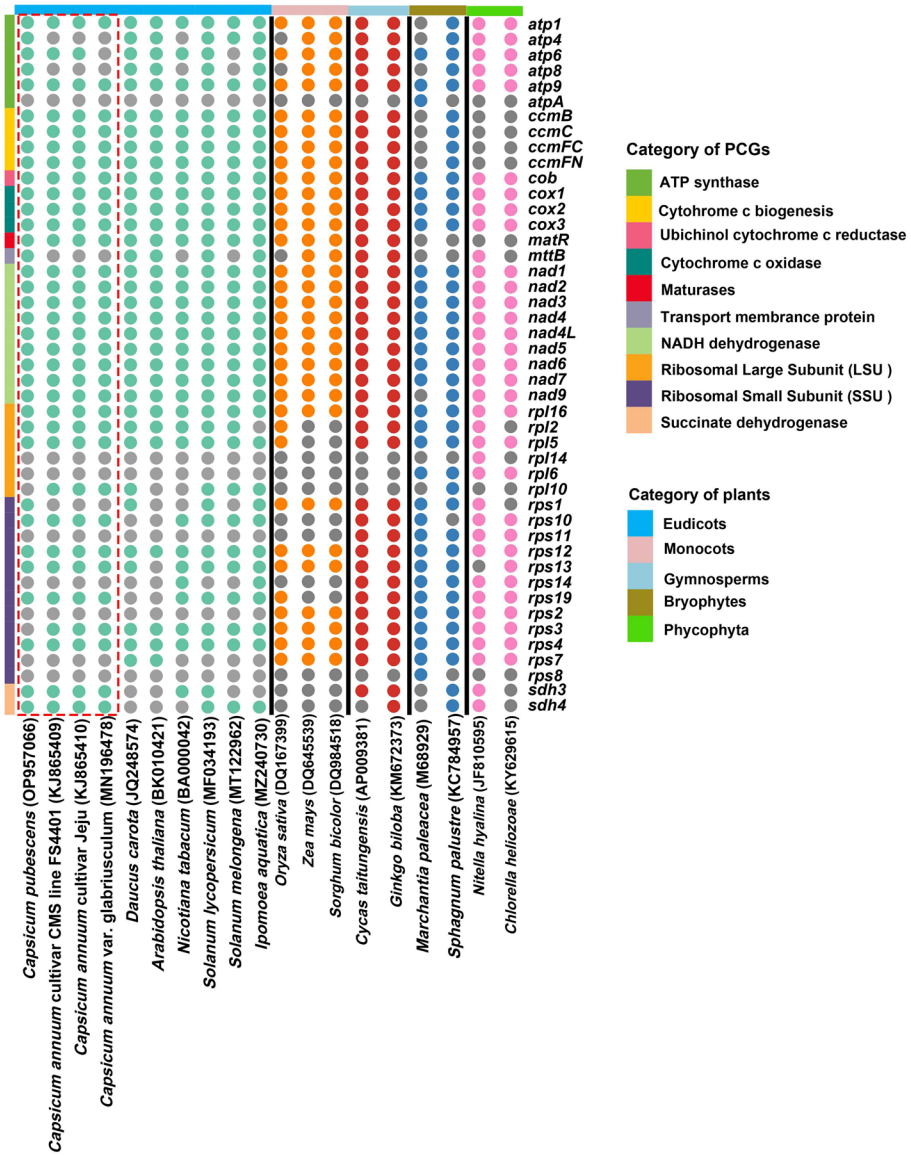

Fig. 7 Distribution of PCGs in plant mitogenomes. The gray circles indicate that the gene does not exist in the corresponding mitogenomes. The light green, orange, rose-red, light blue, and pink circles indicate that PCGs exist in the mitogenomes, corresponding to eudicots, monocots, gymnosperms, bryophytes, and phycophyta, respectively

of BLAST alignment of the DNA sequences of *C. pubescens* with those of *A. thaliana* (BK010421), *S. melongena* (MT122962), *S. lycopersicum* (MF034193), *C. annuum* cultivar Jeju (KJ865410), *C. annuum* var. *glabriusculum* (MN196478), and *C. annuum* cultivar CMS line FS4401 (KJ865409) (Fig. 9a). *C. pubescens* exhibited greater sequence similarity to KJ865409 than to several other groups. Additionally, KJ865410 shows a high degree of sequence similarity with MN196478, which is consistent with the results demonstrated by the phylogenetic tree constructed using the conserved PCGs between them (Fig. 5). The figure clearly illustrates that, in comparison to BK010421 of Brassicaceae, *C. pubescens* shares greater sequence similarity with other species of Solanaceae at the family and genus levels.

The sequence of *C. pubescens* was compared with that of three other pepper mitogenomes, KJ865409, MN196478, and KJ865410, using mauve software (with default parameters) to analyze the sequence homology and synteny relationships between them (Fig. 9b). Using *C. pubescens* as the reference genome, the results indicated that *C. pubescens* exhibited extensive

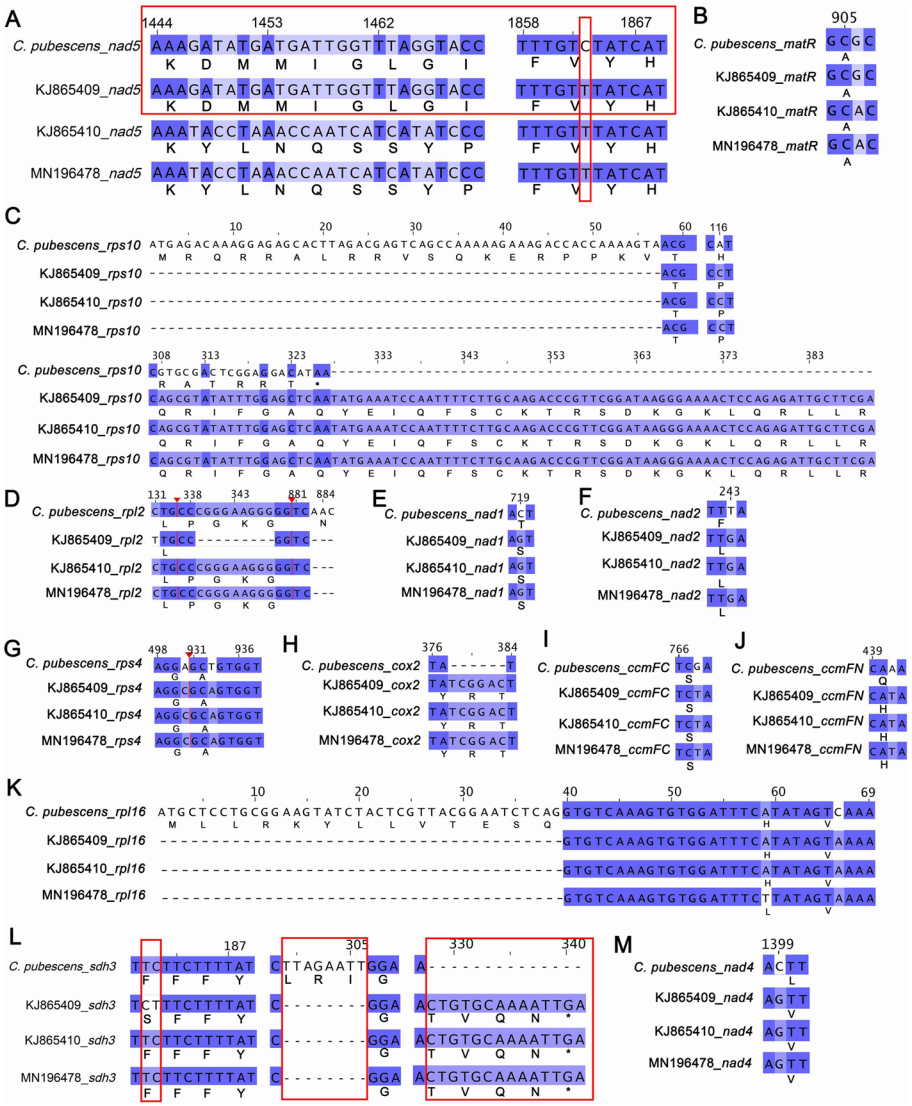

Fig. 8 Alignments of the nucleotide and amino acid sequences of PCGs from four pepper mitogenomes in C. pubescens and C. annuum. We performed multiple sequence alignments using MAFFT v7.490 software with the default parameters. (a-m) Represented as nad5, matR, rps10, rpl2, nad1, nad2, rps4, cox2, ccmFC, ccmFN, rpl16, sdh3, and nad4 genes, respectively. KJ865409, KJ865410, and MN196478 are the mitogenomes of C. annuum

recombination with the other three pepper mitogenomes. The relative positions and orders of the homologous blocks of KJ865410 and MN196478 were mostly identical, indicating a high degree of homology between them. Furthermore, KJ865409, MN196478, and KJ865410 displayed numerous inverted homologous blocks, suggesting that rearrangement events may have occurred.

Homologous sequence analysis of the cp genome and mitogenome

The cp genome is highly conserved compared with the mitogenome. Furthermore, gene transfer from the mitogenome to the cp genome, known as PTMT, rarely occurs, while gene transfer from the cp genome to the mitogenome, known as MTPT, is relatively common. Finally, we detected 72 fragments with a combined length of 57,207 bp that migrated from the *C. pubescens* cp genome to the mitogenome, accounting for 12.60% of the mitogenome size (454,165 bp) (Fig. 10a, Table S13). These fragments contained 13 annotated genes, including 12 tRNA genes (*trnL-CAA*, *trnD-GTC*, *trnE-TTC*, *trnT-GGT*, *trnY-GTA*, *trnR-ACG*, *trnP-TGG* , *trnW-CCA*, *trnM-CAT*, *trnS-GGA*, *trnN-GTT*, and *trnH-GTG*) and one rRNA gene (*rrn18*). Our analysis

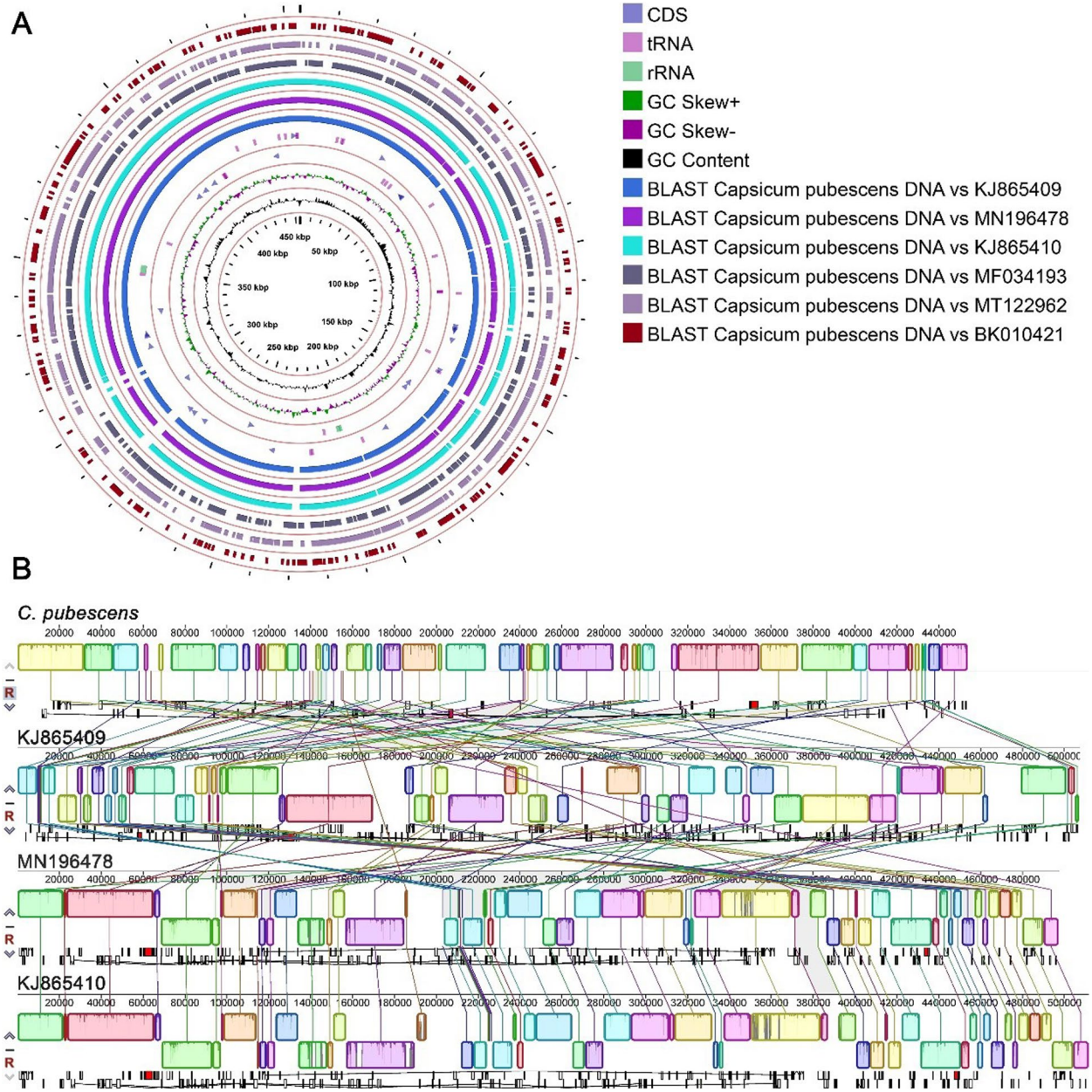

Fig. 9 Comparison of mitogenome structure and collinearity analysis. (a) The six outermost tracks represent the similarity results of the mitogenome alignment of C. pubescens and several other species. Furthermore, the four tracks represent positive-strand genes, minus-strand genes, the GC Skew, and the GC content in the mitogenome of C. pubescens. The innermost circle represents the C. pubescens genome size. (b) The rectangles represent the similarity between genomes, and the lines between the rectangles represent a collinear relationship. The short white squares represent CDSs, the short green squares represent tRNAs, and the short red squares represent rRNAs. The two graphs were drawn using the Proksee (https://proksee.ca/) online tool and Mauve (https://darlinglab.org/mauve/mauve.html) software

also revealed that some genes, such as *ycf3*, *atpF*, *rpl2*, and *ndhB*, migrated from the cp genome to the mitogenome. However, during migration, most of these genes lost their integrity, and only partial fragments could be found in the mitogenome (Table S13). Figures 10b and c show the circular map of the cp genome and a schematic diagram of the partial fragments of the *ycf3* gene migrating to the mitogenome, respectively. After the *ycf3* gene (1,991 bp) in the cp genome migrated to the mitogenome, only a 282 bp fragment remained, while the remaining 1,709 bp fragment was lost.

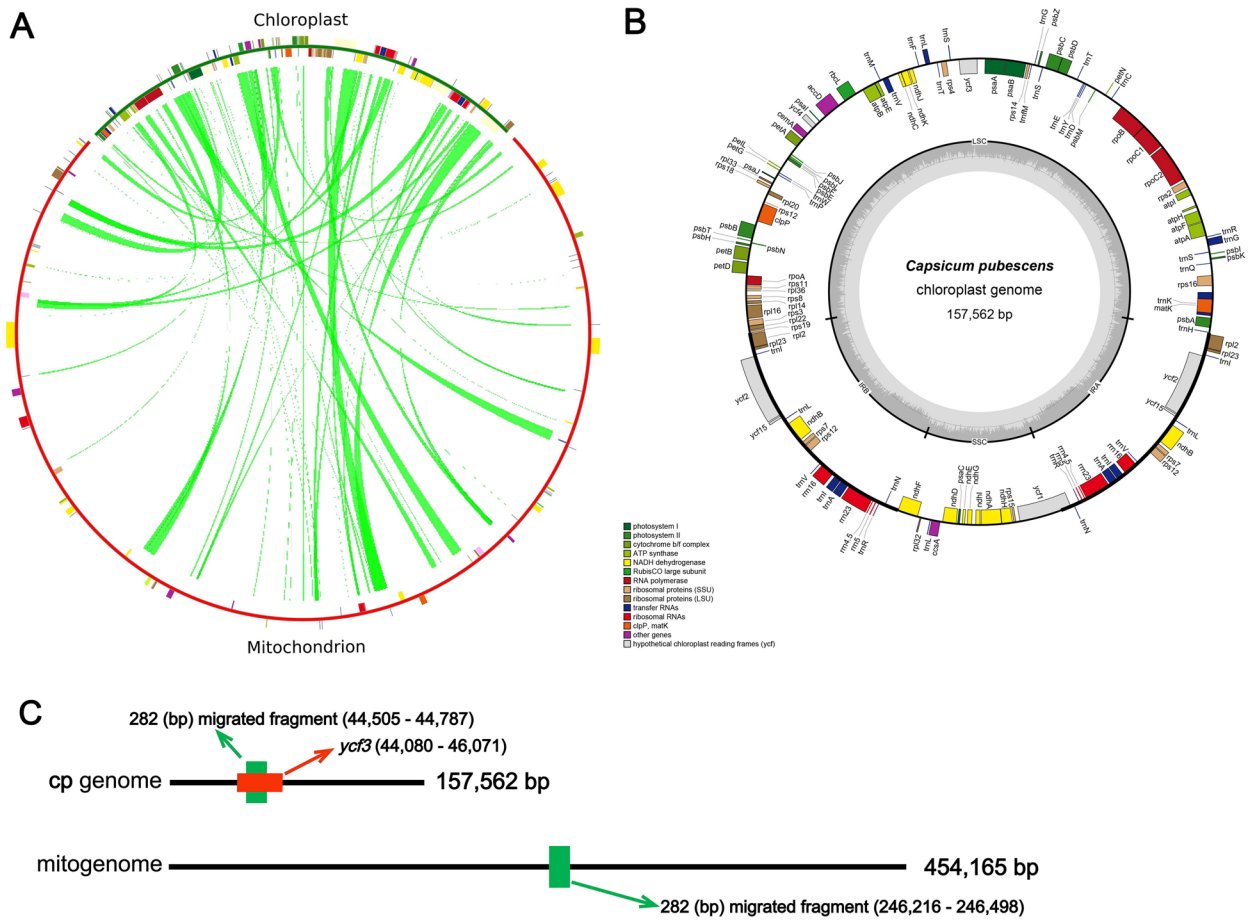

Fig. 10 Homologous sequence analysis of the cp genome and mitogenome. (a) Homologous fragments of chloroplast and mitochondrial sequences. "Chloroplast" represents chloroplast sequences, and "Mitochondrion" represents mitochondrial sequences. Genes from the same complex are marked with the same color, and the middle green line joins the homologous sequences; (b) Circular map of the cp genome. The genes encoded by the upper strand are located outside the circle, while the genes encoded by the lower strand are on the inside. The inner gray circle represents the GC content. (c) The schematic diagram illustrates the migration of the ycf3 gene of the cp genome to the mitogenome. The upper and lower black lines represent the cp genome and mitogenome, respectively. For visual observation, the two genomes are depicted as linear. Circos v0.69–5 software was used to visualize and map the homologous sequences between chloroplasts and mitochondria. Additionally, a circular map of the chloroplast genome was drawn using the OGDRAW online site

Discussion

Genomic characteristics, PCGs, and intronic genes of the *C. pubescens* mitogenome

Mitochondria, often referred to as the "powerhouses" of cells, play a vital role in plant life. In recent years, advances in sequencing technology have facilitated the gradual assembly of numerous plant mitogenomes. Compared to animal mitogenomes, plant mitogenomes tend to exhibit greater complexity, as reflected by their variable sizes and multiple repetitive sequences [8, 16, 33]. In this study, we conducted a detailed characterization of the *C. pubescens* mitogenome, providing the first comprehensive account of its features. The mitogenome of *C. pubescens* possesses a circular structure with a length of 454,165 bp and includes 35 PCGs. The GC content is significant for evaluating the application of species. The GC content of the *C. pubescens* mitogenome was 44.29%, which is basically consistent with the GC content of several peppers that have been sequenced (*C. annuum* var. *glabriusculum*, 44.50% [34]; *C. annuum* cultivar Jeju, 44.50% [32]; *C. annuum* cultivar CMS line FS4401, 44.50%) [35]. During the course of evolution, plant mitogenomes often experience the loss of PCGs [16, 36, 37]. When we compared the mitogenomes of *C. pubescens* with those of three other pepper species (*C. annuum* var. *glabriusculum*, *C. annuum* cultivar Jeju, and *C. annuum* cultivar CMS line FS4401), we observed that *C. pubescens* possessed four unique PCGs (*atp4*, *atp8*, *mttB*, and *rps1*), while two genes (*rpl10* and *rps3*) were absent (Fig. 7). Interestingly, our findings align with previous

research showing that the *rpl10* gene was lost in monocots and gymnosperms but was later recovered during evolutionary processes in eudicots [8, 22]. Additionally, we observed that the *rps7* gene was present solely in *D. carota* (JQ248574) and *A. thaliana* (BK010421) of the eudicots, suggesting that it was lost after the common ancestor of Solanaceae and Convolvulaceae began to diverge. Conversely, the *rps10* gene appears in all Solanaceae and Convolvulaceae, but it is lost in *D. carota* of Apiaceae and *A. thaliana* of Brassicaceae. These findings indicate that the *rps10* gene might have been lost during the separation of Umbelliferae and Brassicaceae. However, during subsequent evolutionary processes, the common ancestor of Solanaceae and Convolvulaceae began to diverge and reacquire the gene.

Introns are indeed prevalent in fully sequenced eukaryotic genomes [38, 39]. Despite the fact that introns need to be removed during RNA splicing in eukaryotic cells, they possess significant functions. One notable function of introns in eukaryotes is their contribution to increased protein abundance in intronic genes [40, 41]. As the number of introns increases, so does the number of intronic genes. Moreover, certain introns can greatly enhance gene expression levels. Typically, substances containing introns tend to exhibit much higher expression levels than substances without introns [42–44]. In our study, we identified ten intronic genes (*ccmFC*, *cox2*, *nad1*, *nad2*, *nad4*, *nad5*, *nad7*, *rpl2*, *rps1*, and *trnY-GTA*) in the *C. pubescens* mitogenome, totaling twenty-four intronic genes. Among them, four genes (*nad1*, *nad2*, *nad5*, and *nad7*) contained four introns, while only the *nad4* gene contained three introns. Additionally, five genes (*ccmFC*, *cox2*, *rpl2*, *rps1*, and *trnY-GTA*) contained a single intron. The presence of introns in these PCGs within the mitogenome of *C. pubescens* suggests that their functional significance warrants further study.

Repeat sequences

Repeated sequences are indeed widespread in plant mitogenomes and play an important role in the intermolecular recombination that occurs within mitochondria [45]. Among these repeated sequences, one type that is particularly notable is Microsatellites or Simple Sequence Repeats (SSRs). SSRs consist of short stretches of DNA composed of repetitive units, typically ranging from 1 to 6 nucleotides in length [46]. SSRs have a variety of functions and have been widely used to identify molecular markers of species, for QTL mapping, for marker-assisted breeding, for diversity analysis, and for establishing evolutionary relationships [47–51]. Our analysis of the *C. pubescens* mitogenome revealed a diverse array of SSRs. Dinucleotide repeats were the most abundant, accounting for 42.90% of the total SSRs, with 154 occurrences. Monomeric repeats were the second most frequent, accounting for 39.83% (143 occurrences), while hexanucleotide repeats were the least common, accounting for only 0.56% (2 occurrences). Previous studies have indicated that SSRs containing AT base repeats are more prevalent in organelle genomes due to the relative ease of breaking AT bonds compared to GC bonds [52]. By applying a matching threshold of 95%, we identified 17 tandem repeat sequences in the *C. pubescens* mitogenome, ranging in length from 7 to 39 bp. Tandem repeats, which are found in many plant mitogenomes, can be utilized as molecular markers for population identification [53]. Additionally, we identified 473 interspersed repeats longer than 30 bp, with a cumulative length of 38,393 bp. These larger interspersed repeats are primarily responsible for genome rearrangements [54, 55]. Notably, we found five repeats exceeding 1 kb in length, ranging from 1,347 bp to 4,425 bp.

RNA editing

RNA editing plays a crucial role in gene expression in plants [56]. Specifically, the conversion of cytosine (C) to uracil (U) at specific sites in the cp genomes and mitogenomes alters genetic information [57]. In our analysis of the *C. pubescens* mitogenome, we identified 448 RNA editing sites in 35 PCGs. These editing sites encompassed 30 different codon transfer types. Among them, the TCA to TTA transition had the greatest number of RNA editing sites, with 70 occurring. Interestingly, this type of editing results in amino acid changes from hydrophilic to hydrophobic properties. Previous studies have shown that RNA editing at the second position within a codon is widespread, often accounting for more than half of the total editing events [52, 57]. Similarly, we found that 359 RNA editing sites (80.13%) occurred at the second base position, consistent with these previous findings. Furthermore, we aligned the RNA sequencing (RNA-seq) data from project number PRJNA822667 with the *C. pubescens* mitogenome data. This analysis revealed a high frequency of RNA editing events at the second codon position. Out of a total of 454 editing sites, 236 were observed at the second codon position, comprising more than half of the total editing sites (Table S10). Notably, all the observed editing sites were of the CT type, which is the most common form of RNA editing in plant mitogenomes [58].

Curiously, we observed no RNA editing events at the third base position in our study. Some researchers suggest that this absence may be due to limitations in predicting RNA editing using the PREP-mt program rather than the actual absence of editing [16, 59]. To address this, we rigorously aligned the RNA-seq data (project number PRJNA822667) with the mitogenome data

and identified potential RNA editing sites (Table S10). Through this analysis, we revealed 71 RNA editing events at the third codon position out of the 454 identified editing sites. This highlights the fact that these results are not solely obtained through software prediction but also emphasizes the importance of validating them using RNA-seq data.

Non-synonymous mutation rate (Ka) and synonymous mutation rate (Ks)

Calculating the Ka and Ks of PCGs provides valuable insights into phylogenetic reconstruction and understanding of gene evolutionary dynamics [60]. The Ka/Ks ratio can indicate different selective pressures acting on genes. A Ka/Ks > 1 suggests positive selection, a Ka/Ks = 1 implies neutral evolution, and a Ka/Ks < 1 indicates negative selection. In our study, the Ka/Ks values of the *sdh3* and *rps10* genes in the three comparison combinations (*C. pubescens* vs. KJ865409, KJ865410, and MN196478) were less than 1. This signifies that these genes experienced negative selection. Mitochondrial genes undergoing negative selection likely play a crucial role in maintaining the normal function of mitochondria [16]. Additionally, we observed that only the Ka/Ks value of the *nad5* gene was greater than 1 when comparing *C. pubescens* with KJ865410 and MN196478. This indicates that the *nad5* gene may have undergone positive selection since its divergence from its last common ancestor. These results suggest that different mitochondrial genes may have experienced various selection pressures throughout evolution [16]. To visually observe the changes in Ka and Ks caused by sequence differences among the four pepper species, we conducted multiple sequence alignments for the 30 PCGs (Fig. 8). Among them, *rps10*, *rpl16*, and *sdh3* had a wide range of base deletions (more than 10 bases) in the four peppers. However, more deletions occurred in KJ865409, KJ865410, and MN196478. For the remaining PCGs, there were a few base substitutions. Some PCGs did not change their amino acids due to base substitution, such as *matR*, *rpl2*, *rps4*, and *ccmFC*. Some PCGs, such as *nad1*, *nad2*, *ccmFN*, and *nad4*, exhibit amino acid changes due to base substitutions (Fig. 8). Typically, a change in the third base of a codon results in a synonymous mutation (Ks), while changes in the first and second bases lead to non-synonymous mutations (Ka) [61]. According to the definitions of Ka and Ks, when an amino acid change occurs due to a base change, it generally causes a change in Ka. Conversely, when a base change does not result in an amino acid change, it generally causes Ks. Regardless of whether it is Ka or Ks, it often affects the mRNA expression level of the mutated gene [62]. According to the sequence alignment of the *nad5* gene, there is a C/T base substitution at the 1,863rd base between *C. pubescens* and KJ865409, which is located at the third base of the Val (V) amino acid. However, this base change did not result in an amino acid change, so the difference in the *nad5* gene between *C. pubescens* and KJ865409 only caused Ks (Ks = 0.0021, Fig. 8 and Table S12). For the *nad5* genes of KJ865410 and MN196478, Ks also exists at the 1,863rd base, and there are multiple base substitutions between the 1,447th and 1,468th bases, resulting in amino acid changes, thus causing Ka. Sequence alignment also revealed that the number of synonymous mutation sites was less than the number of non-synonymous mutation sites. Overall, Ka/Ks > 1 indicates that the gene is under positive selection. Such a gene is a gene that is currently undergoing evolution, similar to the *nad5* gene identified in this study, which has paramount significance for species evolution research in *C. pubescens*.

Migration of homologous sequence fragments from the cp genome to the mitogenome

The transfer of DNA fragments from the organelle genome to the nuclear genome often occurs, and it is also common for DNA fragments to transfer from the cp genome to the mitogenome [25, 63]. Nevertheless, the transfer of DNA fragments from the mitogenome to the cp genome has been reported infrequently. Researchers believe that this phenomenon may be because the plant cp genome is highly conserved, so foreign DNA rarely enters the cp genome [64, 65]. Previous studies have shown that the migration of mitochondrial DNA to chloroplasts occurs in four species (*D. carota*, *Asclepias syriaca*, *Anacardium occidentale*, and herbaceous bamboos). In the mitogenome of *D. carota*, a 74 bp *cox1* gene sequence migrates from the mitogenome to the cp genome, and this sequence is named *Dc*MP [66–68]. An ~ 2.7 kb sequence was inserted into the IR region of the herbaceous bamboos cp genome, which originated from the mitogenome and was subsequently confirmed [69]. *A. syriaca* transfers the *rpl2* pseudogene (*ψrpl2*) from the mitogenome to the cp genome [70]. In *A. occidentale*, *ccmB* is transferred [71]. Through identification, 72 fragments totaling 57,207 bp migrated from the cp genome to the mitogenome of *C. pubescens*, accounting for 12.60% of the mitogenome (454,165 bp). The transfer of tRNA genes from the cp genome to the mitogenome is more common in angiosperms than in bryophytes and gymnosperms [22, [54, 72, 73]. We found 13 annotated genes on these transferred fragments, including 12 tRNA genes (*trnL-CAA*, *trnD-GTC*, *trnE-TTC*, *trnT-GGT*, *trnY-GTA*, *trnR-ACG*, *trnP-TGG*, *trnW-CCA*, *trnM-CAT*, *trnS-GGA*, *trnN-GTT*, and *trnH-GTG*) and one rRNA gene (*rrn18*). Some chloroplast PCGs, such as *ycf3*, *atpF*, *rpl2*, and *ndhB*, have also migrated from the

cp genome to the mitogenome, although many of them have lost their integrity through evolution. Researchers suggest that these lost genes are likely to be found in the nuclear genome, where they are regulated as coding genes through related expression regulation [74, 75]. Let us take the *ycf3* gene of the cp genome as an example. We searched for homologous gene sequences in the nuclear genomes of three peppers (*C. annuum*, *C. baccatum*, and *C. chinense*) using the *ycf3* gene as a query sequence (Table S14). The obtained sequences can then be screened for GO enrichment function analysis. We found that these genes are mainly involved in the generation of precursor metabolites and energy (Biological Process, GO:0006091), NADH dehydrogenase activity (Molecular Function, GO:0003954), and Membrane protein complex (Cellular Component, GO:0098796). We identified only one GO term for each functional enrichment process, and additional results can be found in Table S15.

Conclusion

In this study, we assembled the first mitogenome of *C. pubescens* (GenBank accession number OP957066), which is 454,165 bp in length and has a typical circular structure. The mitogenome of *C. pubescens* had a GC content of 44.29%, and 70 genes were annotated. Compared to those of the other three pepper mitogenomes (KJ865409, KJ865410, and MN196478), the mitogenome of *C. pubescens* contained four unique PCGs (*atp4*, *atp8*, *mttB*, and *rps1*), while two PCGs (*rpl10* and *rps3*) were absent. Structural comparison and synteny analysis of these four pepper mitogenomes revealed that *C. pubescens* is most similar to KJ865409 but exhibits rearrangements with the other three mitogenomes. We identified 72 homologous sequences, accounting for 12.60% (57,207 bp) of the mitogenome, between the mitochondrial and chloroplast genomes. These sequences included 12 tRNA genes and the rRNA gene (*rrn18*). Notably, the *nad5* gene shows signs of positive selection based on selective pressure analysis. Additionally, three genes (*nad1*, *nad2*, and *nad4*) exhibit single-base mutations that result in amino acid changes. Through comparative genome analysis, we can gain insight into the genetic information exchange between pepper organelle genomes. This exploration will be helpful for future studies.

Methods

Plant material, DNA extraction, and sequencing

The pepper (HNUCP0006) seeds used in this study were obtained from the Pepper Germplasm Bank of Hainan University (E: 110°33′52.29", N: 20°06′22.48"). The seeds were wrapped with gauze and placed in water at 50 °C for treatment, after which the water was allowed to cool naturally. After 6 h, the seeds were removed from the water bath and kept in a dark environment for 5–7 days while ensuring that they remained moist. The treated seeds were sown in 50-cell plug trays and planted in a plant growth room at Hainan University, where the temperature was maintained at 26 °C and the light–dark cycle was 16/8 h. After a period of two months, the healthy pepper plants were transferred from the plug trays to a dark environment and allowed to grow for 7 days. During this time, the temperature was maintained at room temperature, and watering was performed once on the 0th day, the third day, and the sixth day. On the eighth day, the dark treatment of the pepper plants was discontinued, resulting in pepper plants with etiolated leaves. Ten grams of etiolated leaves were collected, 5 g was used as a backup, and the isolation process was conducted at Nanjing Jisihuiyuan Biotechnology Co., Ltd. Mitochondrial genomic DNA was extracted from the samples, and the Illumina NovaSeq 6000 platform was used for next-generation sequencing. The second-generation raw data were filtered to obtain high-quality reads using fastp software (v0.20.0). Third-generation sequencing was performed using an Oxford Nanopore PromethION sequencer, and the third-generation sequencing data were filtered out using Filtlong software (v0.2.1).

Assembly and annotation of the mitogenome

All three-generation sequencing data for the mitogenome were acquired using Minimap2 software (v2.1). The obtained three-generation data were then corrected using Canu software with the default settings. The correction process with Canu involved error correction, trimming, and assembly of the raw sequencing reads.

Next, the second-generation sequencing data were aligned to the corrected three-generation sequences using Bowtie2 software (v2.3.5.1). The alignment was performed with the default parameters to ensure accurate mapping of the second-generation reads to the corrected sequences.

For the assembly step, Unicycler (v0.4.8) was used to splice and compare the second-generation sequencing data with the corrected third-generation data. Unicycler was run with the default parameter settings, which include initial read alignment, error correction, and hybrid assembly using both second and third-generation data.

To ensure accurate extraction of mitochondrial reads, the mitochondrial reference genome from *C. annuum* (GenBank accession number KJ865410) was used. The reference genome facilitated the identification and extraction of mitochondrial sequences from the sequencing data, ensuring that only relevant reads were included in the assembly process.

The tRNA of the *C. pubescens* mitogenome was annotated using tRNAscanSE software. Open Reading Frame Finder was used to annotate the open reading frames (ORFs), with the minimum sequence length set to 102 bp. Redundant sequences were removed and sequences with a length exceeding 300 bp were aligned with the NR database for annotation. Finally, the encoded protein and rRNA sequences were compared with published plant mitogenome sequences through a BLAST search, and manual adjustments were made based on closely related species for further refinement. The OGDRAW online tool was used for mapping the mitogenome.

Prediction of RNA editing using software and validation with RNA-seq data

The online website PmtREP (http://cloud.genepioneer.com:9929/#/tool/alltool/detail/336) was used to predict RNA editing sites. The RNA-seq data (project number PRJNA822667) were aligned with the mitochondrial genome data to identify potential RNA editing sites. First, Bowtie2 software (version 2.3.5.1) was used to align the RNA sequencing data with the coding sequences (CDSs) of the PCGs in the mitogenome. Next, SAMtools software (version 1.9) was used to filter out sequences with alignment quality scores greater than or equal to 40. Sites with single nucleotide polymorphisms (SNPs) between the sequencing data and the reference genome were identified.

For a site to be considered a potential RNA editing site, it must meet the following criteria: SNP depth exceeding 3×and at least 20% of the total depth at that site.

Codon composition and RSCU analysis

Codons exhibit degeneracy, with each amino acid having 1–6 corresponding codons. The uneven usage of synonymous codons is commonly referred to as codon bias or "relative synonymous codon usage (RSCU)". RSCU analysis was completed by the data analysis system of Nanjing Jisi Huiyuan Biotechnology Co., Ltd (http://cloud.genepioneer.com:9929/#/tool/alltool/detail/214).

Repeat sequence analysis

We utilized vmatch software (v2.3.0) to identify interspersed repeats. The criteria we set are a minimum length of 30 base pairs and a Hamming distance of 3. To investigate simple sequence repeats (SSRs), we used MISA online software. We set a minimum distance of 100 base pairs between two SSRs. Through this analysis, we identified 8, 4, 4, 3, 3, and 3 repeats, corresponding to 1, 2, 3, 4, 5, and 6 bases, respectively. For the detection of tandem repeats, we employed Tandem Repeats Finder v4.09 software, which identifies matches more significantly than 95%.

Phylogenetic analysis and comparison of genome size and GC content

We downloaded 18 green plant mitogenomes from NCBI and performed phylogenetic analysis. This analysis included the mitogenomes of *C. pubescens* and 14 other species, including 9 eudicots, 3 monocots, and 2 gymnosperms. Multiple sequence alignments of 23 common PCGs in 15 species were performed using MAFFT v7.490 software (using default parameters) [76]. The sequences were manually trimmed or retained based on their similarity and identity in the alignment results. For the construction of the phylogenetic tree, we utilized MEGA 11 and set the bootstrap value to 1,000 using the Maximum-likelihood method. The amino acid substitution model selected was GTR+G. To visualize the evolutionary tree, we utilized Evolview v2 software [77]. Furthermore, we compared the size and GC content of the *C. pubescens* mitogenome with those of 18 other green plant mitogenomes and visualized the results using Origin 2023 software.

Analysis of non-synonymous (Ka) and synonymous (Ks) mutation rates

The Ka and Ks values for 30 common PCGs in *C. pubescens* and three other pepper species were calculated using TBtools (v1.0987663) software [78]. The pepper species included *C. annuum* cultivar Jeju (KJ865410), *C. annuum* cultivar CMS line FS4401 (KJ865409), and *C. annuum* var. glabriusculum (MN196478). The coding sequence (CDS) was utilized for the calculations, and the analysis was conducted using the Simple Ka/Ks Calculator (NG) tool within the TBtools software, applying default parameters.

Comparison of mitogenome structures and collinearity analysis

The online tool Proksee (https://proksee.ca/) was used for comparative analysis of the mitogenome structure between *C. pubescens* and other species. The species included in the analysis were *A. thaliana*, *S. melongena*, *S. lycopersicum*, the *C. annuum* cultivar Jeju, *C. annuum* var. *glabriusculum*, and the *C. annuum* cultivar CMS line FS4401. Mauve (https://darlinglab.org/mauve/mauve.html) software was used for sequence homology and collinearity analysis between the *C. pubescens* mitogenomes and those of three other peppers, namely, the *C. annuum* cultivar CMS line FS4401, *C. annuum* var. *glabriusculum*, and *C. annuum* cultivar Jeju.

Chloroplast and mitochondrial homologous sequence analysis

BLAST (v2.10.1) software was used to detect homologous sequence fragments (identity% > 70%, E-value = 10E-5) between the mitogenome and cp genome of *C. pubescens*. The cp genome of *C. pubescens*

was assembled using the next-generation sequencing results obtained in this study. Among the software used, the chloroplast genome was assembled using GetOrganelle (https://github.com/Kinggerm/GetOrganelle), which required calling SPAdes, Bowtie2, BLAST+, and Bandage. A circular map of the cp genome was created using the web application OGDRAW. The nuclear genomes of *C. baccatum* and *C. chinense* were obtained from the NCBI database, while the nuclear genome of *C. annuum* was retrieved from the Ensembl Plants Database.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10985-w.

Additional file 1: Table S1: Gene profile and organization of the *C. pubescens* mitogenome. Table S2: Gene annotation profile of the *C. pubescens* mitogenome. Table S3: Protein-coding gene start codon and stop codon utilization. Table S4: Relative synonymous codon usage (RSCU) in the *C. pubescens* mitogenome. Table S5: Long interspersed repeated sequences in the *C. pubescens* mitogenome. Table S6: Distribution of SSRs in the *C. pubescens* mitogenome. Table S7: Statistics of tandem repeat sequences in the mitogenome of *C. pubescens*. Table S8: Statistics on the RNA editing sites of protein-coding genes in *C. pubescens*. Table S9: Statistics of RNA editing sites in the *C. pubescens* mitogenome. Table S10: RNA editing sites predicted by alignment of RNA-seq data with mitogenome data. Table S11: The abbreviations, GenBank accession numbers, genome sizes, and GC contents of the mitogenomes used in this study. Table S12: Nonsynonymous mutation rate (Ka) and nonsynonymous mutation rate (Ks) calculation of 23 common PCGs. Table S13: Homologous sequence analysis of chloroplasts and mitochondria. Table S14: Using the *ycf3* gene as a query sequence, we searched for homologous sequences in three pepper genomes. Table S15: GO enrichment analysis.

Acknowledgements
Thanks to our group members (Yuanyuan Hao, Yu Zhang, Muhammad Ali Mumtaz, Huangying Shu, Liping Zhang, and Shanhan Cheng) for their excellent suggestions for the research and writing of this article. In addition, special thanks to the anonymous reviewers for their valuable time and suggestions on this manuscript.

Authors' contributions
L.L. and X.L. planned and designed the research; Data curation, L.L. and X.L.; Formal analysis, H.F.; Investigation, L.L. and M.A.; Methodology, M.A. and Z.W.; Software, H.F.; Visualization, L.L.; Writing—original draft, L.L.; Writing—review & editing, X.L. and Z.W.. All authors reviewed the manuscript.

Funding
This research was funded by grants from the Collaborative Innovation Center of Nanfan and High-Efficiency Tropical Agriculture (XTCX2022NYB03), Hainan University, the Hainan Province Science and Technology Special Fund (ZDYF2023XDNY028), and the National Key Research and Development Program of China (2018YFD1000800).

Data availability
Sequence data that support the findings of this study have been deposited in the National Center for Biotechnology Infommation with the primary accession code OP957066.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]National Key Laboratory for Tropical Crop Breeding, School of Breeding and Multiplication (Sanya Institute of Breeding and Multiplication) , Hainan University, Sanya Hainan 572025, China. [2]Key Laboratory for Quality Regulation of Tropical Horticultural Crops of Hainan Province, School of Tropical Agriculture and Forestry, Hainan University, Haikou 570228, China.

Received: 4 June 2024 Accepted: 30 October 2024
Published online: 11 November 2024

References

1. Ďúranová H, Ivanišová E, Galovičová L, Godočíková L, Borotová P, Kunová S, Miklášová K, Lopašovský Ľ, Mňahončáková E. Antioxidant and antimicrobial activities of fruit extracts from different fresh chili peppers. Acta Sci Pol Technol Aliment. 2021;20(4):465–72.
2. Jarret RL, Barboza GE, da Costa Batista FR, Berke T, Chou Y-Y, Hulse-Kemp A, Ochoa-Alejo N, Tripodi P, Veres A, Garcia CC. *Capsicum*—An abbreviated compendium. J Am Soc Hortic Sci. 2019;144(1):3–22.
3. Palombo NE, Carrizo GC. Geographical patterns of genetic variation in Locoto chile (*Capsicum pubescens*) in the americas inferred by genome-wide data analysis. Plants. 2022;11(21):2911.
4. Ibiza VP, Blanca J, Canizares J, Nuez F. Taxonomy and genetic diversity of domesticated *Capsicum* species in the Andean region. Genet Resour Crop Ev. 2012;59(6):1077–88.
5. Rodríguez-Burruezo A, Prohens J, Raigón MD, Nuez F. Variation for bioactive compounds in ají (Capsicum baccatum L) and rocoto (C pubescens R & P) and implications for breeding. Euphytica. 2009;170:169–81.
6. Barboza GE, García CC, de Bem Bianchetti L, Romero MV, Scaldaferro M. Monograph of wild and cultivated chili peppers (Capsicum L, Solanaceae). PhytoKeys. 2022;200:1.
7. Gao C, Mumtaz MA, Zhou Y, Yang Z, Shu H, Zhu J, Bao W, Cheng S, Yin L, Huang J. Integrated transcriptomic and metabolomic analyses of cold-tolerant and cold-sensitive pepper species reveal key genes and essential metabolic pathways involved in response to cold stress. Int J Mol Sci. 2022;23(12):6683.
8. Ye N, Wang X, Li J, Bi C, Xu Y, Wu D, Ye Q. Assembly and comparative analysis of complete mitochondrial genome sequence of an economic plant *Salix suchowensis*. PeerJ. 2017;5:e3148.
9. Fan W, Liu F, Jia Q, Du H, Chen W, Ruan J, Lei J, Li DZ, Mower JP, Zhu A. Fragaria mitogenomes evolve rapidly in structure but slowly in sequence and incur frequent multinucleotide mutations mediated by microinversions. New Phytol. 2022;236(2):745–59.
10. Zhang X, Zhou W, Cheng X, Chen X, Hu X, Hu Y, Wang X. Assessing the ecological and evolutionary processes underlyingcytonuclear interactions. Scientia Sinica Vitae. 2019;49(8):951–64.
11. Wang X, Li LL, Xiao Y, Chen XY, Chen JH, Hu XS. A complete sequence of mitochondrial genome of *Neolamarckia cadamba* and its use for systematic analysis. Sci Rep. 2021;11(1):21452.
12. Wang Y, Chen S, Chen J, Chen C, Lin X, Peng H, Zhao Q, Wang X. Characterization and phylogenetic analysis of the complete mitochondrial genome sequence of *Photinia serratifolia*. Sci Rep. 2023;13(1):770.
13 Wang J, Kan S, Liao X, Zhou J, Tembrock LR, Daniell H, Jin S, Wu Z. Plant organellar genomes: much done, much more to do. Trends Plant Sci. 2024;29(7):754–69.
14. Mahapatra K, Banerjee S, De S, Mitra M, Roy P, Roy S. An insight into the mechanism of plant organelle genome maintenance and implications of organelle genome in crop improvement: an update. Front Cell Dev Biol. 2021;9:671698.
15. Odahara M, Nakamura K, Sekine Y, Oshima T. Ultra-deep sequencing reveals dramatic alteration of organellar genomes in Physcomitrella patens due to biased asymmetric recombination. Commun Biol. 2021;4(1):633.

16. Bi C, Lu N, Xu Y, He C, Lu Z. Characterization and analysis of the mitochondrial genome of common bean (*Phaseolus vulgaris*) by comparative genomic approaches. Int J Mol Sci. 2020;21(11):3778.
17. Xiong Y, Yu Q, Xiong Y, Zhao J, Lei X, Liu L, Liu W, Peng Y, Zhang J, Li D, Bai S, Ma X. The complete mitogenome of *Elymus sibiricus* and insights into its evolutionary pattern based on simple repeat sequences of seed pant mitogenomes. Front Plant Sci. 2021;12:802321.
18. Bi C, Shen F, Han F, Qu Y, Hou J, Xu K, Xu L-a, He W, Wu Z, Yin T. PMAT: an efficient plant mitogenome assembly toolkit using low-coverage HiFi sequencing data. Hortic Res. 2024;11(3):uhae023.
19. Bi C, Sun N, Han F, Xu K, Yang Y, Ferguson DK. The first mitogenome of Lauraceae (Cinnamomum chekiangense). Plant Diversity. 2024;46(1):144.
20. Woloszynska M, Trojanowski D. Counting mtDNA molecules in Phaseolus vulgaris: sublimons are constantly produced by recombination via short repeats and undergo rigorous selection during substoichiometric shifting. Plant Mol Biol. 2009;70(5):511–21.
21 Wu K, Yang M, Liu H, Tao Y, Mei J, Zhao Y. Genetic analysis and molecular characterization of Chinese sesame (Sesamum indicum L) cultivars using insertion-deletion (InDel) and simple sequence repeat (SSR) markers. BMC Genet. 2014;15(1):35.
22 Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K. The complete sequence of the rice (Oryza sativa L) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. Mol Genet Genomics. 2002;268:434–45.
23. Skippington E, Barkman TJ, Rice DW, Palmer JD. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes. Proc Natl Acad Sci U S A. 2015;112(27):E3515–24.
24. Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol. 2012;10(1):e1001241.
25. Bergthorsson U, Adams KL, Thomason B, Palmer JD. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature. 2003;424(6945):197–201.
26. Wynn EL, Christensen AC. Repeats of unusual size in plant mitochondrial genomes: identification, incidence and evolution. G3 (Bethesda). 2019;9(2):549–59.
27. Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD. The "fossilized" mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. BMC Biol. 2013;11(1):29.
28. Liu YC, Liu S, Liu DC, Wei YX, Liu C, Yang YM, Tao CG, Liu WS. Exploiting EST databases for the development and characterization of EST-SSR markers in blueberry (Vaccinium) and their cross-species transferability in *Vaccinium* spp. Sci Hortic-Amsterdam. 2014;176:319–29.
29. Grover A, Sharma PC. Development and use of molecular markers: past and present. Crit Rev Biotechnol. 2016;36(2):290–302.
30. Wang H. pepReap: a peptide identification algorithm using support vector machines. J Comp Res Dev. 2005;42(9):555–64.
31. Chang S, Wang Y, Lu J, Gai J, Li J, Chu P, Guan R, Zhao T. The mitochondrial genome of soybean reveals complex genome structures and gene evolution at intercellular and phylogenetic levels. PLoS ONE. 2013;8(2):e56502.
32. Androsiuk P, Okorski A, Paukszto L, Jastrzebski JP, Ciesielski S, Pszczolkowska A. Characterization and phylogenetic analysis of the complete mitochondrial genome of the pathogenic fungus *Ilyonectria destructans*. Sci Rep. 2022;12(1):2359.
33. Chevigny N, Schatz-Daas D, Lotfi F, Gualberto JM. DNA repair and the stability of the plant mitochondrial genome. Int J Mol Sci. 2020;21(1):328.
34. Magdy M, Ouyang B. The complete mitochondrial genome of the chiltepin pepper (Capsicum annuum var glabriusculum), the wild progenitor of Capsicum annuum L. Mitochondrial DNA B Resour. 2020;5(1):683–4.
35. Jo YD, Choi Y, Kim DH, Kim BD, Kang BC. Extensive structural variations between mitochondrial genomes of CMS and normal peppers (Capsicum annuum L) revealed by complete nucleotide sequencing. BMC Genomics. 2014;15(1):561.
36. Kan S-L, Shen T-T, Gong P, Ran J-H, Wang X-Q. The complete mitochondrial genome of *Taxus cuspidata* (Taxaceae): eight protein-coding genes have transferred to the nuclear genome. BMC Evol Biol. 2020;20:1–17.
37. Sullivan AR, Eldfjell Y, Schiffthaler B, Delhomme N, Asp T, Hebelstrup KH, Keech O, Öberg L, Møller IM, Arvestad L. The mitogenome of *Norway spruce* and a reappraisal of mitochondrial recombination in plants. Genome Biol Evol. 2020;12(1):3586–98.
38. Gilson PR, Su V, Slamovits CH, Reith ME, Keeling PJ, McFadden GI. Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. Proc Natl Acad Sci U S A. 2006;103(25):9566–71.
39. Gozashti L, Roy SW, Thornlow B, Kramer A, Ares M Jr, Corbett-Detig R. Transposable elements drive intron gain in diverse eukaryotes. Proc Natl Acad Sci U S A. 2022;119(48):e2209766119.
40. Chorev M, Carmel L. The function of introns Front Genet. 2012;3:55.
41. Grabski DF, Broseus L, Kumari B, Rekosh D, Hammarskjold ML, Ritchie W. Intron retention and its impact on gene expression and protein diversity: A review and a practical guide. WIREs RNA. 2021;12(1):e1631.
42. Buchman AR, Berg P. Comparison of intron-dependent and intron-independent gene expression. Mol Cell Biol. 1988;8(10):4395–405.
43. Clark AJ, Archibald AL, McClenaghan M, Simons JP, Wallace R, Whitelaw CB. Enhancing the efficiency of transgene expression. Philos Trans R Soc Lond B Biol Sci. 1993;339(1288):225–32.
44. Fu J, Fu Y-W, Zhao J-J, Yang Z-X, Li S-A, Li G-H, Quan Z-J, Zhang F, Zhang J-P, Zhang X-B. Improved and flexible HDR editing by targeting introns in iPSCs. Stem Cell Reviews and Reports. 2022;18(5):1822–33.
45. Dong S, Zhao C, Chen F, Liu Y, Zhang S, Wu H, Zhang L, Liu Y. The complete mitochondrial genome of the early flowering plant *Nymphaea colorata* is highly repetitive with low recombination. BMC Genomics. 2018;19(1):614.
46. Li Q, Su X, Ma H, Du K, Yang M, Chen B, Fu S, Fu T, Xiang C, Zhao Q. Development of genic SSR marker resources from RNA-seq data in *Camellia japonica* and their application in the genus *Camellia*. Sci Rep. 2021;11(1):1–11.
47 Cao Z, Wang P, Zhu X, Chen H, Zhang T. SSR marker-assisted improvement of fiber qualities in Gossypium hirsutum using G barbadense introgression lines. Theor Appl Genet. 2014;127(3):587–94.
48 El-Esawi MA. SSR analysis of genetic diversity and structure of the germplasm of faba bean (Vicia faba L). C R Biol. 2017;340(11–12):474–80.
49. Ma Q, Li S, Bi C, Hao Z, Sun C, Ye N. Complete chloroplast genome sequence of a major economic species, *Ziziphus jujuba* (Rhamnaceae). Curr Genet. 2017;63(1):117–29.
50. Li D, Long C, Pang X, Ning D, Wu T, Dong M, Han X, Guo H. The newly developed genomic-SSR markers uncover the genetic characteristics and relationships of olive accessions. PeerJ. 2020;8:e8573.
51. Shukla RP, Tiwari GJ, Joshi B, Song-Beng K, Tamta S, Boopathi NM, Jena SN. GBS-SNP and SSR based genetic mapping and QTL analysis for drought tolerance in upland cotton. Physiol Mol Biol Plants. 2021;27(8):1731–45.
52. Qiao Y, Zhang X, Li Z, Song Y, Sun Z. Assembly and comparative analysis of the complete mitochondrial genome of *Bupleurum chinense* DC. BMC Genomics. 2022;23(1):1–17.
53. Sperisen C, Buchler U, Gugerli F, Matyas G, Geburek T, Vendramin GG. Tandem repeats in plant mitochondrial genomes: application to the analysis of population differentiation in the conifer *Norway spruce*. Mol Ecol. 2001;10(1):257–63.
54. Ogihara Y, Yamazaki Y, Murai K, Kanno A, Terachi T, Shiina T, Miyashita N, Nasuda S, Nakamura C, Mori N, Takumi S, Murata M, Futo S, Tsunewaki K. Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. Nucleic Acids Res. 2005;33(19):6235–50.
55. Alverson AJ, Rice DW, Dickinson S, Barry K, Palmer JD. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. Plant Cell. 2011;23(7):2499–513.
56. Ma Q, Wang Y, Li S, Wen J, Zhu L, Yan K, Du Y, Ren J, Li S, Chen Z, Bi C, Li Q. Assembly and comparative analysis of the first complete mitochondrial genome of *Acer truncatum* Bunge: a woody oil-tree species producing nervonic acid. BMC Plant Biol. 2022;22(1):29.
57. Cheng Y, He X, Priyadarshani S, Wang Y, Ye L, Shi C, Ye K, Zhou Q, Luo Z, Deng F, Cao L, Zheng P, Aslam M, Qin Y. Assembly and comparative analysis of the complete mitochondrial genome of *Suaeda glauca*. BMC Genomics. 2021;22(1):167.
58. Edera AA, Sanchez-Puerta MV. Computational detection of plant RNA editing events. RNA Editing: Methods and Protocols. 2021;2181:13–34.
59. Mower JP. PREP-Mt: predictive RNA editor for plant mitochondrial genes. BMC Bioinf. 2005;6:96.

60. Fay JC, Wu CI. Sequence divergence, functional constraint, and selection in protein evolution. Annu Rev Genomics Hum Genet. 2003;4(1):213–35.
61. Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet. 2002;18(9):486.
62. Shen X, Song S, Li C, Zhang J. Synonymous mutations in representative yeast genes are mostly strongly non-neutral. Nature. 2022;606(7915):725–31.
63. Martin W, Stoebe B, Goremykin V, Hapsmann S, Hasegawa M, Kowallik KV. Gene transfer to the nucleus and the evolution of chloroplasts. Nature. 1998;393(6681):162–5.
64. Richardson AO, Palmer JD. Horizontal gene transfer in plants. J Exp Bot. 2007;58(1):1–9.
65. Smith DR. Extending the limited transfer window hypothesis to inter-organelle DNA migration. Genome Biol Evol. 2011;3:743–8.
66. Goremykin VV, Salamini F, Velasco R, Viola R. Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. Mol Biol Evol. 2009;26(1):99–110.
67. Iorizzo M, Grzebelus D, Senalik D, Szklarczyk M, Spooner D, Simon P. Against the traffic: The first evidence for mitochondrial DNA transfer into the plastid genome. Mob Genet Elements. 2012;2(6):261–6.
68. Iorizzo M, Senalik D, Szklarczyk M, Grzebelus D, Spooner D, Simon P. De novo assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome. BMC Plant Biol. 2012;12:1–17.
69. Ma PF, Zhang YX, Guo ZH, Li DZ. Evidence for horizontal transfer of mitochondrial DNA to the plastid genome in a *bamboo* genus. Sci Rep. 2015;5(1):11608.
70. Straub SC, Cronn RC, Edwards C, Fishbein M, Liston A. Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (apocynaceae). Genome Biol Evol. 2013;5(10):1872–85.
71. Rabah SO, Lee C, Hajrah NH, Makki RM, Alharby HF, Alhebshi AM, Sabir JSM, Jansen RK, Ruhlman TA. Plastome Sequencing of Ten Nonmodel Crop Species Uncovers a Large Insertion of Mitochondrial DNA in Cashew. Plant Genome. 2017;10(3):plantgenome2017.03.0020.
72. Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Ohyama K. Transfer RNA genes in the mitochondrial genome from a liverwort, *Marchantia polymorpha*: the absence of chloroplast-like tRNAs. Nucleic Acids Res. 1992;20(14):3773–7.
73. Chaw S-M, Chun-Chieh Shih A, Wang D, Wu Y-W, Liu S-M, Chou T-Y. The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short interspersed elements, Bpu sequences, and abundant RNA editing sites. Mol Biol Evol. 2008;25(3):603–15.
74. Bryant N, Lloyd J, Sweeney C, Myouga F, Meinke D. Identification of nuclear genes encoding chloroplast-localized proteins required for embryo development in Arabidopsis. Plant Physiol. 2011;155(4):1678–89.
75. Savage LJ, Imre KM, Hall DA, Last RL. Analysis of essential Arabidopsis nuclear genes encoding plastid-targeted proteins. PLoS ONE. 2013;8(9):e73291.
76. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
77. He Z, Zhang H, Gao S, Lercher MJ, Chen WH, Hu S. Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. Nucleic Acids Res. 2016;44(W1):W236–41.
78. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 2020;13(8):1194–202.

Publisher's Note