BMC Genomics

RESEARCH

Open Access

Comparative mitochondrial genomics of *Terniopsis yongtaiensis* in Malpighiales: structural, sequential, and phylogenetic perspectives

Miao Zhang[1†], Xiaohui Zhang[1,2†], Yinglin Huang[1,2], Zhangxue Chen[1,2] and Binghua Chen[1,2*]

Abstract

Background *Terniopsis yongtaiensis*, a member of the Podostemaceae family, is an aquatic flowering plant displaying remarkable adaptive traits that enable survival in submerged, turbulent habitats. Despite the progressive expansion of chloroplast genomic information within this family, mitochondrial genome sequences have yet to be reported.

Results In current study, the mitochondrial genome of the *T. yongtaiensis* was characterized by a circular genome of 426,928 bp encoding 31 protein-coding genes (PCGs), 18 tRNAs, and 3 rRNA genes. Our comprehensive analysis focused on gene content, repeat sequences, RNA editing processes, intracellular gene transfer, phylogeny, and codon usage bias. Numerous repeat sequences were identified, including 130 simple sequence repeats, 22 tandem repeats, and 220 dispersed repeats. Phylogenetic analysis positioned *T. yongtaiensis* (Podostemaceae) within the Malpighiales order, showing a close relationship with the Calophyllaceae family, which was consistent with the APG IV classification. A comparative analysis with nine other Malpighiales species revealed both variable and conserved regions, providing insights into the genomic evolution within this order. Notably, the GC content of *T. yongtaiensis* was distinctively lower compared to other Malpighilales, primarily due to variations in non-coding regions and specific protein-coding genes, particularly the *nad* genes. Remarkably, the number of RNA editing sites was low (276), distributed unevenly across 27 PCGs. The dN/dS analysis showed only the *ccmB* gene of *T. yongtaiensis* was positively selected, which plays a crucial role in cytochrome *c* biosynthesis. Additionally, there were 13 gene-containing homologous regions between the mitochondrial and chloroplast genomes of *T. yongtaiensis*, suggesting the gene transfer events between these organellar genomes.

Conclusions This study assembled and annotated the first mitochondrial genome of the Podostemaceae family. The comparison results of mitochondrial gene composition, GC content, and RNA editing sites provided novel insights into the adaptive traits and genetic reprogramming of this aquatic eudicot group and offered a foundation for future

†Miao Zhang and Xiaohui Zhang contributed equally to this work and share first authorship.

*Correspondence:
Binghua Chen
bhchen@fjnu.edu.cn

Full list of author information is available at the end of the article

research on the genomic evolution and adaptive mechanisms of Podostemaceae and related plant families in the Malpighiales order.

Keywords *Terniopsis yongtaiensis*, Genome size variation, Mitochondrial genome, Evolution, Phylogenetic

Background

Podostemaceae, commonly referred to as "river-weeds", are a unique group of aquatic eudicots found in various wetlands across tropical and subtropical regions worldwide. These plants have undergone remarkable adaptive evolution, enabling them to survive in submerged, turbulent environments with a lifecycle synchronized to seasonal water level fluctuations, resulting in blooming, fruiting, and withering during the dry season [1]. The extreme habitat and atypical lifecycle of Podostemaceae have resulted in morphological deviations from the typical root-shoot structure observed in most seed plants, and often show the thalloid vegetative body due to the dorsiventral flattening of roots, shoots (stems), or a combination of both [2]. This unique feature underscores the profound genetic reprogramming that has occurred during their evolution from terrestrial to aquatic habitats [3]. In this study, we focused on the mitochondrial genome of *Terniopsis yongtaiensis*, a species belonging to the Tristichoideae subfamily, which is acknowledged as one of the primitive subfamilies within the Podostemaceae family [2]. Given the small size and atypical morphology of Podostemaceae, posing significant challenges for traditional morphological species identification, organelle genomes have emerged as a potent resource for elucidating taxonomic relationships, phylogenetic histories, and adaptive evolution patterns within this intriguing family.

To adapt to their specific aquatic habitats, the organelles of aquatic plants often undergo adaptive evolution to maintain essential life functions. While chloroplast are the primary sites for harnessing solar energy in plants [4], mitochondria play a pivotal role in energy production, especially under conditions where chlorophyll or light is lacking [5]. Plant mitochondrial genomes, renowned for their complexity and diversity, exhibit considerable variations in size, sequence alignment, repeat numbers, and structure [6]. These genomes, often interspersed with the introns [7], contain various types of repeats, including the simple, tandem, and dispersed repeats [8], providing a rich source of genetic information for understanding genome evolution and dynamics.

Interestingly, despite their relatively large size, plant mitochondrial genomes contain a limited number of genes, comprising 24 core genes and 17 variant genes [9]. This restricted gene content is attributed to the evolutionary loss or transfer of genes to the nucleus during angiosperm evolution, contributing to the stability of coding sequences in the retained genes [9]. However, yet, this genetic streamlining masks the mitochondrial genome's significance as a repository of evolutionary history and a valuable tool for species classification. Comparative genomic analysis of closely related species reveals the dynamic nature of these genomes [10], shedding light on the mechanisms underlying genome evolution and species diversification.

According to the Updated List of National Key Protected Wild Plants (Decree No. 15) issued by China's State Forestry and Grassland Administration and the Ministry of Agriculture and Rural Affairs, all known genera of Podostemaceae found in China are classified as secondarily protected species. Chloroplast genomes of approximately 16 Podostemaceae species have been documented in recent studies, including *Apinagia fucoides* [11], *Marathrum foeniculaceum* [12], *Terniopsis yongtaiensis* [13], and *Polypleurum chinense* [14], etc. However, there remains a gap in our knowledge concerning the mitochondrial genome of this family.

In this study, we investigate the mitochondrial genome of *T. yongtaiensis* using both third- and second-generation sequencing techniques. After assembling and annotating the mitochondrial genome, a comprehensive analysis was conducted to explore its genomic characteristics, repetitive sequences, RNA editing patterns, codon usage bias, and phylogenetic relationships. A comparative analysis with nine other Malpighiales species revealed regions of variations and conservation. Significantly, intergenomic gene transfer phenomena were explored between chloroplast and mitochondrial genomes in *T. yongtaiensis*, underscoring the study's dual focus on mitochondrial genome characterization and inter-organelle genetic exchange. This research not only facilitates molecular marker development and genetic engineering applications but also deepens our understanding into the developmental and evolutionary trajectories of vascular plants, bridging the gap in our understanding of organelle interactions and their evolutionary significance.

Materials and methods

Plant materials, DNA extraction and sequencing

Terniopsis yongtaiensis X.X. Su, Miao Zhang & Bing-Hua Chen, sp. nov.

Type China, Fujian Province, Yongtai County, Fuquan Town, elevation 95 m, 25°51'N, 118°52'E, collected on 2 January 2022 by Bing-Hua Chen. Holotype specimen: Designated as CBH 04587, deposited in the Herbarium of College of Life Sciences, Fujian Normal University (FNU), with the barcode FNU0041314. Isotype specimens: Also deposited in the Herbarium of the College of Life Sci-

ences, Fujian Normal University (FNU), barcoded as FNU0041315. All voucher specimens of *Terniopsis yongtaiensis* are maintained in the Herbarium of the College of Life Sciences, Fujian Normal University (FNU) for future reference and study.

The fresh leaves of *Terniopsis yongtaiensis* was collected from Dehua, Fujian for DNA extraction. In this study, total DNA was extracted from fresh sample using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA). The extracted DNA was subjected to both second-generation and third-generation sequencing using Illumina Novaseq6000 and Oxford Nanopore PromethION, respectively. The sequencing and data filtering processes were performed by Genepioneer Biotechnologies Co. Ltd. (Nanjing, China). Fastp v0.20.0 software (https://github.com/OpenGene/fastp) and Filtlong v0.2.1 software (https://link.zhihu.com/?target=https%3 A//github.com/rrwick/Filtlong) were utilized for filtering the raw sequencing data from both second-generation and third-generation sequencing (Table S2).

Sequence assembly and annotation

Minimap2 v2.1 [15], configured with the parameters -t20-ax map-ont, was used to align the raw third-generation data with a reference gene sequence (plant mitochondrial core genes). Sequences with alignment lengths exceeding 50 bp were selected as candidate sequences for subsequent analysis via a Perl command: perl -ane 'print if(/^@/);if(/NM: i:(\d+)/){$n=$1;$l=0;$l+=$1 while $F[5]=~ /(\d+)[M]/g; if($l>50){print}}'. Among these candidates, sequences exhibiting a higher number of aligned genes and superior alignment quality were chosen as seed sequences. Then, the original third-generation sequencing data was aligned to the seed sequences using Minimap2 v2.1, and sequences with overlaps larger than 1 kb and similarity greater than 70% were added to the seed sequences. This process enabled iterative alignment and the acquisition of all third-generation sequencing data of the mitochondrial genome. Canu snapshot [16] was used to correct the obtained third-generation data, using the parameters -correct genomeSize=500k useGrid=false -nanopore-raw. Subsequently, the second-generation data was aligned to these corrected sequences using bowtie2 v2.3.5.1 [17], with the parameters: --very-sensitive-local -p 20. Both generations of data were then assembled using Unicycler v0.4.8 [18] with default parameters. Finally, Minimap2 was used to align the corrected third-generation sequencing data to the contigs obtained from the second step of Unicycler. Manual determination of branch direction was performed to obtain the final assembly result.

The coding proteins and rRNAs were annotated through comparison with published plant mitochondrial sequences using Blast. tRNAs were annotated using tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/). For ORFs, OpenReading Frame Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) was employed with a minimum length setting of 102 bp, excluding redundant sequences and those with overlaps with known genes. Sequences longer than 300 bp were annotated by aligning them to the Non-Redundant Database. Third-generation sequencing data was aligned to the assembly using Minimap2 to inspect coverage and circular status. Second-generation sequencing data was aligned to the assembly via Bowtie2 to investigate coverage and base correctness. The mitochondrial genome data of *Terniopsis yongtaiensis* has been uploaded to the NCBI database, with Accession number OR818323.

Analysis of repeat sequences

Microsatellite sequence repeats were identified using MISA v2.1 [19] with the parameters "1-10 2-5 3-4 4-3 5-3 6-3". Tandem repeats were identified using TRF v4.09 [20] with the parameters "2 7 7 80 10 50 500 -f -d -m". Dispersed repeats were identified using Blastn v2.10.1 [21] with the parameters "-word_size 7, evalue 1e-5". The results were visualized using the Circos package implemented in TBtools v2.003 [22].

DNA transfer between the chloroplast and the mitochondrion

The chloroplast genome of *Terniopsis yongtaiensis* (NC_066797.1) has been reported in our previous study [13]. Sequence similarities between the chloroplast and mitochondrial genomes were analyzed to identify transferred DNA fragments using the Blast package implemented in TBtools v2.003 with an e-value cut-off of 1e-5 [23]. The results were visualized using the Circos package implemented in TBtools v2.003 [22].

Prediction of RNA editing sites

The PmtREP tool (http://112.86.217.82:9919/#/tool/alltool/detail/336) was utilized to predict RNA editing sites in plant chloroplast and mitochondrial genomes, with a threshold set at 0.2. The number and location of RNA editing sites in 10 plants species within Malpighiales were analyzed, and the stacked bar chart was plotted using Origin 2018. The density of RNA editing sites of each gene (site/kb) was calculated and normalized, and the heatmap was created using Origin 2018.

Analysis of codon usage bias

The protein-coding sequences were extracted using Phylosuite software v.1.1.15, and the relative synonymous codon usage (RSCU) and effective number of codons (ENC) values of the amino acid composition of protein-coding genes from mitochondrial

genome were determined using Genepioneer (http://112.86.217.82:9919/#/tool/alltool/detail/214).

Sequence alignment and phylogenetic analyses

Phylogenetic analyses were performed using Maximum likelihood (ML) and Bayesian Inference (BI) analyses, based on 31 conserved protein-coding genes (PCGs) from mitochondrial genomes (Table S1). Our analysis included 38 species of Alga, Bryophytes, Pteridophyte, Gymnosperm and Angiosperms (Podostemaceae, Calophyllaceae, Euphorbiaceae, Passifloraceae and Salicaceae) to construct the phylogenetic tree. Alga plants was selected as the outgroup. Each individual sequence was aligned using MAFFT 7.310 [24] with default settings. A concatenated supermatrix of the sequences was generated using PhyloSuite v.1.1.15 [25] for the phylogenetic analysis. All missing data were treated as gaps. The best nucleotide substitution model, according to the Bayesian Information Criterion (BIC), was GTR+F+R3, which was selected by Model Finder [26] implemented in IQ-TREE v.1.6.8. Maximum likelihood phylogenies were inferred using IQ-TREE [27] under the model automatically selected by IQ-TREE ("Auto"option in IQ-TREE) for 1000 ultrafast bootstraps [28]. Bayesian Inference phylogenies were inferred using MrBayes 3.2.6 [29] under the GTR+F+G4 model (2 parallel runs, 2000000 generations), with the initial 25% of sampled data discarded as burn-in. The phylograms were visualized in iTOL v.5 [30].

Selective pressure analysis

The dN/dS ratios of 24 common protein-coding sequences among mitochondrial genomes from *Terniopsis yongtaiensis* and nine other plant species in Malpighiales were calculated using PAMLX v1.3.1 [31]. The YN00 module was selected to estimate nonsynonymous substitution rate (dN) and synonymous substitution rate (dS) with the following parameters: "verbose=0, icode=0, ndata=1". A boxplot of pairwise dN/dS values was generated using Origin 2018. For further analysis of the *ccmB* gene, the coding sequences (CDS) were extracted from the mitochondrial genomes of the 10 Malpighiales species. Multiple sequence alignment of the *ccmB* gene was conducted using PhyloSuite v.1.1.15, and the results were visualized with ESPript 3.0 (ESPript 3.x / ENDscript 2.x (ibcp.fr)) to highlight conserved regions and potential RNA editing sites.

Comparison of mitochondrial genomes of related species

The GC content and size of mitochondrial genomes of *Terniopsis yongtaiensis* and 37 other plant species, collected from the NCBI database (accessed in September 2023) (Table S1), were compared using TBtools v2.003 software [22]. The results were then visualized with bar graphs and line graphs created in Origin 2018.

Results

General features of mitochondrial genome of *Terniopsis yongtaiensis*

The mitochondrial genome can be arranged in one of many chromosomes as either circular or linear molecules [32]. This study showed that the mitochondrial genome of *T. yongtaiensis* (OR818323) is a circular sequence with a length of 426,928 bp (Fig. 1), featuring a nucleotide composition of 29.04% A, 21.05% T, 21.03% C, 28.88% G, and a GC content of 42.09%. The mitochondrial genome contains a total of 31 protein-coding genes, wherein 24 represent unique core genes essential for mitochondrial function, and 7 represent variable genes indicative of evolutionary adjustments or species-specific roles. Additionally, the genome includes 18 tRNAs and 3 rRNA genes, along with one pseudogene (*sdh4*) (Table 1). Notably, *T. yongtaiensis* contains the *rrn26* (2770 bp), *rrn18* (1771 bp), and *rrn5* (111 bp) genes, consistent with the presence of three rRNA genes commonly observed in most terrestrial plants [33]. Furthermore, we compared the gene content of mitochondrial genomes across the 10 plant species of Malpighiales, revealing varied pattern of gene lost among these species (Fig. 2). Additionally, we observed considerable variation in the number and compositions of introns within plant mitochondrial genomes. Specifically, in *T. yongtaiensis*, eight of the annotated genes were found to contain type II introns, with detailed composition shown in Table S3. Notably, the introns of *nad1*, *nad2*, and *nad5* genes were found to be trans-spliced.

Repeat analysis

Simple sequence repeats (SSRs), ranging in length from one to six base pairs, are notable for their polymorphism, ease of detection through PCR, codominant inheritance, and extensive coverage across the genome [34, 35]. In this study, we identified 105 SSRs in the chloroplast genome and 130 SSRs in the mitochondrial genomes of *Terniopsis yongtaiensis* (Fig. 3). Within the mitochondrial genome, we found six types of SSRs, including 48 mono-, 19 di-, 16 tri-, 38 tetra-, 7 penta-, and 2 hexanucleotide repeat units (Table S4). However, the penta- and hexa- SSRs were not detected in the chloroplast genome of *T. yongtaiensis* (Table S5). Mononucleotide repeat units were the most abundant SSRs in both genomes, accounting for 82.86% of all SSRs in the chloroplast genome and 36.92% in the mitochondrial genome. Notably, these mononucleotide SSRs primarily consisted of A/T bases, representing 97.7% and 87.5% of the total mononucleotide SSRs in the chloroplast and mitochondrial genome, respectively. Most SSRs were found in intergenic regions, comprising

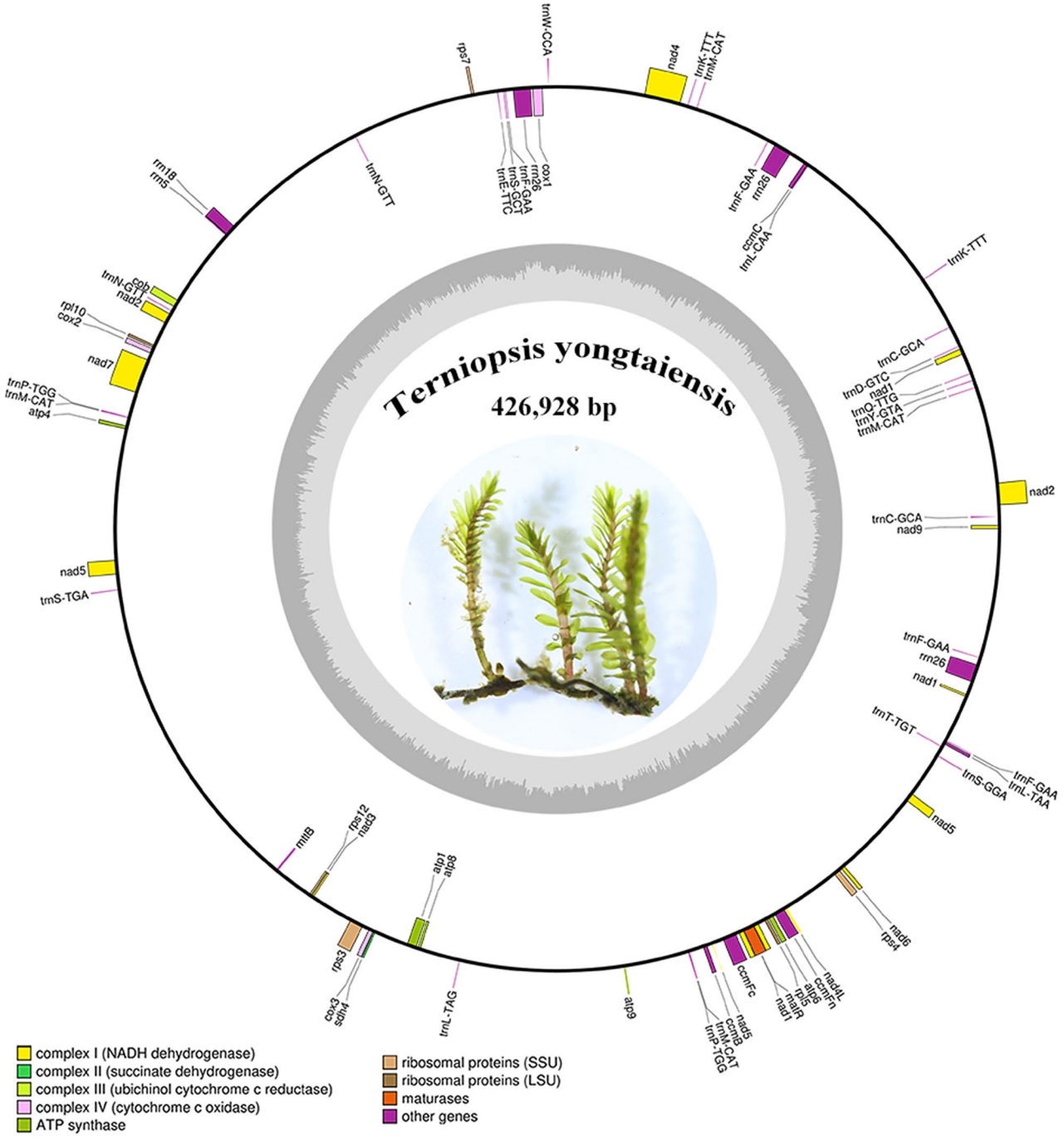

Fig. 1 The circular map of the mitochondrial genome of *Terniopsis yongtaiensis*. Genes are color-coded based on their functional groups. GC content is represented on the inner circle by the dark gray plot

70.5% and 87.7% of the totals SSRs in the chloroplast and mitochondrial genomes of *T. yongtaiensis*, respectively. Given their abundance and distinctive distribution, these SSRs have the potential to serve as valuable markers for identifying *T. yongtaiensis*.

Tandem repeats DNA sequences, which consist of units longer than 6 bp, are highly dynamic components of genomes. Our study identified a total of 14 and 22 tandem repeats in the chloroplast and mitochondrial genomes of *T. yongtaiensis*, respectively (Fig. 3). Furthermore, the length range of these tandem repeats varied between the two types of organelle genomes. In the mitochondrial genome, the length range was wider, spanning from 9 to 40 bp (Table S6), while in the chloroplast genome, it was narrower, ranging from 11 to 33 bp (Table S7). Concerning their distribution, the majority (71.4%)

Table 1 Gene compositions of the mitochondrial genome of *Terniopsis yongtaiensis*

Group of genes	Gene name
ATP synthase	*atp1 atp4 atp6 atp8 atp9*
Cytochrome *c* biogenesis	*ccmB ccmC ccmFc* ccmFn*
Ubichinol cytochrome *c* reductase	*cob*
Cytochrome *c* oxidase	*cox1 cox2 cox3*
Maturases	*matR*
Transport membrance protein	*mttB*
NADH dehydrogenase	*nad1**** nad2**** nad3 nad4*** nad4L nad5**** nad6 nad7**** nad9*
Ribosomal proteins (LSU)	*rpl10 rpl5*
Ribosomal proteins (SSU)	*rps12 rps3* rps4 rps7*
Succinate dehydrogenase	*#sdh4 sdh4*
Ribosomal RNAs	*rrn18 rrn26(3) rrn5*
Transfer RNAs	*trnC-GCA(2) trnD-GTC trnE-TTC trnF-GAA(4) trnK-TTT(2) trnL-CAA trnL-TAA* trnL-TAG trnM-CAT(4) trnN-GTT(2) trnP-TGG(2) trnQ-TTG trnS-GCT trnS-GGA trnS-TGA trnT-TGT trnW-CCA trnY-GTA*
Other	-

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

of tandem repeats in the chloroplast genome were located in intergenic regions, although some were also found within coding sequences. Conversely, all the tandem repeats identified in the mitochondrial genome were exclusively occurred in intergenic regions.

Dispersed repeats play an essential role in generating genetic diversity, and significantly contribute to the evolution of plant genomes [36]. These repeats can be classified into four types: forward repeats, reverse repeats, complement repeats, and palindromic repeats. Our study revealed that the chloroplast genome of *T. yongtaiensis* contained 28 repeats dispersed, which were classified into three out of the four types: forward repeats, reverse repeats, and palindromic repeats (Table S8). Conversely, the mitochondrial genome of *T. yongtaiensis* contained 220 repeats, all of which were either the forward or palindromic repeats (Fig. 3, Table S9). In both genomes, forward repeats were the most abundant repeats, accounts for 60.7% and 56.4% of the total repeats in the chloroplast and mitochondrial genomes, respectively. This dominance of forward repeats suggests a shared evolutionary mechanism that favors this type of repeat in these organelle genomes. Within the mitochondrial genome, the longest fragment had a length of 6,218 bp, while most repeats fell within the range of 30 to 39 bp. These repeats were predominantly located in intergenic regions, constituting 86% of the total. In contrast, the distribution of dispersed repeats in the chloroplast genome was more balanced, with repeats evenly spread across protein-coding regions, intron regions, and intergenic regions.

The identification of potential RNA editing sites in PCGs

RNA editing is a prevalent biochemical process observed across all eukaryotes, involving modifications such as nucleotide additions, deletions, or substitutions within the coding region of transcribed RNA. Within the mitochondria and chloroplasts of plants, RNA editing specifically entails the conversion of cytosines to uracils, resulting in an alteration of the genetic information encoded within the genome [37]. In current study, we employed the PmtREP tool to predict RNA editing events using a cutoff value of 0.2. Our analysis revealed a total of 22 and 267 RNA editing sites within 8 and 27 PCGs of the chloroplast and mitochondrial genomes of *Terniopsis yongtaiensis*, respectively (Fig. 4). Among the 27 PCGs of mitochondrial genomes, the *nad4*, *ccmFn*, *nad2*, *ccmB*, and *ccmC* genes contained 21 to 25 RNA editing sites, while *sdh4*, *atp1*, *atp8*, and *cob* genes only contained 1 to 2 RNA editing sites (Table S10). In comparison, all 8 genes of the chloroplast genome of *T. yongtaiensis* only contained 1 to 2 RNA editing sites (Table S11), indicating a higher prevalence of RNA editing in the mitochondrial genome compared to the chloroplast genome of *T. yongtaiensis*, albeit with less even distribution. Furthermore, when comparing the total number of editing sites in PCGs across the 10 Malpighiales plants from five families (Euphorbiaceae, Calophyllaceae, Podostemaceae, Passifloraceae and Salicaceae), the results demonstrated that *T. yongtaiensis* had the fewest RNA editing sites (Fig. 5, Table S12). In addition, the *ccm* genes exhibited significantly higher average editing densities compared to other gene types for the 10 species (Fig. S1).

Of note, all identified RNA editing events were of the C-U type, with 32.21% (62) of the editing sites located on the first base of the triplet codon, and 65.54% (175) located on the second base. However, for the *ccmB*, *nad4*, and *nad6* genes, both the first and second bases of the triplet codon were edited, resulting in the conversion of proline (CCC, CCT) to phenylalanine (TTC, TTT). Following RNA editing, 43.9% of amino acids retained their hydrophobicity, while 5.2% of hydrophobic amino acids became hydrophilic, and 50.9% of hydrophilic amino acids became hydrophobic (Table S13). RNA editing not only leads to changes in the encoded amino acids, but may also lead to the premature termination of the coding process [38]. In the *T. yongtaiensis* mitochondrial genome, this phenomenon was observed in the coding gene *ccmFc*. The predicted outcomes also indicated that the amino acids converted to leucine had the highest tendency after RNA editing, with 46.82% (125 positions) of amino acids being converted to leucine, followed by phenylalanine, accounting for 23.22% (62 positions) of all conversions.

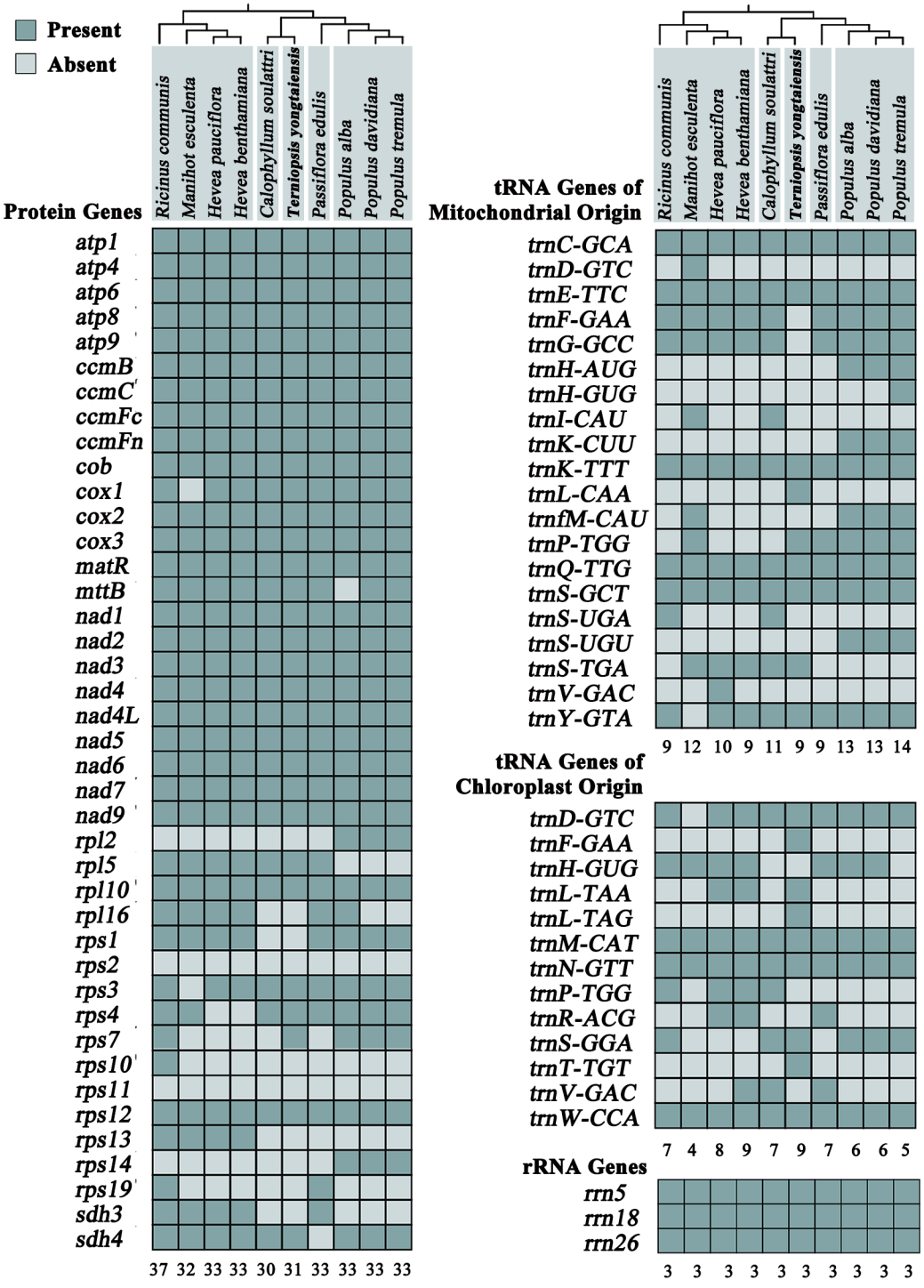

Fig. 2 Gene content in the Malpighiales plant mitochondrial genomes. Dark green boxes indicate the presence of an intact reading frame or folding structure while light gray boxes indicate the absence of an intact reading frame or folding structure. The numbers at the bottom of each gene group indicate the total number of intact genes for that species

Codon usage analysis of PCGs in *Terniopsis yongtaiensis*

Codon usage bias is the preferential or non-random use of synonymous codons, a ubiquitous phenomenon observed in bacteria, plants and animals. Different species have consistent and characteristic codon biases. Codon bias varies not only with species, family or group within kingdom, but also between the genes within an organism. Codon usage bias has evolved through mutation, natural selection, and genetic drift in various organisms [39]. The codon usage of 31 unique PCGs from

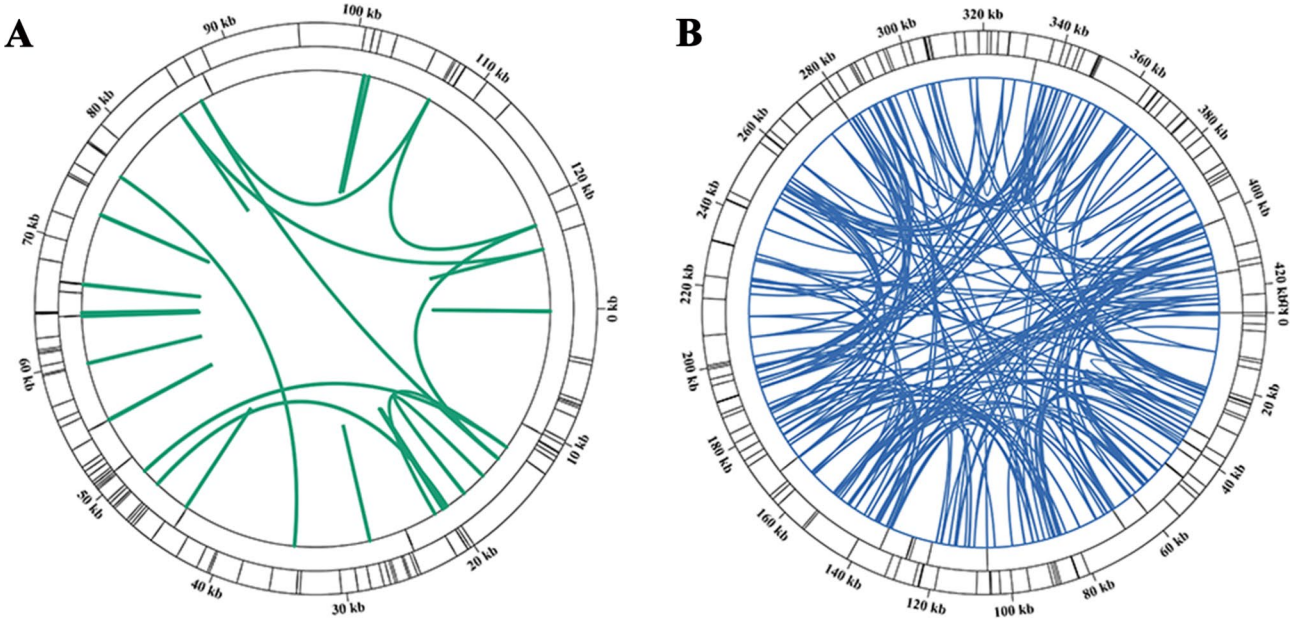

Fig. 3 The repeat analysis of the *Terniopsis yongtaiensis* organelle genomes. (**A**) The repeat sequences identified in the chloroplast genome. (**B**) The repeat sequences identified in the mitochondrial genome. The innermost circle shows the dispersed repeats connected with green (chloroplast genome) and blue (mitochondrial genome) arcs from the center going outward. The center circle shows the tandem repeats as short bars. The outermost circle shows the microsatellite sequences identified using MISA. The scale is shown on the outermost circle, with intervals of 10 kb for chloroplast genome and 20 kb for mitochondrial genome

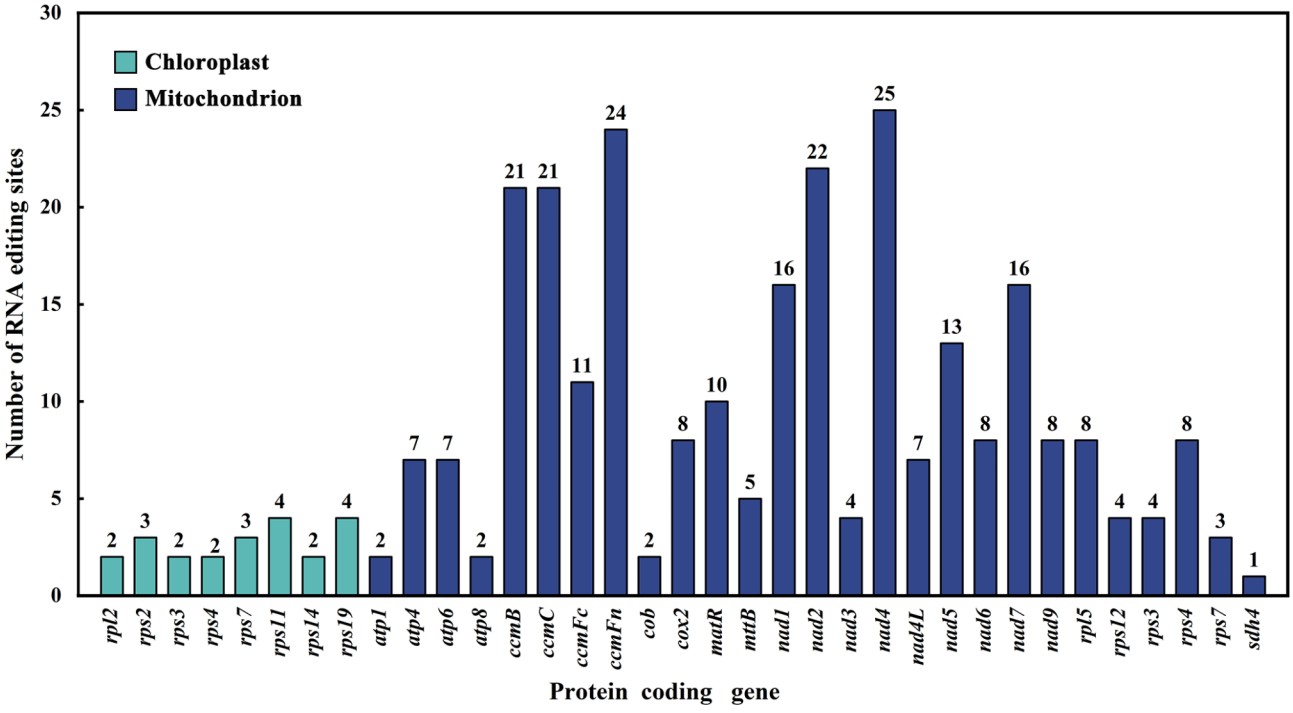

Fig. 4 The distribution of RNA editing sites across different genes of organelle genomes of *Terniopsis yongtaiensis*. The X axis shows the name of protein-coding genes, and the Y axis shows the number of predicted RNA editing sites

Terniopsis yongtaiensis was analyzed to determine their preference for synonymous codons (Fig. 6). Codons exhibiting a relative synonymous codon usage (RSCU) greater than 1 were considered to be preferentially used by amino acids. The analysis revealed that 31 codons had RSCU values greater than 1, with AUG having the highest RSCU value of 3, followed by UAA of 1.83. This indicates a high frequency of usage for methionine (Met) and the

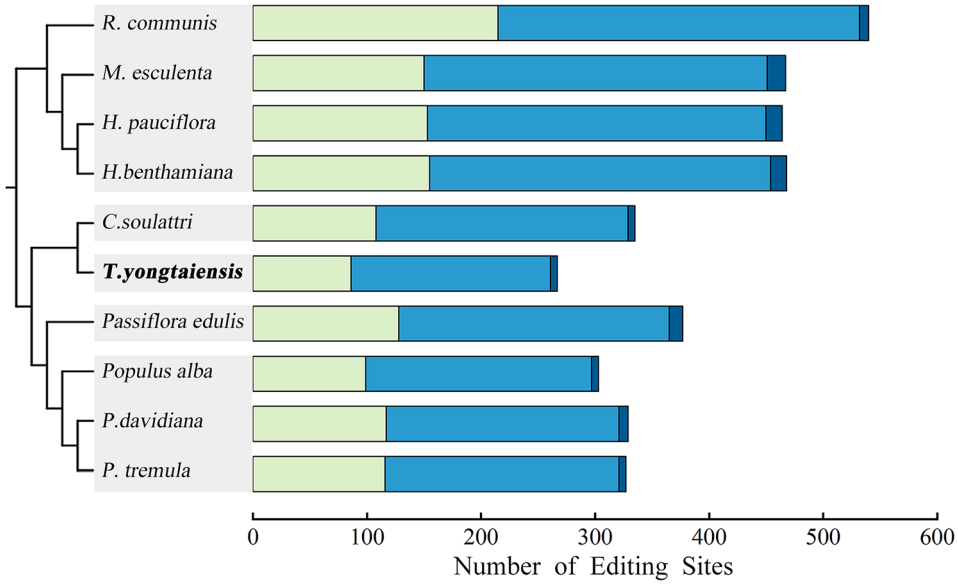

Fig. 5 Total number of editing sites in protein-coding genes across the 10 Malpighiales plants, involving 5 families (from top to bottom: Euphorbiaceae, Calophyllaceae, Podostemaceae, Passifloraceae, Salicaceae). Stacked bars showing numbers of editing sites at the first position (light green), second position (light blue), and the simultaneous occurrence in the first and second positions (dark blue) of codons, respectively

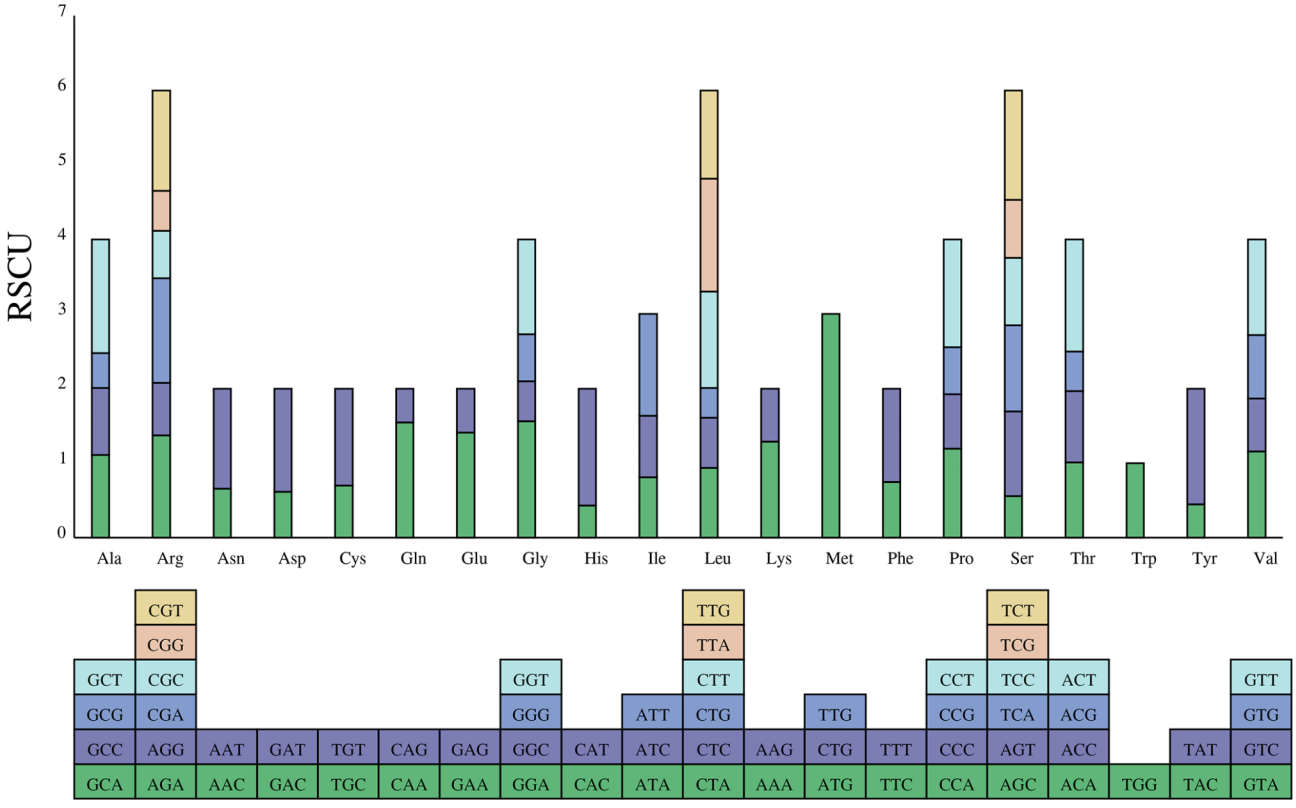

Fig. 6 Analysis of codon usage bias in *Terniopsis yongtaiensis* mitochondrial genomes. X-axis, codon families; Y-axis, the relative synonymous codon usage (RSCU) value. RSCU measures the likelihood of a specific codon being used among synonymous codons that encode the same amino acid and values greater than 1 indicate a higher frequency of usage for the codon

termination codon, respectively. Notably, 29 out of the 31 codons with RSCU values greater than 1 ended with the A/T base, accounting for 93.55% of these codons. This observation suggests a prevalent tendency for frequently used codons to terminate with the A/T base. To further investigate the effect of gene base composition on codon usage preference across all species within Malpighiales examined in this study, we calculated the effective number of codons (ENC) (Table S14). Remarkably, all genes within the mitochondrial genome of the studied species had an ENC value greater than 35, indicating that the observed codon usage bias is most likely due to natural selection or alternative factors [40]. Additionally, gene *nad9* is positioned above the standard curve, while the remaining genes are located below the standard curve line (Fig. S2). These results provide valuable insights into the evolutionary history of plants within the Malpighiales order.

Sequence similarity between the mitochondrial and chloroplast genomes

A total of 78 fragments within the mitochondrial genome showed homology with the chloroplast genome, accounting for 14.6% and 49.6% of the total lengths of the mitochondrial and chloroplast genomes, respectively (Fig. 7). These fragments varied in length from 29 to 6,771 bp, cumulatively amounting to 62,481 bp in length (Table S15). Subsequent analysis revealed that 42 out of the 78 fragments originated from the large single-copy (LSC) regions of the chloroplast genome, collectively representing approximately 72.45% of the total length of homologous fragments, amounting to 45,268 bp. Furthermore, 28 fragments were identified within the inverted repeat (IR) regions of the chloroplast genome, accounting for 20.46% of the total length, which equates to 12,786 bp. The remaining 8 fragments were found in the small single-copy (SSC) regions, contributing only 7.09% of the total length. Upon annotation of these fragments, it was found that 68 out of the 78 fragments were in the coding

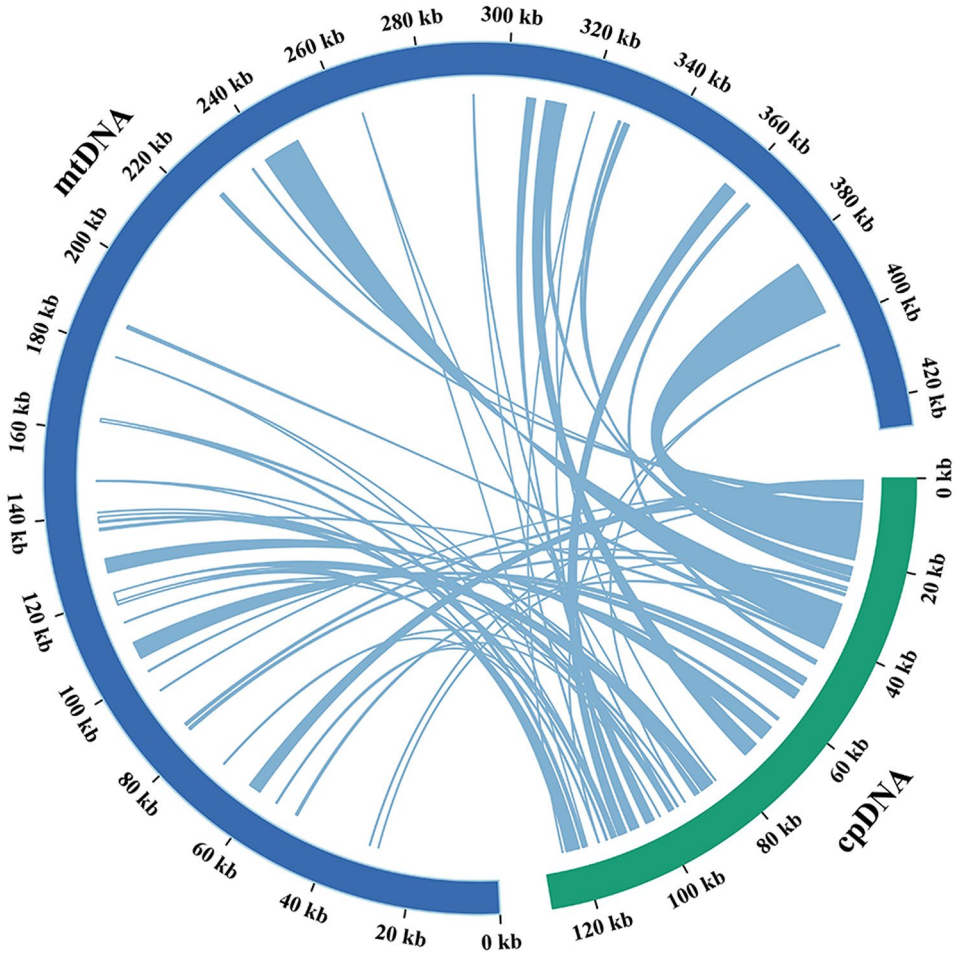

Fig. 7 Comparison of the chloroplast genome and mitochondrial genome of *Terniopsis yongtaiensis*. The blue and green outer arcs represent the mitochondrial genome (mtDNA) and chloroplast genome (cpDNA), respectively, and the inner green arcs show the homologous DNA fragments. The scale is shown on the outer arcs, with intervals of 20 kb

region of the chloroplast genome. This annotation led to the identification of 24 complete genes, including 14 PCGs and 10 tRNA genes (Table S15). In contrast, only 19 fragments were located within the coding region of the mitochondrial genome, encompassing a total of 10 complete genes (*rps7* and 9 tRNA genes) and 3 partial genes (*trnN-GTT*, *rrn18*, *rrn26*). The observation of such high similarity over substantial lengths, particularly when it pertains to intact genes, suggests gene transfer events from the chloroplast to the mitochondrial genome during evolution (Table S15). The exchange of genetic material among different genomic compartments within a cell is referred to as intracellular gene transfer (IGT) [41]. IGTs within plant cells occur continuously and dynamically, and may have great potential for applications [42].

Substitution rates of protein-coding genes

To investigate the evolutionary rate of mitochondrial genes, we calculated the nonsynonymous substitution rate (dN) and the synonymous substitution rate (dS) for the 24 shared PCGs of the 10 Malpighiales plants (Fig. 8). The dN/dS ratio provide insights into whether a specific PCG has been subjected to selective pressure during evolution. Possible outcomes include positive selection (dN/dS ratio > 1), neutral selection (dN/dS ratio = 1), and negative or purifying selection (dN/dS ratio < 1). As shown in Fig. 8, the *ccmB* gene likely experienced positive selection, given its dN/dS ratio > 1. To further explore this evolutionary signal, we performed a multiple sequence alignment of the *ccmB* gene's coding sequences across the 10 Malpighiales species (Fig. S3). This alignment revealed notable sequence differences in *Terniopsis yongtaiensis*, indicating that the *ccmB* gene may be a relatively evolved gene within its mitochondrial genome. These findings provide additional context of the dN/dS analysis, suggesting that the observed positive selection in *ccmB* could be linked to specific adaptive changes in *T. yongtaiensis*. Conversely, the *atp1*, *atp9*, and *cox3* genes exhibit low dN/dS ratios ranging from 0 to 0.3, indicating that they have been under purifying selection. Notably, the *atp9* gene demonstrated a particularly low dN/dS ratio of 0.019 with minimal variation, suggesting its role as a highly conserved gene crucial to the functionality of mitochondrial genomes (Table S16).

Phylogenetic analysis

To investigate the evolutionary origins of *Terniopsis yongtaiensis*, we conducted a taxonomic analysis utilizing the 31 conserved PCGs from its mitochondrial genome alongside those of 37 previously published plant species. Phylogenetic trees were constructed using both maximum likelihood (ML) and Bayesian methods, with *Ulva pertus*, *Chlorella heliozoae*, and *Chara vulgaris* selected as outgroups. The results obtained from both ML and Bayesian analyses demonstrated overall consistency in the classification, with the exception of Fagaceae and Rosaceae (Fig. 9), where divergent family relationships were observed. Notably, the Bayesian tree strongly supported (PP = 1.00) the phylogenetic relationships of both families, while the ML tree provided weaker support (BP = 55). Therefore, the topological structure of the Bayesian tree was used to elucidate the phylogenetic

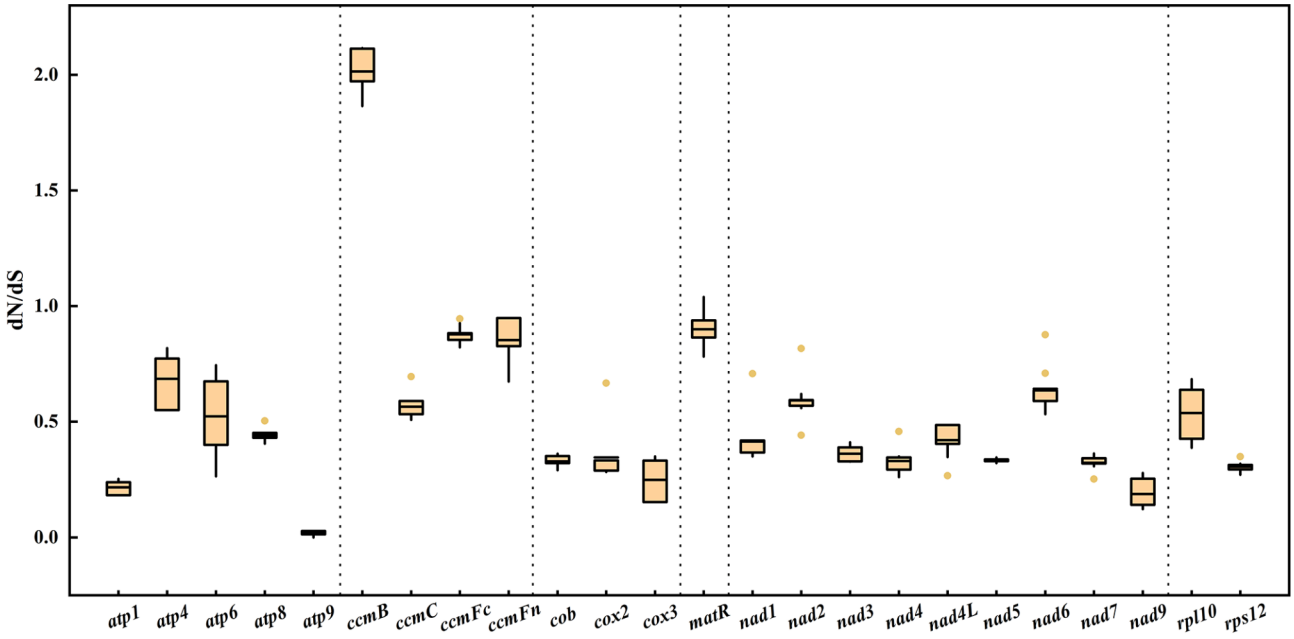

Fig. 8 The boxplots of dN/dS values of each mitochondrial gene in the 10 Malphighiales plants. The X axis shows the names of protein-coding genes, and the Y axis shows the dN/dS values

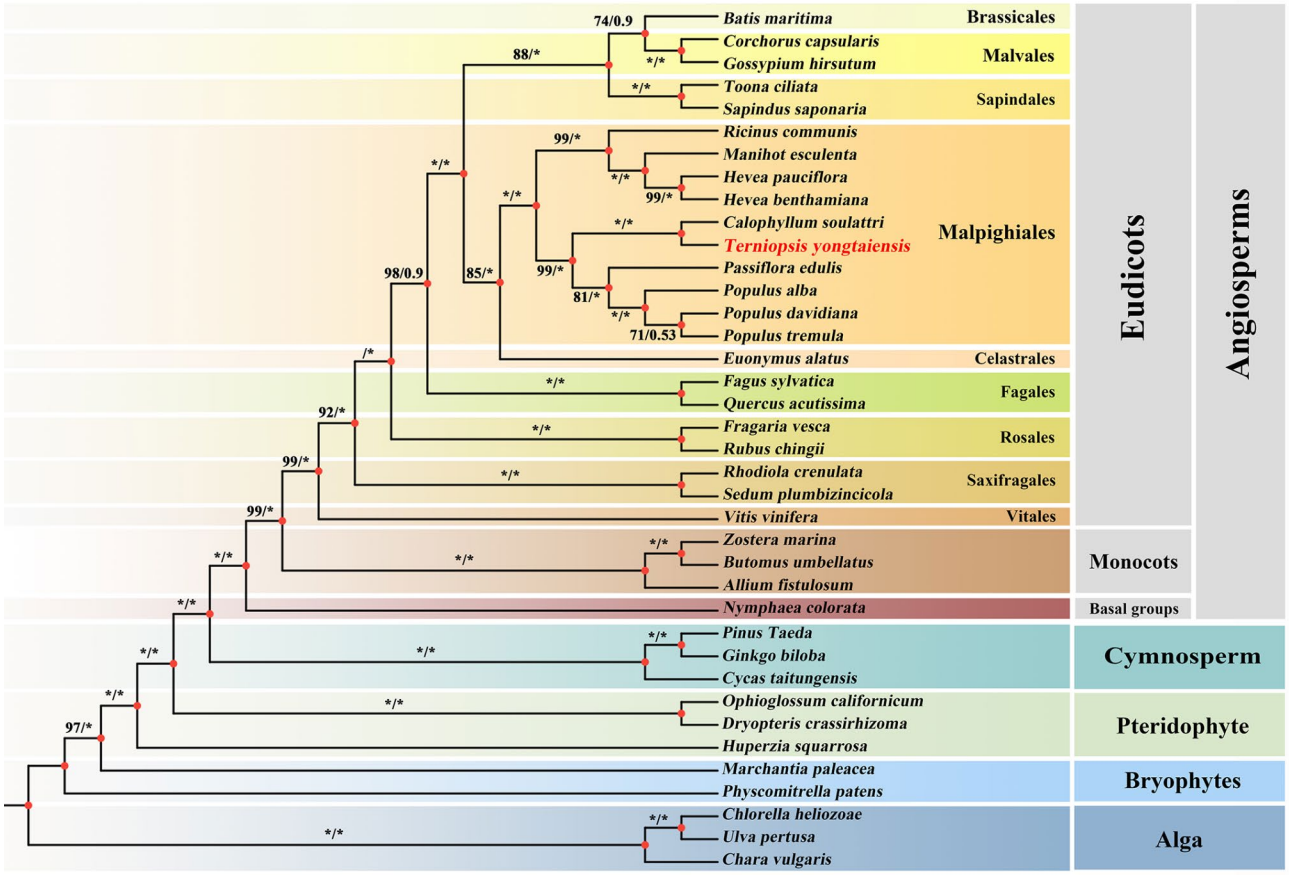

Fig. 9 Phylogenetic tree constructed using the data from 38 taxa based on 31 mitochondrial protein-coding genes. Different phyla and classes are highlighted in different colors, with the name of each phylum or class marked at the right of each highlight. Numbers above and below branches indicate RAxML (left) bootstrap probabilities (BP) and Bayesian (right) posterior probabilities (PP), respectively. Where * means BP = 100 or PP = 1.00

relationships among the 38 plant species. Based on the phylogenetic analysis constructed using the Bayesian method, we discovered that the Podostemaceae family was positioned within Malpighiales, showing a closer affinity to the Calophyllaceae family. The results are consistent with the classification proposed by the Angiosperm Phylogeny Group (APG IV). Our mitochondrial genome study provides a solid foundation for further investigations into Podostemaceae plant relatedness.

Genome size and GC content of *Terniopsis yongtaiensis* and other species

Plant organelles are influenced by various factors, with genome size and GC content being two of the most critical. To explore this further, we conducted an analysis of mitochondrial genomes from 37 plant species, comparing their sizes and GC contents to that of *Terniopsis yongtaiensis* (Fig. 10). The results revealed a significant variation in the sizes of mitochondrial genomes across the plant species investigated. This variability was evident even among closely related species, with genome size ranging from 62,477 bp (*Chlorella heliozoae*) to 1,999,602 bp (*Corchorus capsularis*). Notably, within the species investigated in Malpighiales, *T. yongtainesis* had the second smallest genome size (Table S1).

Furthermore, our investigation into GC content uncovered intriguing patterns across different plant groups. Specifically, we observed variation in GC content among the species of Alga, Bryophytes, Pteridophyte, and Gymnosperms included in our study. However, for the angiosperms investigated in our research, GC contents remained generally stable at approximately 44%, except for *T. yongtaiensis*, which exhibited a GC content of 42.09% (Table S1). This observation suggested that angiosperms have developed mechanisms to maintain a relatively consistent level of GC content level in their mitochondrial genome over time [43].

Discussion

Mitochondrial genome structure and size variations

In this study, we report for the first time the mitochondrial genome of *Terniopsis yongtaiensis*, making the initial sequencing of a mitochondrial genome within the Podostemaceae family. Mitochondria serve as critical

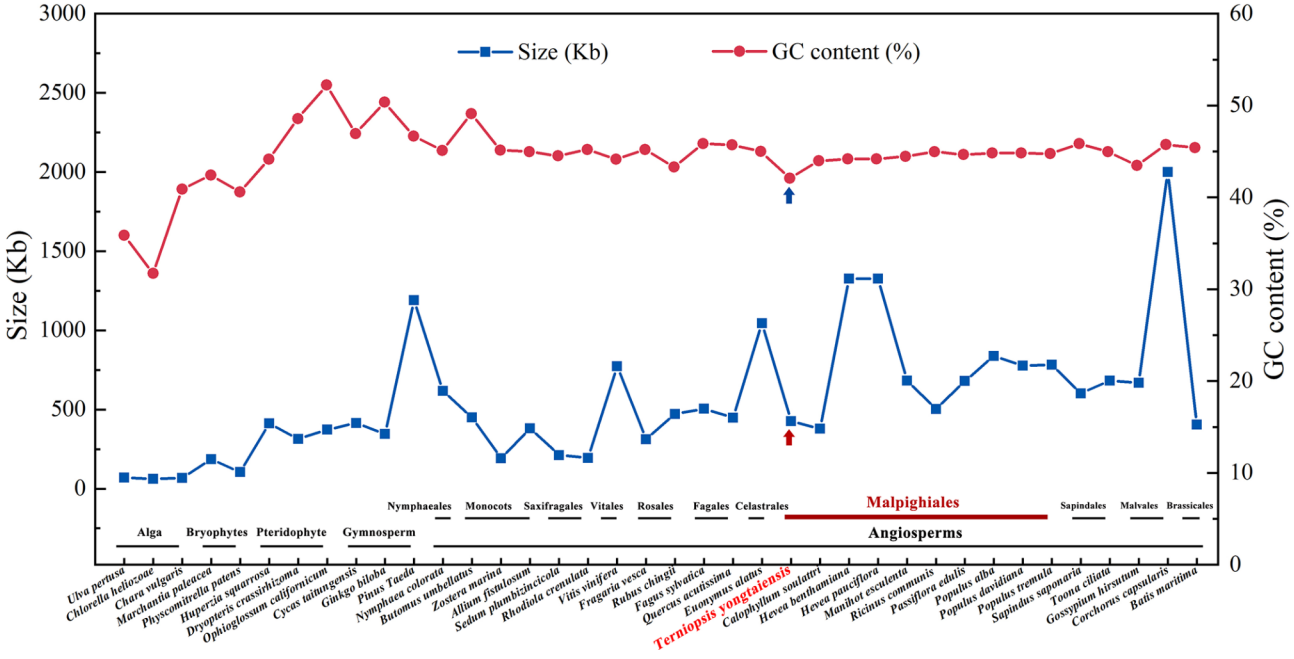

Fig. 10 Sizes and GC contents of mitochondrial genomes of 38 plants

cellular powerhouses, providing energy for various cellular activities. Notably, eukaryotic mitochondrial genome size exhibit a remarkable range, spanning over three orders of magnitude, from >11 Mb for *Silene conica* [44] to only 6 kb for apicomplexans [45]. Our results show that the circular mitochondrial genome of *T. yongtaiensis* measures 426,928 bp, which is distinctively smaller in comparison to other Malphighiales plants analysed in this study. Given the observed correlation between mitochondrial genome length variation among Malpighiales plants and alterations in the length of the non-coding regions, coupled with the preponderance of repetitive sequences within these non-coding regions, we hypothesize that a substantial reduction of the non-coding regions, primarily driven by repetitive sequences, is the primary factor influencing the compact size of *T. yongtaiensis*'s mitochondrial genome. However, to determine whether this observed mitochondrial genome size is a prevalent characteristic among Podostemaceae plants, a comprehensive analysis encompassing the mitochondrial and chloroplast genomes of additional Podostemaceae species is crucial.

Regarding the GC content of Malphighiales plant mitochondrial genomes, a relatively consistent range of 44–45% is typically observed. However, the mitochondrial genome of *T. yongtaiensis* exhibits a notably lower GC content of 42.09%, the lowest among the studied Malpighiales species. Upon comparing the GC content across various genomic components, including protein-coding genes (PCGs), tRNA, rRNA, and non-coding regions, it becomes apparent that the primary differences arise from the non-coding regions and PCGs, with a lesser contribution from the tRNA and rRNA. Specifically, the GC content of non-coding regions is influenced by factors such as repetitive sequences and intracellular gene transfer events [46]. Within the PCGs, the lower GC content observed in *T. yongtaiensis* can be primarily attributed to specific genes, including *ccm*, *matR*, and *nad*. Among these, the *nad* genes are particularly significant as they play a crucial role in mitigating oxidative stress and providing ATP for essential cellular activities, serving as strong evidence of *T. yongtaiensis*'s adaptation to aquatic environments. The unique characteristics of the mitochondrial genome of *T. yongtaiensis*, in terms of both its length and GC content, provide strong evidence for significant independent evolution within the Malpighiales order of plants.

Furthermore the mitochondrial genome of *T. yongtaiensis* exhibits another notable characteristic common to angiosperm mitochondrial genomes: the presence of extensive repetitive sequences. These repetitive sequences are known to facilitate intra- and intermolecular recombination, contributing to the observed variation in genome size among different plant species [8]. Our current study demonstrated the rich abundance of repetitive sequences, including tandem, simple, and dispersed repeats, within the mitochondrial genome of *T. yongtaiensis*. This suggests that frequent intermolecular recombination events have dynamically shaped the structure and organization of the mitochondrial genome during evolution. Of particular interest is the identification of 130 SSRs within the *T. yongtaiensis*

mitochondrial genome, which are highly valuable genetic markers widely used in assessing genetic diversity in aquatic plants due to their high abundance, variability, and codominance [47].

Phylogenetic and mitochondrial genome comparison

The utilization of organelle genomes in plant phylogeny studies has garnered increasing attention in recent years [48, 49]. In this study, the mitochondrial phylogenetic tree reconstructed for *Terniopsis yongtaiensis*, along with 37 other plant species, confirmed its placement within the Malpighiales order of angiosperm. This finding aligns with the phylogenetic position determined by Xue et al. (2020) [50] based on whole-genome analysis, highlighting the value of organelle genomes in plant phylogeny research. Specifically, our analysis revealed a close evolutionary relationship between *T. yongtaiensis* and *Calophyllum soulattri* (Fig. 9). However, while chloroplast genome data of various Podostemaceae plants are available in the NCBI database, mitochondrial genome data remain scarce. The *T. yongtaiensis* studied here providing a valuable phylogenetic framework for understanding the adaptive evolution of Podostemaceae. Nevertheless, the limited data make it challenging to reveal regular changes in the mitochondrial genome during evolution. Therefore, it is urgent need to expand our knowledge of organelle genome within the Podostemaceae family and explore their potential as novel DNA super-barcodes. Although both mitochondrial and whole-genome based phylogenetic analyses support the inclusion of Podostemaceae within the Malpighiales order, consistent with the systematic classification of APG IV, the lack of consistent morphological synapomorphies between Podostemaceae and other Malpighiales plants remains a challenge. Hence, a comprehensive approach integrating both morphological and molecular evidence is necessary to further refine the taxonomic status of Podostemaceae and Malpighiales plants.

To further explore the patterns of gene losses in *T. yongtaiensis*, we conducted a comparative analysis within Malpighiales, revealing varying degrees of phylogenetic depth in gene loss among the studied angiosperms. Notably, while three ribosomal RNA genes were found to be universally present, approximately 20 tRNA genes, most of which are frequently lost, were identified (Fig. 2). Furthermore, certain genes, such as *rps2*, *rps10* (except for *Ricinus communis*) and *rps11*, exhibited consistent loss across all the species studied within the order, suggesting ancient and deep losses. Conversely, some genes display a more restricted distribution in terms of phylogenetic occurrence, indicating relatively recent losses. Significantly, we found exclusive loss of certain genes, including *rpl16*, *rps1*, and *sdh3*, in *T. yongtaiensis* and its closely related species *C. soulattri* highlighted their potential roles in species-specific evolution and adaptation. The results further support the notion that ribosomal proteins are subjected to more rapid rates of loss than genes involved in bioenergetics [9]. Additionally, the majority of gene losses appear to have arisen from gene transfer to the nucleus, as evidenced by previous studies [9, 51–53]. These findings highlight the dynamic nature of gene loss during plant evolution and provide valuable insights into the complex mechanisms. As we looked to the future, the study of pan-genomes becomes imperative to capture the full spectrum of genetic diversity, variation, and evolution within and between species, unlocking deeper insights into adaptive mechanisms and evolutionary trajectories.

Variability of RNA editing among Malphighiales plants and genes

RNA editing plays an essential role in the post-transcriptional process in plant organellar genomes, particularly within the protein-coding regions [54]. This process characterized by adjustability in RNA editing sites, contributes significantly to genetic diversity, adaptability, and environmental fitness [55]. Our results show that RNA editing is more frequently observed in mitochondrial genomes compared to chloroplast genomes in plants, aligning with the previous report [48]. This difference is attributed to the presence of a particular PPR protein that facilitates targeted binding of effector enzymes to specific sites for editing. To maintain the stability and controllability of the RNA editing process, there is usually a correspondence between the number of PPR protein families and the RNA editing sites within the organelle genome. Substantial evidence indicates that approximately half of the PPR proteins are localized in mitochondria, while a quarter reside in chloroplasts [56].

Recent research has indicated that while RNA editing sites are generally conserved in angiosperms [57], certain species exhibit unique and specific editing sites. To explore this further, our study compared the mitochondrial genomes of *T. yongtaiensis* with nine other Malpighiales species, focusing on the quantity, distribution, and types of RNA editing sites. The results show that the number of RNA editing sites remains relatively conserved at the family level. However, *T. yongtaiensis* has the lowest number of predicted RNA editing sites (267), even lower than that reported for Salicaceae plants known for their few RNA editing sites [58]. This variation in RNA editing sites among different species may by influenced by environmental conditions and developmental stages [59]. The results suggest that the hygrophilous plants, such as *T. yongtaiensis*, *Calophyllum soulattri*, *Populus alba*, *P. davidiana*, and *P. tremula*, may have significantly fewer predicted RNA editing sites in their mitochondrial genomes compared to other species within Malpighiales.

The distribution of RNA editing sites in mitochondrial genes appears to be uneven compared to chloroplast genes. Among the 27 mitochondrial genes examined, three cytochrome *c* biogenesis genes (*ccmFn*, *ccmB*, *ccmC*) and two NADH dehydrogenase genes (*nad4*, *nad2*) exhibited the highest number and density of RNA editing sites. This observation corroborate findings from previous study [60] and exhibits a consistent pattern across all ten Malpighiales species studied herein. Moreover, this phenomenon aligns with data from mitochondrial genomes across other plant families, such as Poaceae [61] and Cucurbitaceae [62], suggesting a prevalent characteristic amongst angiosperm mitochondrial genomes.

Notably, the dN/dS analysis of the mitochondrial genomes in the ten Malpighiales plants shows that the PCGs of the *T. yongtaiensis* mitochondrial genome are predominantly under purifying selection. An exception to this trend is observed in the *ccmB* gene, which exhibited a dN/dS ratio > 1, indicative of the potential influence of positive selection on this gene throughout its evolutionary history. The *ccmB* gene encodes an inner membrane protein that plays a role in heme delivery to the matrix for cytochrome *c* maturation [63]. The identification of high dN/dS ratios in specific genes holds particular significance when studying evolution within the Podostemaceae family.

Considering the significant RNA editing density in the *ccmB*, *ccmFc*, and *ccmFn* genes, which are involved in heme transport and the synthesis of *c*-type cytochromes, it is reasonable to deduce that these genes collectively enhance the stress resistance of *T. yongtaiensis*. This inference is further supported by the adaptive evolution analysis, which suggests that most *ccm* and *nad* genes have undergone positive selection during the adaptation of *T. yongtaiensis* to aquatic environments. The identification of high dN/dS ratios in specific genes, particularly in the *ccmB* gene, underscors their importance in the evolution adaptation of *T. yongtaiensis* and highlights the potential role of positive selection in shaping the mitochondrial genome's functionality under aquatic conditions.

On the other hand, *nad* and *ndh* are two essential coenzymes in plants that participate in respiratory and photosynthetic processes through oxidation-reduction reactions. They play crucial roles in mitochondria and chloroplasts, alleviating oxidative stress and providing energy [64]. In this study, it was found that the *nad* genes in the mitochondrial genome of *T. yongtaiensis* were subject to strong environmental selection pressure. Xue et al. (2020) [50] also illuminated that most of the expanded genes in the *Cladopus chinensis* genome are involved in plant energy metabolism, particularly in oxidative phosphorylation. Therefore, we hypothesize that genes related to energy metabolism in Podostemaceae have mostly undergone positive selection during evolution, enhancing their survival and growth in hypoxic and low-light aquatic environments by improving the efficiency of oxidative phosphorylation.

Furthermore, homology analysis revealed 3 RNA editing sites in the *rps7* gene in both mitochondrial and chloroplast genomes of *T. yongtaiensis*. Further analysis confirmed that the mitochondrial *rps7* gene had been entirely transferred from the chloroplast genome through intracellular gene transfer process. Additionally, RNA editing was observed to impact the hydrophobicity of amino acids, consistent with previous study [58]. Approximately 50% of hydrophilic amino acids within the *T. yongtaiensis* mitochondrial genome were converted to hydrophobic amino acids due to RNA editing, indicating a significant impact of RNA editing on protein functional properties in plant organellar genomes.

Gene transfer between mitochondrial and chloroplast genomes

Recent researches have drawn attention to the occurrence of DNA transfers in plants. In addition to transfers from organelles to the nucleus, documented evidences also indicate transfers from chloroplast to mitochondria [65, 66]. This mechanism contributes to the expansion of the mitochondrial genome [6]. Chloroplast-derived mitochondrial genes exhibit various characteristics, including the presence of pseudogenes among protein-coding sequences [67], and nonfunctional rRNA sequences [68]. Our study identified 78 mitochondrial genome fragments, totaling 62,481 bp, which exhibited homology to the chloroplast genome of *Terniopsis yongtaiensis*. Within these fragments, we detected 13 genes, suggesting the potential for intracellular gene transfer between organelle genomes. Specifically, 10 out of the 13 identified genes were tRNA genes, including the *trnM-CAT*, *trnN-GTT*, and *trnW-CCA* homologous genes shared by 10 species of Malpighiales (Fig. 2). This finding aligns with previous research indicating the frequent transfer of tRNA genes from the chloroplast to the mitochondria genome in angiosperms, presumably to maintain essential functions [60]. Furthermore, Yue et al. (2012) [69] found through functional annotation of homologous gene families in the genome of *Physcomitrella patens* that these genes are involved in critical biological processes such as xylem formation, hormone synthesis, and nitrogen cycling. This suggests that gene transfer events play a crucial role in plants' adaptation to their environments. Looking ahead, a comprehensive analysis of the nuclear genome of *T. yongtaiensis* holds promise. Such an investigation could shed light on intracellular gene transfer events between the nuclear and organelle genomes, potentially providing insights into the evolutionary

trajectory of Podostemaceae plants as they transitioned from terrestrial to amphibious lifestyles. Unraveling the functional activity of these migrated genes may hold the key to understanding this fascinating transformation.

Conclusion

Podostemaceae, known for their extraordinary physiological adaptations to their habitats, had long been a focus of botanical research. In this study, we presented a comprehensive analysis of the mitochondrial genome of *Terniopsis yongtaiensis*, representing the first instance of mitochondrial genome sequencing within the Podostemaceae family. The circular mitochondrial genome of *T. yongtaiensis* was 426,928 bp in length and contained 31 PCGs, 18 tRNAs, and 3 rRNA genes. We subsequently analyzed the repeat sequences, RNA editing processes, and codon usage bias of the mitochondrial genome of *T. yongtaiensis*. Our results showed substantial variability in mitochondrial genome size, even among species within the same order. Although the GC content had remained relatively conserved throughout the evolutionary process, it was notably observed to be the lowest for *T. yongtaiensis* among the ten species that had been studied in Malpighiales to date, potentially influenced by recent DNA transfer from the plastome. The results of dN/dS analysis, based on coding substitutions, indicated that the majority of coding genes, excluding *ccmB*, had undergone negative selection, indicating the evolutionary conservation of mitochondrial genes. Moreover, we identified 13 homologous gene-containing regions between the mitochondrial and chloroplast genomes of *T. yongtaiensis*, suggesting gene transfer events between these organellar genomes. This study not only provided valuable insights into the genetic variation and systematic evolution of plants of Malpighiales, but also established a foundation for future research in this field.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10765-6.

Supplementary Material 1

Author contributions

MZ and BHC: design; validation; resources; database gathering; writing; preparation; analysis; finance acquisition, and editing of initial drafts. XHZ and YLH: methods. XHZ, ZXC, YLH: all types of software. All authors contributed to the article and approved the submitted version.

Funding

This work was financially supported by Special Project of Orchid Survey of National Forestry and Grassland Administration (contract no. 2020070705), the National Special Fund for Chinese medicine resources Research in the Public Interest of China (Grant No.2019-39), the National Natural Science Foundation of China (NSFC) (#32470215), the Natural Science Foundation of Fujian Province (2020J05037 to MZ), the Foundation of Fujian Educational Committee (JAT190089 to MZ).

Data availability

The mitochondrial genome data of Terniopsis yongtaiensis has been uploaded to the NCBI database, with accession number: OR818323.

Declarations

Ethics approval and consent to participate

The necessary permissions for collecting *Terniopsis yongtaiensis* has been obtained.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]College of Life Sciences, Fujian Normal University, Fuzhou 350117, China
[2]Fujian Key Laboratory of Special Marine Bioresource Sustainable Utilization, Southern Institute of Oceanography, College of Life Sciences, The Public Service Platform for Industrialization Development Technology of Marine Biological Medicine and Products of the State Oceanic Administration, Fujian Normal University, Fuzhou 350117, China

Received: 19 April 2024 / Accepted: 3 September 2024
Published online: 12 September 2024

References

1. Tǎng HT, Kato M. Culture of river-weed *Terniopsis chanthaburiensis* (Podostemaceae). Aquat Bot. 2020;166:103–255. https://doi.org/10.1016/j.aquabot.2020.103255.
2. Fujinami R, Imaichi R. Developmental anatomy of *Terniopsis malayana* (Podostemaceae, subfamily Tristichoideae), with implications for body plan evolution. J Plant Res. 2009;122:551–8. https://doi.org/10.1007/s10265-009-0243-7.
3. Rutishauser R. Evolution of unusual morphologies in Lentibulariaceae (bladderworts and allies) and Podostemaceae (river-weeds): a pictorial report at the interface of developmental biology and morphological diversification. Ann Bot. 2016;117:811–32. https://doi.org/10.1093/aob/mcv172.
4. Taiz L, Zeiger E, Møller IM. Murphy, AS. Plant physiology and development. Sunderland, MA: Sinauer Associates; 2015.
5. Møller IM, Rasmusson AG, Aken OV. Plant mitochondria – past, present and future. Plant J. 2021;108:912–59. https://doi.org/10.1111/tpj.15495.
6. Mower JP, Sloan DB, Alverson AJ. Plant mitochondrial genome diversity: the genomics revolution. In: Wendel JF, Greilhuber J, Dolezel J, Leitch AR, editors. Plant genome diversity: plant genomes, their residents, and their evolutionary dynamics. Vienna: Springer; 2012. pp. 123–44.
7. Fox TD, Leaver CJ. The *Zea mays* mitochondrial gene coding cytochrome oxidase subunit II has an intervening sequence and does not contain TGA codons. Cell. 1981;26:315–23. https://doi.org/10.1016/0092-8674(81)9020-2.
8. 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.
9. Adams KL, Qiu YL, Stoutemyer M, Palmer JD. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. PNAS. 2002;99(15):9905–12. https://doi.org/10.1073/pnas.042694899.
10. Zubaer A, Wai A, Hausner G. The mitochondrial genome of *Endoconidiophora resinifera* is intron rich. Sci Rep. 2018;8:17591. https://doi.org/10.1038/s41598-018-35926-y.
11. Bedoya AM, Ruhfel BR, Philbrick CT, Madriñán S, Bove CP, Mesterházy A, Olmstead RG. Plastid genomes of five species of riverweeds (Podostemaceae): structural organization and comparative analysis in Malpighiales. Front Plant Sci. 2019;10:1035. https://doi.org/10.3389/fpls.2019.01035.
12. Jin DM, Jin JJ, Yi TS. Plastome structural conservation and evolution in the clusioid clade of Malpighiales. Sci Rep. 2020;10(1):9091. https://doi.org/10.1038/s41598-020-66024-7.
13. Zhang M, Zhang XH, Ge CL, Chen BH. *Terniopsis Yongtaiensis* (Podostemaceae), a new species from South East China based on morphological

and genomic data. PhytoKeys. 2022;194:105–22. https://doi.org/10.3897/phytokeys.194.83080.
14. Chen BH, Zhang M, Zhao K, Zhang XH, Ge CL. *Polypleurum chinense* (Podostemaceae), a new species from Fujian, China, based on morphological and genomic evidence. PhytoKeys. 2022;199:167–86. https://doi.org/10.3897/phytokeys.199.85679.
15. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–100. https://doi.org/10.1093/bioinformatics/bty191.
16. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27(5):722–36. https://doi.org/10.1101/gr.215087.116. http://www.genome.org/cgi/doi/.
17. Langdon WB. Which is faster: bowtie2GP bowtie > bowtie2 > BWA. Proceedings of the 15th annual conference companion on Genetic and evolutionary computation. 2013: 1741–1742. https://doi.org/10.1145/2464576.2480772
18. Wick RR, Judd LM, Gorrie CL, Holt KF. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13(6):e1005595. https://doi.org/10.1371/journal.pcbi.1005595.
19. Beier S, Thiel T, Münch T, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33(16):2583–5. https://doi.org/10.1093/bioinformatics/btx198.
20. Benson G. Tandem repeats finder: a program to analyze DNA sequences. NAR. 1999;27(2):573–80. https://doi.org/10.1093/nar/27.2.573.
21. Chen Y, Ye WC, Zhang YD, Xu YS. High speed BLASTN: an accelerated MegaBLAST search tool. NAR. 2015;43(16):7762–8. https://doi.org/10.1093/nar/gkv784.
22. Chen CJ, Chen H, Zhang Y, Thomas HR, Frank MH, Frank MH, He YH, Xia R. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 2020;13(8):1194–202. https://doi.org/10.1016/j.molp.2020.06.009.
23. Song Y, Du XR, Li AX, Fan AM, He LJ, Sun Z, Niu YB, Qiao YG. Assembly and analysis of the complete mitochondrial genome of *Forsythia suspensa*. (Thunb) Vahl BMC Genomics. 2023;1:708–708. https://doi.org/10.1186/s12864-023-09821-4.
24. 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.
25. Zhang D, Gao FL, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT. Mol Ecol Resour. 2020;20(1):348–55. https://doi.org/10.1111/1755-0998.13096. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies.
26. Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler AV, Jermiin L. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9. https://doi.org/10.1038/nmeth.4285.
27. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74. https://doi.org/10.1093/molbev/msu300.
28. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 2013;30:1188–95. https://doi.org/10.1093/molbev/mst024.
29. Ronquist F, Teslenko M, Van Der Mark P, Aryes DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61(3):539–42. https://doi.org/10.1093/sysbio/sys029.
30. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. NAR. 2021; 49(W1): W293–W296. https://doi.org/10.1093/nar/gkab301
31. Xu B, Yang ZH. PAMLX: a graphical user interface for PAML. Mol Biol Evol. 2013;30(12):2723–4. https://doi.org/10.1093/molbev/mst179.
32. Mower JP. Variation in protein gene and intron content among land plant mitogenomes. Mitochondrion. 2020;53:203–13. https://doi.org/10.1016/j.mito.2020.06.002.
33. Archibald JM. Origin of eukaryotic cells: 40 years on. Symniosis. 2011;54:69–86. https://doi.org/10.1007/s13199-011-0129-z.
34. Powell W, Machray GG, Provan J. Polymorphism revealed by simple sequence repeats. Trends Plant Sci. 1996;1:215–22. https://doi.org/10.1016/1360-1385(96)86898-1.
35. Qiu LJ, Yang C, Tian B, Yang JB, Liu AZ. Exploiting EST databases for the development and characterization of EST-SSR markers in castor bean (*Ricinus communis* L). BMC Plant Biol. 2010;10:278. https://doi.org/10.1186/1471-2229-10-278.
36. Smyth DR. Dispersed repeats in plant genomes. Chromosoma. 1991;100(6):355–9. https://doi.org/10.1007/BF00337513.
37. Lukeš J, Kaur B, Speijer D. RNA editing in mitochondria and plastids: weird and widespread. Trends Genet. 2021;37(2):99–102. https://doi.org/10.1016/j.tig.2020.10.004.
38. Ichinose M, Sugita M. RNA editing and its molecular mechanism in plant organelles. Genes. 2017;8(1):5. https://doi.org/10.3390/genes8010005.
39. Parvathy ST, Udayasuriyan V, Bhadana V. Codon usage bias. Mol Biol Rep. 2022;49:539–65. https://doi.org/10.1007/s11033-021-06749-4.
40. Wright F. The 'effective number of codons' used in a gene. Gene. 1990; 87(1): 23–29. https://doi.org/10.1016/0378-1119(90)90491-9
41. Choi IS, Schwarz EN, Ruhlman TA, Khiyami MA, Sabir JSM, Hajarah NH, Sabir MJ, Rabah SO, Jansen RK. Fluctuations in Fabaceae mitochondrial genome size and content are both ancient and recent. BMC Plant Biol. 2019;19:448. https://doi.org/10.1186/s12870-019-2064-8.
42. Wang J, Kan SL, Liao XZ, Zhou JW, Tembrock LR, Daniell H, Jin SX, Wu ZQ. Plant organellar genomes: much done, much more to do. Trends Plant Sci. 2024. https://doi.org/10.1016/j.tplants.2023.12.014.
43. Cheng Y, He XX, Priyadarshani SVGN, Wang Y, Ye L, Shi C, Ye KZ, Zhou Q, Luo ZQ, 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;22:167. https://doi.org/10.1186/s12864-021-07490-9.
44. 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:e1001241. https://doi.org/10.1371/journal.pbio.1001241.
45. Hikosaka K, Watanabe Y, Tsuji N, Kita K, Kishine H, Arisue N, Palacpac NM, Kawazu S, Sawai H, Horii T, Igarashi I, Tanabe K. Divergence of the mitochondrial genome structure in the apicomplexan parasites, Babesia and Theileria. Mol Biol Evol. 2010;27:1107–16. https://doi.org/10.1093/molbev/msp320.
46. Sloan DB, Wu ZQ. 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.
47. Hu SQ, Li GJ, Yang JJ, Hou HW. Aquatic plant genomics: advances, applications, and prospects. Int J Genomics. 2017;27:6347874. https://doi.org/10.1155/2017/6347874.
48. Ni Y, Li JL, Chen HM, Yue JW, Chen PH, Liu C. Comparative analysis of the chloroplast and mitochondrial genomes of *Saposhnikovia divaricata* revealed the possible transfer of plastome repeat regions into the mitogenome. BMC Genomics. 2022;23:570. https://doi.org/10.1186/s12864-022-08821-0.
49. Li J, Tang H, Luo H, Tang J, Zhong N, Xiao LZ. Complete mitochondrial genome assembly and comparison of Camellia sinensis var. Assamica Cv. Duntsa. Front Plant Sci. 2023;14:1117002. https://doi.org/10.3389/fpls.2023.1117002.
50. Xue T, Zheng XH, Chen D, Liang LM, Chen N, Huang Z, Fan WF, Chen JN, Cen W, Chen S, Zhu JM, Chen BH, Zhang XT, Chen YQ. A high-quality genome provides insights into the new taxonomic status and genomic characteristics of *Cladopus chinensis* (Podostemaceae). Hortic Res. 2020;7(1):46. https://doi.org/10.1038/s41438-020-0269-5.
51. Wischmann C, Schuster W. Transfer of *rps10* from the mitochondrion to the nucleus in *Arabidopsis thaliana*: evidence for RNA-mediated transfer and exon shuffling at the integration site. FEBS lett. 1995;374:152–6. https://doi.org/10.1016/0014-5793(95)01100-S.
52. Sánchez H, Fester T, Kloska S, Schröder W, Schuster W. Transfer of *rps19* to the nucleus involves the gain of an RNP-binding motif which may functionally replace RPS13 in *Arabidopsis* mitochondria. EMBO J. 1996;15:2138–49. https://doi.org/10.1002/j.1460-2075.1996.tb00567.x.
53. Adams KL, Ong HC, Palmer JD. Mitochondrial gene transfer in pieces: fission of the ribosomal protein gene *rpl2* and partial or complete gene transfer to the nucleus. Mol Biol Evol. 2001;18:2289–97. https://doi.org/10.1093/oxfordjournals.molbev.a003775.
54. Steinhauser S, Beckert S, Capesius I, Malek O, Knoop V. Plant mitochondrial RNA editing. J Mol Evol. 1999;48:303–12. https://doi.org/10.1007/pl00006473.
55. Rosenthal JJC. The emerging role of RNA editing in plasticity. J Exp Biol. 2015;218:1812–21. https://doi.org/10.1242/jeb.119065.
56. Takenaka M, Zehrmann A, Brennicke A, Graichen K. Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. PLoS ONE. 2013;8:e65343. https://doi.org/10.1371/journal.pone.0065343.
57. Brenner WG, Mader M, Müller NA, Hoenicka H, Schroeder H, Zorn I, Fladung M, Kerste B. High level of conservation of mitochondrial RNA editing sites among four *Populus* species. G3-Genes Genom. Genet. 2019;9:709–17. https://doi.org/10.1534/g3.118.200763.

58. Edera AA, Gandini CL, Sanchez-Puerta MV. Towards a comprehensive picture of C-to-U RNA editing sites in angiosperm mitochondria. Plant Mol Biol. 2018;97:215–31. https://doi.org/10.1007/s11103-018-0734-9.
59. Okuda K, Hammani K, Tanz SK, Peng L, Fukao Y, Myouga F, Motohashi R, Shinozaki K, Small I, Shikanai T. The pentatricopeptide repeat protein OTP82 is required for RNA editing of plastid *ndhB* and *ndhG* transcripts. Plant J. 2010;61:339–49. https://doi.org/10.1111/j.1365-313X.2009.04059.x.
60. Bi CW, Paterson AH, Wang XL, Xu YQ, Wu DY, Qu YS, Jiang A, Ye QL, Ye N. Corrigendum to Analysis of the complete mitochondrial genome sequence of the diploid cotton *Gossypium raimondii* by comparative genomics approaches. BioMed. Res. Int. 2019; 2019: 9691253. https://doi.org/10.1155/2019/9691253
61. Zhou DG, Liu Y, Yao JZ, Yin Z, Wang XW, Xu LP, Que YX, Mo P, Liu XL. Characterization and phylogenetic analyses of the complete mitochondrial genome of sugarcane (*Saccharum* Spp. Hybrids) line A1. Diversity. 2022;14:333. https://doi.org/10.3390/d14050333.
62. Niu Y, Zhang T, Chen MX, Chen GJ, Liu ZH, Yu RB, Han X, Chen KH, Huang AZ, Chen CM, Yang Y. Analysis of the complete mitochondrial genome of the bitter gourd (*Momordica charantia*). Plants. 2023;12:1686. https://doi.org/10.3390/plants12081686.
63. Thöny-Meyer L, Fischer F, Künzler P, Ritz D, Hennecke H. Escherichia coli genes required for cytochrome c maturation. J Bacteriol. 1995;177(15):4321–6. https://doi.org/10.1128/jb.177.15.4321-4326.1995.
64. Kory N, uit de Bos J, van der Rijt S, Jankovic N, Güra M, Arp N, Pena IA, Prakash G, Chan SH, Kunchok T, Lewis CA, Sabatini ADM. MCART1/SLC25A51 is required for mitochondrial NAD transport. Sci. Adv. 2020;6(43):eabe5310. https://doi.org/10.1126/sciadv.abe5310.
65. Kubo T, Nishizawa S, Sugawara A, Itchoda N, Estiati A, Mikami T. The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for $tRNA^{Cys}$ (GCA). Nucleic Acids Res. 2000;28:2571–6. https://doi.org/10.1093/nar/28.13.2571.
66. Cui HN, Ding Z, Zhu QL, Wu Y, Qiu BY, Gao P. Comparative analysis of nuclear, chloroplast, and mitochondrial genomes of watermelon and melon provides evidence of gene transfer. Sci Rep. 2021;11:1595. https://doi.org/10.1038/s41598-020-80149-9.
67. Cummings MP, Nugent JM, Olmstead RG, Palmer JD. Phylogenetic analysis reveals five independent transfers of the chloroplast gene *rbcL* to the mitochondrial genome in angiosperms. Curr Genet. 2003;43:131–8. https://doi.org/10.1007/s00294-003-0378-3.
68. 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. https://doi.org/10.1007/s00438-002-0767-1.
69. Yue JP, Hu XY, Sun H, Yang YP, Huang JL. Widespread impact of horizontal gene transfer on plant colonization of land. Nat Commun. 2012;3:1152. https://doi.org/10.1038/ncomms2148.

Publisher's note