

Comparative genomic and phylogenetic analyses of mitochondrial genomes of *Graptopetalum paraguayense* (N. E. Br.) Walth.1938

Xue Zhou

NanJing Normal University TaiZhou

ChuQi Lin

NanJing Normal University TaiZhou

ZhiRui Li

NanJing Normal University TaiZhou

Lin Zhou

NanJing Normal University TaiZhou

Xi Du

20101012@nnutc.edu.cn

NanJing Normal University TaiZhou

Article

Keywords: Graptopetalum paraguayense, Mitochondrial genome, Phylogeny

Posted Date: October 16th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-7592618/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on January 3rd, 2026.
See the published version at <https://doi.org/10.1038/s41598-025-34236-4>.

Abstract

Graptopetalum paraguayense is a perennial succulent plant with ornamental value, functional value and medicinal activity. This study presents the first complete mitochondrial genome of *G. paraguayense*. We assembled and annotated the complete mitochondrial genome of *G. paraguayense*. we characterized a 242,059bp circular mitogenome with 43.65% GC content, harboring 50 functional genes (31 PCGs, 13 tRNAs, 3 rRNAs) and 3 pseudogenes. Identification of 599 RNA editing sites, predominantly altering amino acid hydrophobicity (47.41% hydrophilic-to-hydrophobic conversions). Codon usage bias analysis revealing preferential use of GCU (Ala), CGA (Arg), and UUA (Leu), with UAA stop codon exhibiting highest RSCU (1.9355). A total of 122 repetitive sequences were identified, including 59 Simple sequence repeat, 1 tandem repeat, and 62 dispersed repeat. Ka/Ks analysis indicating positive selection on *ccmB* and *nad7* genes. *G. paraguayense* shared 57.28% sequence similarity with *Sedum plumbizincicola*. Evidence of chloroplast-to-mitochondrial DNA transfer involving *psaC*, *ndhE*, *ndhG*, *ndhI*, *ndhA*, *ndhH* genes. Eleven divergent hotspot regions were identified by comparative analyses, were *atp9*, *atp8*, *rpl5*, *cox2*, *ccmFn*, *rps7*, *ccmC*, *ccmFc*, *mttB*, *nad6* and *rps13*. Based on phylogenetic tree analysis, *G. paraguayense* is highly related to *Sedum plumbizincicola*. These results provide foundational insights into mitogenome evolution in Crassulaceae, highlighting adaptive genetic mechanisms and interorganellar gene transfer in succulent plants.

1. Introduction

Graptopetalum paraguayense (N. E. Br.) Walth. 1938, a perennial succulent plant in the *Graptopetalum* genus of the *Crassulaceae* family, is commonly known as Ghost Plant or Mother of Pearl. Native to arid regions such as Mexico, *G. paraguayense* features thick, flat leaves that are slightly pointed at the top, with an overall lotus-like shape that makes it highly ornamental. The leaf color varies depending on sunlight and environmental conditions: it appears blue-gray in the shade and pink in the sun. In addition to its aesthetic appeal, *G. paraguayense* is also regarded as a functional food and is used for its antihypertensive and hepatoprotective properties. It has been reported to exhibit a range of other bioactive effects, including anticancer, antibacterial, antiasthma, skin-whitening, and anti-Alzheimer's activities^[1]. Moreover, *G. paraguayense* extract (GP extract) has been shown to inhibit liver cancer cell growth and significantly prolong the lifespan of *C. elegans*^[2].

The *Crassulaceae* family comprises 34 genera and over 1,500 species, including *Aeonium*, and *Sedum*. The subfamily *Crassulaceae* is divided into three subfamilies: *Crassuloideae*, *Kalanchoideae*, and *Sempervivoideae*. Within *Sempervivoideae*, there are five clades: the *Telephium*, *Sempervivu*, *Aeonium*, *Leucosedum*, *Acre* clades^[3]. The focus of this study is *G. paraguayense*, which belongs to the *Graptopetalum* genus of the *Acre* clade. Previous studies have utilized various molecular markers, such as nuclear ETS, internal transcribed spacer (ITS), and plastid DNA sequences (*rpl16*, *trnL-F*), to explore phylogenetic relationships within closely related genera, including *Cyperus*^[4].

Mitochondria are semi-autonomous organelles found in most eukaryotic cells. Plant mitochondria typically range from 1 to 3 μm in length and about 0.5 μm in diameter, with each plant cell containing several hundred mitochondria^[5]. The study of plant mitochondria began in earnest around the 1950s, following the first successful isolations of mitochondria from animal and plant tissues^[6]. Plant mitogenomes are characterized by high rearrangement, low mutation rates, and large size^[7]. The structural features of plant mitochondrial genomes are complex and can include various conformations such as Y-type, H-type, single-ring, linear, and multi-ring structures. The mitochondria of *G. paraguayense* in this study are of a ring-type configuration, with the size and form of the mitochondrial genome varying across plant species^[8]. The structure and function of mitochondria are intricately linked^[9], and the mitochondrial genomes of different species exhibit considerable variability, with frequent recombination and horizontal gene transfer^[10]. Plant mitochondrial are key organelles responsible for important cellular processes such as metabolism, respiration, and energy conversion^[11]. Mitochondria play a key role in essential cellular processes such as metabolism, respiration, and energy conversion, and are often referred to as the "powerhouses" of the cell. As such, mitochondrial genomes are invaluable tools for studying eukaryotic evolution, genetic diversity, species identification, and new cultivar breeding. The mitochondrial DNA structure can also be used to trace common ancestry among different species^[12].

Hengwu Ding et al. sequenced the mitogenome of *Sedum plumbizincicola* (Crassulaceae) using high-throughput sequencing technologies, combining long reads (Oxford Nanopore Technology) and short reads (Illumina). The mitogenome of *S. plumbizincicola* is 212,159 bp in size, with a GC content of 44.5%. It includes 31 protein-coding genes (PCGs), 14 tRNAs, 3 rRNAs, 2 ORFs, and 11 pseudogenes. Notably, there are 508 RNA editing sites, with 496 codon changes, most of which are nonsynonymous conversions. This RNA editing significantly mitigates the impact of DNA mutations. Phylogenetic analysis showed that *S. plumbizincicola* is closely related to *Sedum album* within the *Crassulaceae* family^[7]. Similarly, Xiaolei Yu et al. sequenced the organelle genomes of three *Rhodiola* species (*R. wallichiana*, *R. crenulata*, and *R. sacra*) from the Qinghai-Tibet Plateau and compared their mitogenomes. The study revealed that although these species exhibited structural diversity, their mitogenomes showed slower rates of sequence evolution, with *R. wallichiana* and *R. sacra* being more closely related than *R. crenulata*^[13].

While significant progress has been made in studying the mitogenomes of *Crassulaceae* plants, there remains a lack of research, especially concerning *G. paraguayense*. Given the ecological significance of *Crassulaceae* plants, understanding the mitogenomes of these species is crucial for elucidating their classification, phylogenetic relationships, and unique genetic phenomena. Furthermore, such studies could offer new insights into the evolutionary dynamics of *Crassulaceae* mitogenomes, which may also benefit future plant mitogenome research and breeding programs.

In this study, we utilized Real-Time DNA Sequencing of Single Polymerase Molecules to sequence the full mitochondrial genome of *G. paraguayense*. The mitochondrial genome was then assembled and annotated, followed by sequence collinearity analysis and the construction of a mitochondrial genome

map for *G. paraguayense*. We also performed codon usage bias analysis, repeat sequence identification, and Ka/Ks ratio analysis of the mitogenome. Additionally, we conducted phylogenetic tree analysis and examined chloroplast, nuclear, and mitochondrial homologous sequences. The mitochondrial genome data of *G. paraguayense* obtained in this study will help improve our understanding of the species' evolutionary relationships and provide valuable insights for its classification.

2. Materials and methods

2.1. Plant samples and mitochondrial genome sequencing

Fresh leaves of *Graptopetalum paraguayense* were collected from the Nanjing Ligan Flower Base, Nanjing, Jiangsu Province, China (32°12'N, 119°10'E). Pick the succulent leaves that are growing well, fresh and undamaged. Healthy, undamaged succulent leaves were carefully selected, thoroughly washed with sterile water, and stored at -80°C. We commissioned Nanjing Genepioneer Biotechnologies (Nanjing, China) use Real-Time DNA Sequence from Single Polymerase Molecules. After quality assessment of the DNA samples, the DNA was fragmented and repair of DNA damage, end repair, adapter ligation, enzymatic digestion, and purification. Sequencing was performed on the PacBio Sequel II platform in HiFi mode. Then use the perl script to count the third-generation sequencing data.

HiFi raw reads were assembled using PMAT2 software (v2.0.2, parameter: -g 5G). Assembly results were visualized and manually adjusted using Bandage (v0.8.1). HiFi reads were subsequently aligned to the assembled sequence using minimap2, and sequence polishing was performed with NextPolish to generate the final mitogenome assembly and adjust the corresponding GFA file.

Protein-coding genes (PCGs) and ribosomal RNA (rRNA) genes were first aligned against published plant mitochondrial reference sequences using BLAST (v2.13.0), and the annotations were then manually refined based on information from closely related species. Transfer RNA (tRNAs) genes were annotated using tRNAscanSE (<http://lowelab.ucsc.edu/tRNAscan-SE/>)^[14]. Open reading frames (ORFs) were annotated by OpenReading Frame Finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) with a minimum length of 102 bp, excluding redundant sequences or those overlapping known genes. Sequences >300bp were annotated by comparison with the nr database. RNA editing sites were predicted using PmtREP (<http://112.86.217.82:9919/#/tool/alltool/detail/336>)^[15], and final annotations were verified and manually curated. Finally. The circular mitochondrial genome map was drawn using OGDRAW (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>).

2.2. Relative synonymous codon usage (RSCU) analyses

Due to the principle of codon degeneracy, there are great differences in the usage of genome codons among different species, and the inequality of synonymous codon usage is called relative synonymous codon usage (RSCU), which is usually considered to be the comprehensive result of natural selection, species mutation and genetic drift. We use the perl script of Nanjing Genepioneer Biotechnologies (Nanjing,China) filter uniq CDS and calculate it. The calculation method of codon preference is: (the

number of one codon encoding an amino acid / the number of all codons encoding the amino acid) / (1 / the type of codon encoding the amino acid), that is, (the actual use frequency of the codon / the theoretical use frequency of the codon). If RSCU > 1, it indicates that the codon is widely used. If RSCU < 1, it indicates that the codon is not often used. If RSCU = 1, it indicates that the codon is not biased in use.

2.3. Repeat sequences analysis

Repeat sequences included simple sequence repeats (SSR), tandem and dispersed. SSR was identified by MISA software (v1.0, parameter: 1–10 2–5 3–4 4 – 3 5 – 3 6 – 3)^[16], and tandem was identified by TRF software (trf409.linux64, parameter: 27 7 80 10 50 2000-f-d-m)^[17]. Dispersed was identified using BLASTN (v2.10.1, parameters: -word _ size 7, evaluate 1e-5, remove redundancy, remove tandem duplication) software and visualized using CIRCOs v0.69-5^[18].

2.4. Ka/Ks analyses

The ratio of Ka to Ks indicates the selection effect. Ka/Ks > 1 indicating a positive selection effect, Ka/Ks < 1 indicating a negative selection effect. We grouped the species selected for advanced analysis in pairs and extracted homologous gene pairs. These homologous gene pairs were then aligned using MAFFT v7.427 (<https://mafft.cbrc.jp/alignment/software/>)^[19]. After alignment, KaKs_Calculator v2.0 (<https://sourceforge.net/projects/kakscalculator2/>) was used to calculate the Ka and Ks values of each gene pair, with the MLWL method selected for the calculations^[20]. Finally, the ka/ks values of each pair of genes were counted and the box plot was drawn.

2.5. Mitogenome syntenic analysis

Using the nucmer (4.0.0beta2) software and the maxmatch parameter, genomic alignment between other sequences and assembly sequences was performed to generate Dot-plot plots^[21].

2.6. Nucleotide Diversity and Phylogenetic Analysis

The complete mitogenomes of 14 species (Table S1) were downloaded from GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>). Homologous gene sequences were globally aligned using MAFFT (v7.427, –auto mode), and nucleotide diversity (pi values) was calculated with DnaSP v^[19].

For phylogenetic analysis, coding sequences (CDSs) were aligned with MAFFT and concatenated. The alignment was trimmed using trimAl (v1.4.rev15, parameter: –gt 0.7)^[22]. The best-fit substitution model (GTR) was selected using jModelTest (v2.1.10). A maximum likelihood (ML) phylogenetic tree was then constructed using RAxML (v8.2.10, GTRGAMMA model, 1,000 bootstraps)^[23].

2.7. Migration analyses of chloroplast-to-mitochondrial DNA

Homologous sequences between chloroplast and mitochondrial genomes were identified using BLAST with an E-value threshold of 1e–5 and a minimum sequence identity of 70%.

3. Results

3.1. Mitochondrial gene organization and features

The complete mitochondrial genome of *Graptopetalum paraguayense* has been deposited in NCBI under accession number PV256627. A circular map of the genome was generated using the OGDRAW platform (Fig. 1). The mitochondrial genome is 242,059 bp in length and has a GC content of 43.65%. A total of 50 genes were annotated, comprising 13 tRNA genes, 3 rRNA genes, 31 protein-coding (mRNA) genes, and 3 pseudogenes (Table 1, TableS2). Detailed statistics are provided in Table S3. These results confirm that the assembled mitochondrial genome of *G. paraguayense* is of high quality.

Table 1
The genes identified in the mitochondrial genome of *G. paraguayense*

Group of genes	Gene name
ATP synthase	<i>atp1 atp4 atp6 atp8 atp9</i>
Cytochrome c biogenesis	<i>ccmB ccmC ccmFc* ccmFn*</i>
Ubiquinol cytochrome c reductase	<i>cob</i>
Cytochrome c oxidase	<i>cox1 cox2** cox3</i>
Maturases	<i>matR</i>
Transport membrane protein	<i>mttB</i>
NADH dehydrogenase	<i>nad1**** nad2**** nad3 nad4*** nad4L nad5**** nad6 nad7**** nad9</i>
Ribosomal proteins (LSU)	<i>rpl10 rpl16 rpl5</i>
Ribosomal proteins (SSU)	<i>#rps3 #rps4 rps12 rps13 rps14 rps7</i>
Succinate dehydrogenase	<i>#sdh4</i>
Ribosomal RNAs	<i>rrn18 rrn26 rrn5</i>
Transfer RNAs	<i>trnE-TTC trnF-GAA(2) trnH-GTG trnM-CAT(6) trnQ-TTG trnW-CCA trnY-GTA</i>
Notes: *: intron number; #Gene: Pseudo gene;Gene(2): Number of copies of multi-copy genes.	

RNA-editing sites in the mitochondrial genome were predicted based on multiple sequence alignment. The amino acid sequence of each target gene was compared with homologous sequences in the database, and changes at corresponding positions were recorded. Sites that fulfilled the criteria for RNA

editing were considered potential RNA-editing sites^[24]. We statistically analyzed the number of RNA edits for each gene (Fig. 2). After RNA editing, the hydrophilic and hydrophobic properties of the amino acids produced by codons may change (Fig. 2, Table 2).

Table 2
Effects of RNA editing on changes hydrophilicity and hydrophobicity of amino acid in the mitogenome of *G. paraguayense*.

Type	RNA-editing	Number	Percentage
hydrophilic-hydrophilic	CAC (H) = > TAC (Y)	8	
	CAT (H) = > TAT (Y)	17	
	CGC (R) = > TGC (C)	11	
	CGT (R) = > TGT (C)	27	
	total	63	10.52%
hydrophilic-hydrophobic	ACA (T) = > ATA (I)	6	
	ACC (T) = > ATC (I)	6	
	ACG (T) = > ATG (M)	9	
	ACT (T) = > ATT (I)	7	
	CGG (R) = > TGG (W)	36	
	TCA (S) = > TTA (L)	78	
	TCC (S) = > TTC (F)	34	
	TCG (S) = > TTG (L)	48	
	TCT (S) = > TTT (F)	60	
	total	284	47.41%
	CAG (Q) = > TAG (X)	1	
	total	2	0.33%
hydrophobic-hydrophilic	CCA(P) = > TCA(S)	8	
	CCC (P) = > TCC (S)	17	
	CCG (P) = > TCG (S)	5	
	CCT (P) = > TCT (S)	19	
	total	49	8.18%
hydrophobic-hydrophobic	CCA (P) = > CTA (L)	45	
	CCC (P) = > CTC (L)	15	
	CCC (P) = > TTC (F)	4	

Type	RNA-editing	Number	Percentage
	CCG (P) = > CTG (L)	35	
	CCT (P) = > CTT (L)	27	
	CCT (P) = > TTT (F)	14	
	CTC (L) = > TTC (F)	15	
	CTT (L) = > TTT (F)	22	
	GCA (A) = > GTA (V)	8	
	GCC (A) = > GTC (V)	4	
	GCG (A) = > GTG (V)	7	
	GCT (A) = > GTT (V)	5	
	total	201	33.56%
	All	599	100%

3.2. Codon use analysis of RSCU

The mitochondrial genome of *Graptopetalum paraguayense* encodes 20 amino acids using the standard set of 64 codons. Our analysis showed that 30 codons have a RSCU value greater than 1, 32 codons have an RSCU value less than 1, and only one or two codons exhibit an RSCU value equal to 1, indicating no usage bias for those codons. These unbiased codons are AUG (Met) and UGG (Trp). Among stop codons, UAA has the highest RSCU value (1.9355), followed by GCU, which encodes alanine (Ala) with an RSCU of 1.6244. In contrast, the codon UAC, which encodes tyrosine (Tyr), shows the lowest RSCU value (0.438), suggesting it is rarely used (Fig. 3, Table S4).

Further analysis of specific amino acids revealed similar patterns of preferential codon usage. Among the four codons encoding Ala, GCU is used most frequently (RSCU = 1.6244). For arginine (Arg), CGA shows the highest usage among its six synonymous codons (RSCU = 1.3297). Among the four codons encoding glycine (Gly), GGA is the most frequent (RSCU = 1.4961). For leucine (Leu), UUA is the most preferred codon among the six options (RSCU = 1.5226). Similarly, for serine (Ser), UCU exhibits the highest usage frequency (RSCU = 1.28) (Fig. 3, Table S4).

3.3. Analysis of repeat structures in the *Graptopetalum.paraguayense* mitochondrial genomes

The repeat structures identified in the mitochondrial genome of *G. paraguayense* comprise three types: simple sequence repeats (SSRs), tandem repeats, and dispersed repeats. A total of 122 repeat elements were detected, including 59 SSRs, 1 tandem repeat, and 62 dispersed repeats (Fig. 4, TableS5-S8). The

circular representation in Fig. 4 clearly illustrates the distribution and connections of these repeats across the mitogenome.

3.4. Ka/Ks analyses

In order to analyze the substitution rate of the mitochondrial genome of the *H. yue* plant, We calculated *G. paraguayense* with *Tetragonia tetragonoides*, *Sedum plumbizincicola*, *Rhodiola crenulata*, *Mammillaria huitzilopochtli*, *Mesembryanthemum crystallinum*, *Nopalea cochenillifera* and *Rhodiola rosea*, Ka/Ks of 30 common genes, a box plot was made. These genes include: *atp1* *atp4* *atp6* *atp8* *atp9* *ccmB* *ccmC* *ccmFc* *ccmFn* *cob* *cox1* *cox2* *cox3* *matR* *mttB* *nad1* *nad2* *nad3* *nad4* *nad4L* *nad5* *nad6* *nad7* *nad9* *rpl10* *rpl16* *rpl5* *rps12* *rps13* *rps7*. Among them, only the median ka/Ks values of *ccmB* and *nad7* genes were greater than 1, showing a positive selection effect. There are 28 genes including *atp1* *atp4* *atp6* *atp8* *atp9* *ccmC* *ccmFc* *ccmFn* *cob* *cox1* *cox2* *cox3* *matR* *mttB* *nad1* *nad2* *nad3* *nad4* *nad4L* *nad5* *nad6* *nad9* *rpl10* *rpl16* *rpl5* *rps12* *rps13* *rps7*. The Ka/Ks values of *rpl10*, *rpl16*, *rpl5*, *rps12*, *rps13* and *rps7* are less than 1, showing a negative selection effect, indicating that they have a purification selection effect. No gene has a Ka/Ks value equal to 1 (Fig. 5, Table S9).

3.5. Mitogenome synteny analysis

The genome sequence of *Graptopetalum paraguayense* was compared pairwise with those of seven other plants to assess mitochondrial genome collinearity. The dot-plot results demonstrate that *G. paraguayense* shares numerous homologous collinear fragments with *Tetragonia tetragonoides*, *Sedum plumbizincicola*, *Rhodiola crenulata*, *Mammillaria huitzilopochtli*, *Mesembryanthemum crystallinum*, *Nopalea cochenillifera*, and *Rhodiola rosea*. Among these comparisons, the homologous sequences between *G. paraguayense* and *S. plumbizincicola* are the longest and exhibit the highest similarity (57.28%). In contrast, the homologous sequences between *G. paraguayense* and *M. crystallinum* account for only 10.72% of the total mitochondrial genome, indicating relatively low collinearity (Fig. 6, Table S10).

3.6. Nucleotide Diversity and Phylogenetic Analysis of *G. paraguayense*

Analysis of nucleotide diversity (P_i) in the protein-coding genes (PCGs) of the *G. paraguayense* mitochondrial genome revealed marked heterogeneity among genes. The *atp9* gene exhibited the highest P_i value (0.15809), indicating the greatest sequence variability and potential evolutionary significance, followed by *atp8* ($P_i = 0.08645$) and *rpl5* ($P_i = 0.07741$). These highly variable regions ($P_i \geq 0.05$) may serve as potential molecular markers for future population-genetic or phylogenetic investigations (Fig. 7, Table S11).

Phylogenetic analysis further revealed that *G. paraguayense* clusters with *Sedum plumbizincicola* as a strongly supported sister taxon. Species of *Rhodiola* formed a well-supported monophyletic clade, although divergence was evident within the genus; for example, *R. crenulata* was more closely related to

R. rosea. This topology provides robust phylogenetic evidence for the taxonomic placement of *G. paraguayense* within the Acre clade and reflects its relatively low mitochondrial sequence variation, as illustrated by the 57.28% collinearity observed with *S. plumbizincicola*.

3.7. Analysis of homologous sequence of chloroplast and mitochondria

BLAST searches were performed to identify homologous sequences shared between the chloroplast and mitochondrial genomes of *G. paraguayense*. The highest sequence similarity score reached 9378. The homologous region spans positions 113,918 – 119,671 bp in the chloroplast genome and positions 152,894 – 147,275 bp in the mitochondrial genome. Within this chloroplast region, the homologous segment contains the genes *psaC*, *ndhE*, *ndhG*, *ndhI*, *ndhA*, and *ndhH* (Fig. 9).

4. Discussion

G. paraguayense is a perennial succulent plant in the genus of the family Crassulaceae. It possesses significant ornamental value, medicinal properties, and is recognized as a functional food^[1]. Mitochondria, essential organelles in living organisms, primarily facilitate energy production and are crucial for sustaining vital activities^{[25][26]}. Composed of dual membranes (outer and inner), mitochondria contain their own DNA and ribosomes^[27]. They play pivotal roles in calcium ion storage and apoptosis regulation^[28]. For secondary quality control, 5,000 randomly selected reads were aligned against the NT database (blastn v2.13.0), revealing 37.92% unmatched to known species, while all matched species constituted $\leq 2.0\%$ (*G. amethystinum* highest at 2.00%).

4.1. Assembly and Evolutionary Conservation of the *G. paraguayense* Mitochondrial Genome

Based on PacBi Revio and other platforms, this study successfully assembled the mitochondrial genome of *G. paraguayense*, ensuring high resolution and validating the reliability of the assembly results. The mitochondrial genome size of *G. paraguayense* is 242,209 bp, which falls within a medium size range. It is larger than *Rhodiola juparensis* (202,019 bp) and *Rhodiola crenulata* (194,106 bp), but smaller than *Sesuvium portulacastrum* (392,221 bp), *Tetragonia tetragonoides* (347,227 bp), *Mesembryanthemum crystallinum* (1,005,707 bp), *Mammillaria huiatzilopochtli* (2,052,004 bp), *Selenicereus monacanthus* (2,290,019 bp), *Nopalea cochenillifera* (1,156,235 bp), *Pereskia aculeata* (515,187 bp), *Hevea spruceana* (935,732 bp), and *Excoecaria agallocha* (704,003 bp). Its size is similar to *Sedum plumbizincicola* (212,159 bp), *Rhodiola tangutica* (257,378 bp), and *Rhodiola rosea* (259,150 bp).

Similar to most plant mitogenomes (e.g., *Helianthus annuus* and *Sedum plumbizincicola*), the *rps2* gene was undetected in the *G. paraguayense* mitochondrial genome^[29]. Additionally, the absence of *rps10* and *rps11* genes was observed in *S. plumbizincicola*, *R. juparensis*, *R. tangutica*, *R. rosea*, and *R. crenulata*, suggesting potential early evolutionary loss of these genes. The *G. paraguayense* mitogenome encodes

13 tRNAs and 3 rRNAs, consistent with counts in related species: *S. plumbizincicola* (14 tRNAs, 3 rRNAs), *R. juparensis* (11 tRNAs, 3 rRNAs), and *R. tangutica* (15 tRNAs, 3 rRNAs). The overall GC content (43.65%) aligns with values in *S. plumbizincicola* (44.51%), *R. juparensis* (44.58%), *R. tangutica* (44.24%), *R. rosea* (44.88%), and *R. crenulata* (45.21%). Notably, the GC content of protein-coding regions (PCGs, 43.1%) was significantly lower than in tRNA-coding regions (50.42%) and rRNA-coding regions (51.56%). Collectively, these findings indicate conserved evolutionary features in the *G. paraguayense* mitogenome shared with related species. The pervasive loss of *rps2*, *rps10*, and *rps11* across crassulaceae plants, combined with stable tRNA/rRNA gene counts and uniform GC content, suggests strong selective constraints on mitogenome architecture, potentially linked to functional optimization.

4.2. Analysis of codon usage

There is a significant codon usage bias in this genome. Of all 64 codons, 30 had an RSCU value greater than 1 (used more frequently than expected by chance) and 32 had an RSCU value less than 1 (used less frequently than expected). Only two codons, AUG-Met and UGG-Trp, had RSCU values equal to 1, indicating no preference for their use. Among them, the stop codon UAA showed the strongest bias (RSCU = 1.9355), while the tyrosine codon UAC showed the weakest bias (RSCU = 0.438). Among the synonymous codons of multiple amino acids, alanine (Ala) preferred GCU (RSCU = 1.6244), arginine (Arg) preferred CGA (RSCU = 1.3297), and glycine (Gly) preferred GGA (RSCU = 1.4961). Leucine (Leu) preferred UUA (RSCU = 1.5226), and serine (Ser) preferred UCU (RSCU = 1.28). These preference patterns reflect the combined effects of natural selection, mutation pressure, and other factors on the mitochondrial genome during evolution^[30]. In the mitochondrial genome of *G. paraguayense*, there were 30 codons with RSCU values greater than 1, of which 12 ended with A base and 16 ended with U base, indicating that the third base of the codon of *G. paraguayense* was preferred to use A/U.

4.3. Analysis of repetitive sequences in the mitochondrial genome of *G. paraguayense* and Ka/Ks ratio assessment

This study identified 122 repeat sequences in the *G. paraguayense* mitochondrial genome, including 59 simple repeats, 1 tandem repeat, and 62 dispersed repeats. The tandem repeat has a length of 39 bp, spanning from position 89779 to 89855, as marked by the red short line in Fig. 4. The shortest dispersed repeat is 30 bp in length, while the longest dispersed repeat measures 3472 bp. Their detailed locations are provided in the supplementary materials (Table S5-S8).

To assess mitochondrial genome substitution rates in *G. paraguayense*, we calculated Ka/Ks ratios for 30 shared protein-coding genes across seven comparative species (*G. paraguayense*, *T. tetragonoides*, *S. plumbizincicola*, *R. crenulata*, *M. huitzilopochtli*, *M. crystallinum*, *N. cochenillifera*, *R. rosea*) (Table S9). Boxplot analysis revealed that only *ccmB* and *nad7* exhibited median Ka/Ks > 1, indicating positive selection (Fig. 5). It is speculated that the *ccmB* and *nad7* genes may be key signals of *G. paraguayense* fast-adaptive evolutionary genes. The remaining 28 genes showed Ka/Ks < 1, This indicates that these PCGs and the key functional genes possess a high degree of conservation (Fig. 5)^[31].

4.4. Phylogenetic and Pi analysis

Based on the nucleotide diversity (Pi) analysis, this study identified several highly variable genes ($P_i \geq 0.05$) in the mitochondrial genome, including *atp9* ($P_i = 0.15809$), *atp8* ($P_i = 0.08645$), *rp15* ($P_i = 0.07741$), *cox2* ($P_i = 0.0662$), *ccmFn* ($P_i = 0.06365$), *rps7* ($P_i = 0.06349$), *ccmC* ($P_i = 0.05701$), *ccmFc* ($P_i = 0.05425$), *mttB* ($P_i = 0.0534$), *nad6* ($P_i = 0.05369$), and *rps13* ($P_i = 0.05299$). These genes not only exhibit significant nucleotide diversity but are also associated with a high number of variant sites (e.g., 66 sites in *atp9* and up to 194 sites in *ccmFn*), indicating faster evolutionary rates and rich genetic polymorphism^[32]. Therefore, they could serve as potential molecular markers for subsequent phylogenetic analyses of closely related species, investigations into population genetic structures, and studies of adaptive evolution. This provides critical insights into understanding the genetic diversity and evolutionary history of *G. paraguayense*^[33].

Phylogenetic analysis strongly supported *G. paraguayense* and *S. plumbizincicola* as sister taxa, confirming *G. paraguayense*'s placement within the Acre clade. Rhodiola formed a monophyletic clade, albeit with internal divergence, such as the close relationship between *R. crenulata* and *R. rosea*.

5. Conclusions

This study successfully assembled and annotated the complete mitochondrial genome of *G. paraguayense* for the first time. The genome measures 242,059 base pairs with a GC content of 43.65 percent, containing 50 functional genes including 31 protein-coding genes, 13 tRNA genes, 3 rRNA genes along with 3 pseudogenes. Analysis identified 599 RNA editing sites in the mitochondrial genome, with nearly half converting hydrophilic amino acids to hydrophobic forms, indicating RNA editing likely plays a significant role in regulating protein function^[34].

Codon usage analysis revealed preferential use of GCU for alanine, CGA for arginine and UUA for leucine, while the stop codon UAA showed the strongest bias with an RSCU value of 1.9355. The genome contained 122 repetitive sequences, predominantly 62 dispersed repeats that may contribute substantially to structural variation. Evolutionary analysis detected positive selection acting on *ccmB* and *nad7* genes, suggesting their importance in adaptive evolution^[35]. Comparative genomic analysis demonstrated closest similarity with *Sedum plumbizincicola* at 57.28 percent sequence identity, supporting their phylogenetic relationship. These eleven highly variable regions (*atp9*, *atp8*, *rp15*, *cox2*, *ccmFn*, *rps7*, *ccmC*, *ccmFc*, *mttB*, *nad6*, and *rps13*) can serve as potential molecular markers in population genetics. The study also found evidence of chloroplast gene transfer to mitochondria involving *psaC* and *ndh* genes, providing new evidence of inter-organellar DNA exchange. Phylogenetic reconstruction confirmed *G. paraguayense* and *Sedum plumbizincicola* as sister species within the Acre clade of Crassulaceae, while Rhodiola species formed a distinct monophyletic group. In summary, this pioneering work characterizes the mitochondrial genome architecture and evolutionary patterns of *G. paraguayense*, advancing our understanding of Crassulaceae genomics. The findings establish a foundation for future investigations into plant organelle evolution, adaptive mechanisms and phylogenetic relationships. The study provides valuable genomic resources for further research on this economically and medicinally important succulent species.

Declarations

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work was supported by the Natural Science Foundation from Jiangsu Province (No. BK20211128) and the project of innovative entrepreneurship training for college students in Jiangsu province (No. 202313843019Y, No. 202313843009Y and No. S202513843029).

Author Contribution

X Z: Data curation, Formal analysis, Funding acquisition, Investigation, Writing –original draft, Writing–review&editing. CQ L: Data curation, Formal analysis, Writing–original draft, Methodology. ZR L: Data curation, Writing–original draft, Methodology. L Z: Methodology. X D: supervised the project and revised the manuscript. All authors reviewed the manuscript.

Data Availability

The data is available at NCBI accession: PV256627.

References

1. V. Jaiswal, H.-J. Lee, A Comprehensive Review on *Graptopetalum paraguayense*'s Phytochemical Profiles, Pharmacological Activities, and Development as a Functional Food, 14(3) (2025) 349.
2. Y.-X. Chen, P.T.N. Le, T.-T. Tzeng, T.-H. Tran, A.T. Nguyen, I.H.-J. Cheng, C.-Y.F. Huang, Y.-J. Shiao, T.-T. Ching, *Graptopetalum paraguayense* Extract Ameliorates Proteotoxicity in Aging and Age-Related Diseases in Model Systems, 13(12) (2021) 4317.
3. T.F.E. Messerschmid, J.T. Klein, G. Kadereit, J.W. Kadereit, Linnaeus's folly – phylogeny, evolution and classification of *Sedum* (Crassulaceae) and Crassulaceae subfamily Sempervivoideae, *Taxon* 69(5) (2020) 892-926.
4. R. Acevedo-Rosas, K. Cameron, V. Sosa, S. Pell, A molecular phylogenetic study of *Graptopetalum* (Crassulaceae) based on ETS, ITS, RPL16, and TRNL-F nucleotide sequences, *Am J Bot* 91(7) (2004) 1099-104.

5. D.C. Logan, The mitochondrial compartment, *Journal of Experimental Botany* 57(6) (2006) 1225-1243.
6. I.M. Møller, A.G. Rasmusson, O. Van Aken, Plant mitochondria – past, present and future, *The Plant Journal* 108(4) (2021) 912-959.
7. H. Ding, D. Bi, S. Zhang, S. Han, Y. Ye, R. Yi, J. Yang, B. Liu, L. Wu, R. Zhuo, X. Kan, The Mitogenome of *Sedum plumbizincicola* (Crassulaceae): Insights into RNA Editing, Lateral Gene Transfer, and Phylogenetic Implications, 11(11) (2022) 1661.
8. H. Štorchová, M. Krüger, Methods for assembling complex mitochondrial genomes in land plants, *Journal of Experimental Botany* 75(17) (2024) 5169-5174.
9. S. Khosravi, M.E. Harner, The MICOS complex, a structural element of mitochondria with versatile functions, 401(6-7) (2020) 765-778.
10. S.-i. Arimura, I. Nakazato, Genome Editing of Plant Mitochondrial and Chloroplast Genomes, *Plant and Cell Physiology* 65(4) (2023) 477-483.
11. X.J. Chen, R.A. Butow, The organization and inheritance of the mitochondrial genome, *Nature Reviews Genetics* 6(11) (2005) 815-825.
12. J. Li, H. Tang, H. Luo, J. Tang, N. Zhong, L. Xiao, Complete mitochondrial genome assembly and comparison of *Camellia sinensis* var. *Assamica* cv. *Duntsa*, *Frontiers in Plant Science* 14 (2023).
13. X. Yu, P. Wei, Z. Chen, X. Li, W. Zhang, Y. Yang, C. Liu, S. Zhao, X. Li, X. Liu, Comparative analysis of the organelle genomes of three *Rhodiola* species provide insights into their structural dynamics and sequence divergences, *BMC Plant Biol* 23(1) (2023) 156.
14. P.P. Chan, T.M. Lowe, tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences, *Methods Mol Biol* 1962 (2019) 1-14.
15. J.P. Mower, PREP-Mt: predictive RNA editor for plant mitochondrial genes, *BMC Bioinformatics* 6(1) (2005) 96.
16. S. Beier, T. Thiel, T. Münch, U. Scholz, M. Mascher, MISA-web: a web server for microsatellite prediction, *Bioinformatics* 33(16) (2017) 2583-2585.
17. G. Benson, Tandem repeats finder: a program to analyze DNA sequences, *Nucleic Acids Research* 27(2) (1999) 573-580.
18. M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones, M.A. Marra, Circos: An information aesthetic for comparative genomics, *Genome Research* 19(9) (2009) 1639-1645.
19. K. Katoh, D.M. Standley, MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability, *Molecular Biology and Evolution* 30(4) (2013) 772-780.
20. D. Wang, Y. Zhang, Z. Zhang, J. Zhu, J. Yu, KaKs_Calculator 2.0: A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies, *Genomics, Proteomics & Bioinformatics* 8(1) (2010) 77-80.

21. A.E. Darling, G. Marçais, A.L. Delcher, A.M. Phillippy, R. Coston, S.L. Salzberg, A. Zimin, MUMmer4: A fast and versatile genome alignment system, *PLOS Computational Biology* 14(1) (2018).
22. S. Capella-Gutiérrez, J.M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses, *Bioinformatics* 25(15) (2009) 1972-1973.
23. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, *Bioinformatics* 30(9) (2014) 1312-1313.
24. Mower, J.P. PREP-Mt: predictive RNA editor for plant mitochondrial genes. *BMC Bioinformatics* 6, 96 (2005).
25. J. Nunnari, A. Suomalainen, Mitochondria: In Sickness and in Health, *Cell* 148(6) (2012) 1145-1159.
26. P. Andrieux, C. Chevillard, E. Cunha-Neto, J.P.S. Nunes, Mitochondria as a Cellular Hub in Infection and Inflammation, 22(21) (2021) 11338.
27. David E. Green, Mitochondria – Structure, Function, and Replication, *The New England Journal of Medicine*, 1983;309:182-183.
28. G. Hajnóczky, E. Davies, M. Madesh, Calcium signaling and apoptosis, *Biochemical and Biophysical Research Communications* 304(3) (2003) 445-454.
29. C.J. Grassa, D.P. Ebert, N.C. Kane, L.H. Rieseberg, Complete Mitochondrial Genome Sequence of Sunflower (*Helianthus annuus* L.), 4(5) (2016) 10.1128/genomea.00981-16.
30. H. Akashi, R.M. Kliman, A. Eyre-Walker, Mutation pressure, natural selection, and the evolution of base composition in *Drosophila*, in: R.C. Woodruff, J.N. Thompson (Eds.), *Mutation and Evolution*, Springer Netherlands, Dordrecht, 1998, pp. 49-60.
31. X. Zhang, P. Li, J. Wang, D. Fu, B. Zhao, W. Dong, Y. Liu, Comparative genomic and phylogenetic analyses of mitochondrial genomes of hawthorn (*Crataegus* spp.) in Northeast China, *International Journal of Biological Macromolecules* 272 (2024) 132795.
32. Z. Wang, R. Wang, Y. Sang, T. Wang, Y. Su, W. Liao, Comparative analysis of mitochondrial genomes of invasive weed *Mikania micrantha* and its indigenous congener *Mikania cordata*, *International Journal of Biological Macromolecules* 281 (2024) 136357.
33. P. Zhou, Q. Zhang, F. Li, J. Huang, M. Zhang, Assembly and comparative analysis of the complete mitochondrial genome of *Ilex metabaptista* (Aquifoliaceae), a Chinese endemic species with a narrow distribution, *BMC Plant Biology* 23(1) (2023) 393.
34. G. Quinones-Valdez, S.S. Tran, H.-I. Jun, J.H. Bahn, E.-W. Yang, L. Zhan, A. Brümmer, X. Wei, E.L. Van Nostrand, G.A. Pratt, G.W. Yeo, B.R. Graveley, X. Xiao, Regulation of RNA editing by RNA-binding proteins in human cells, *Communications Biology* 2(1) (2019) 19.
35. S.E. Faivre-Nitschke, P. Nazoa, J.M. Gualberto, J.M. Grienenberger, G. Bonnard, Wheat mitochondria *ccmB* encodes the membrane domain of a putative ABC transporter involved in cytochrome *c* biogenesis, *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* 1519(3) (2001) 199-208.

Figures

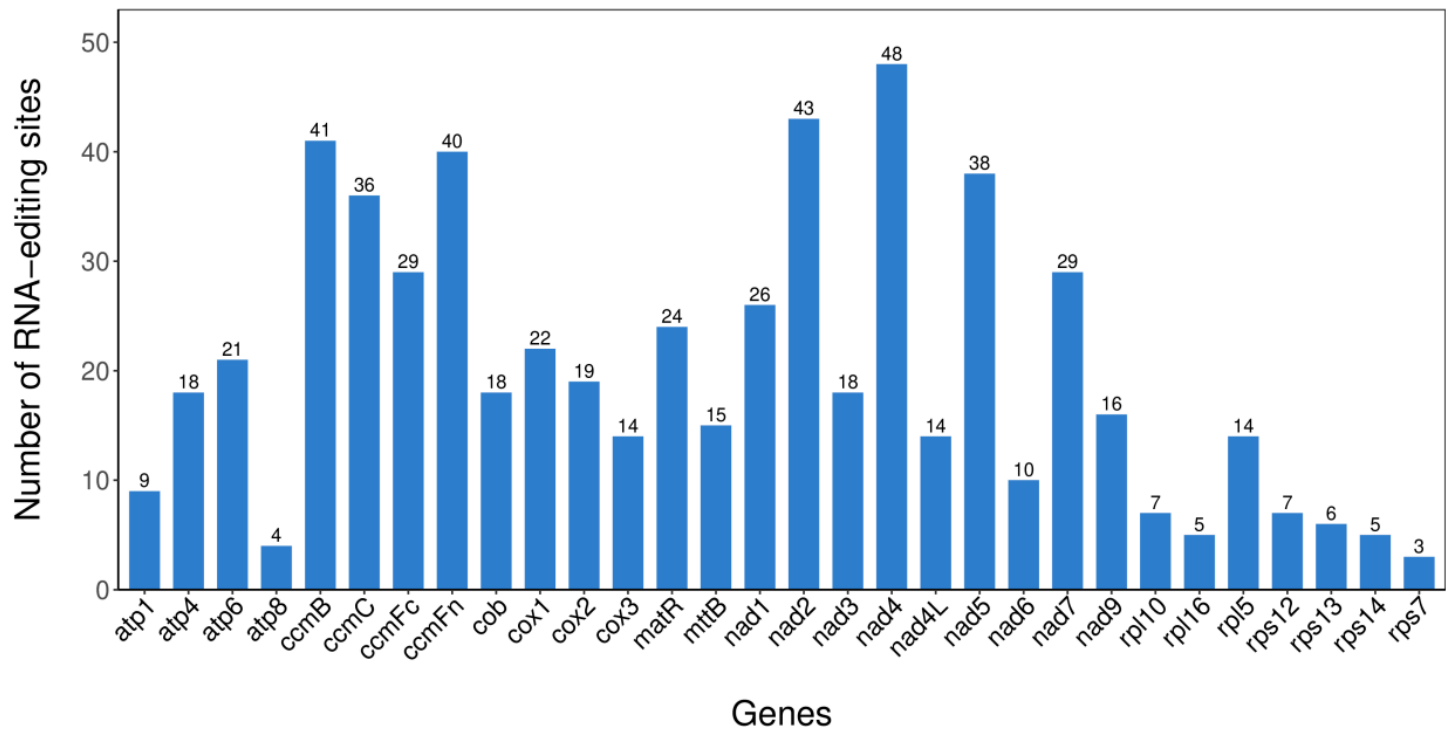


Figure 2

Statistics of the number of RNA editing sites for each gene.

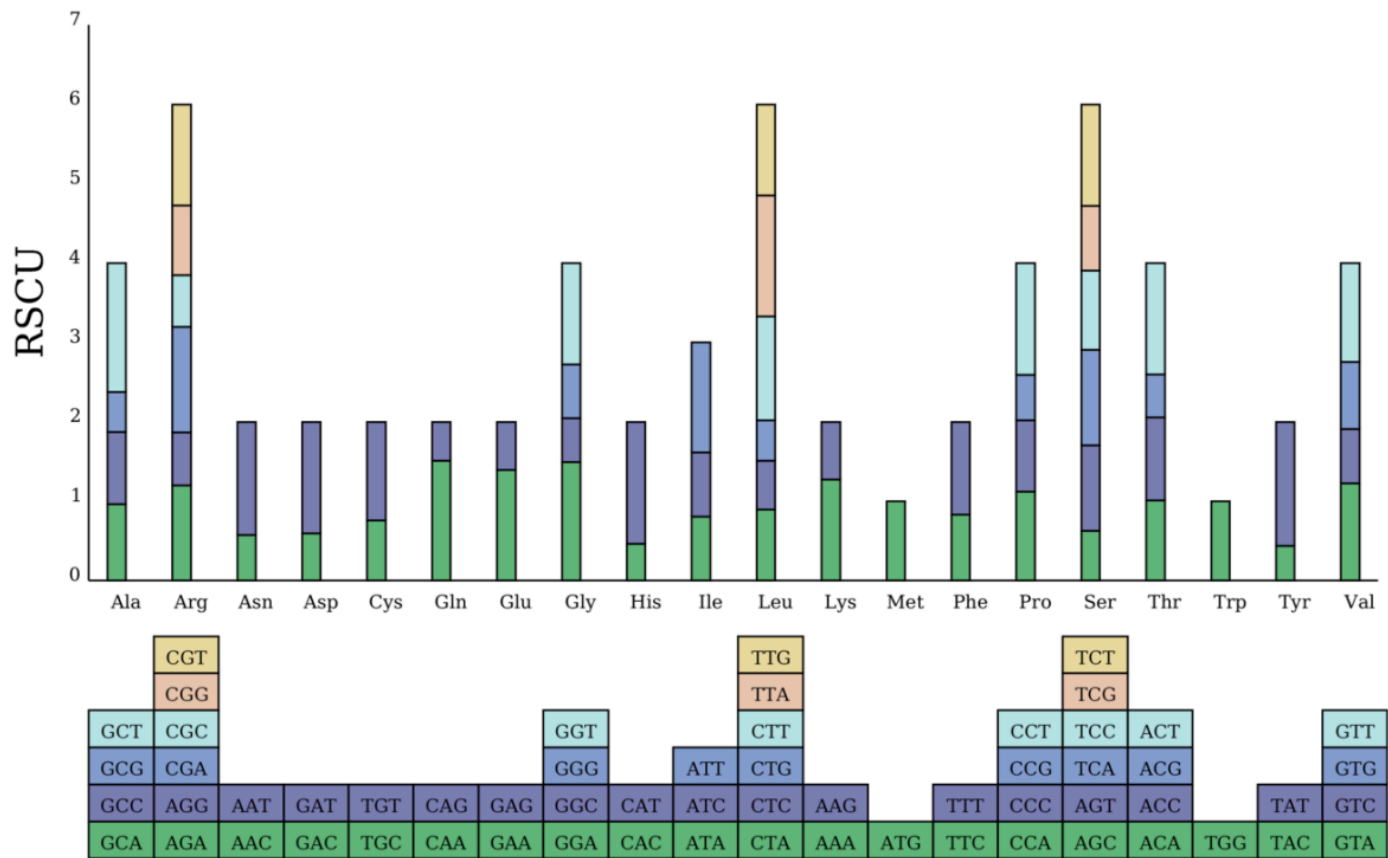


Figure 3

Relative Synonymous Codon Usage (RSCU) Value bar chart. The upper part of the picture shows the total RSCU of the mitochondria of the *G.paraguayense*, and the lower part represents the codons used.

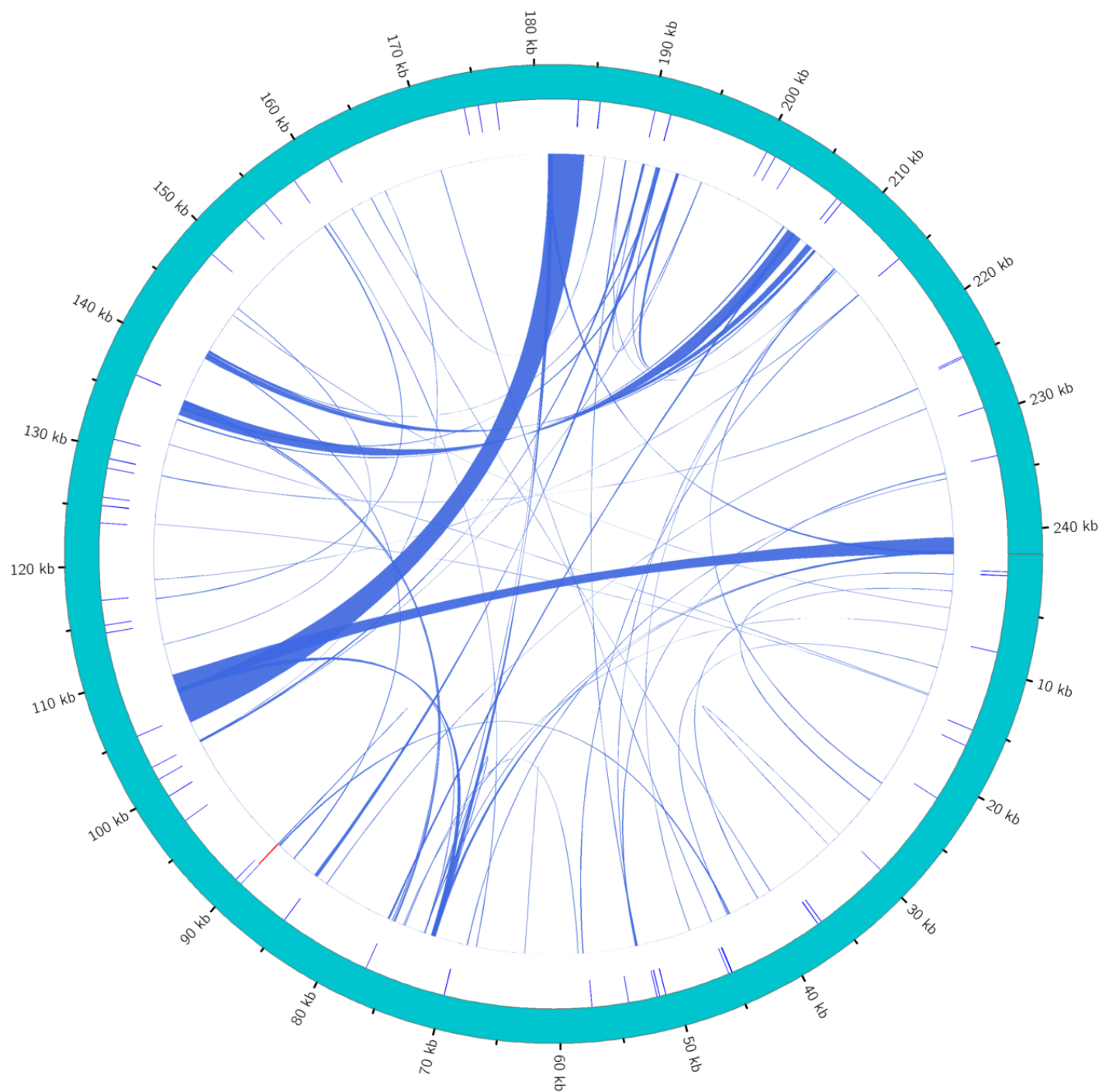


Figure 4

The repeated sequence diagram in the mitogenomes of *G. paraguayense*. The outermost circle represents the mitochondrial genome sequence, followed by simple repeat sequences, tandem repeat

sequences, and scattered repeat sequences in sequence.

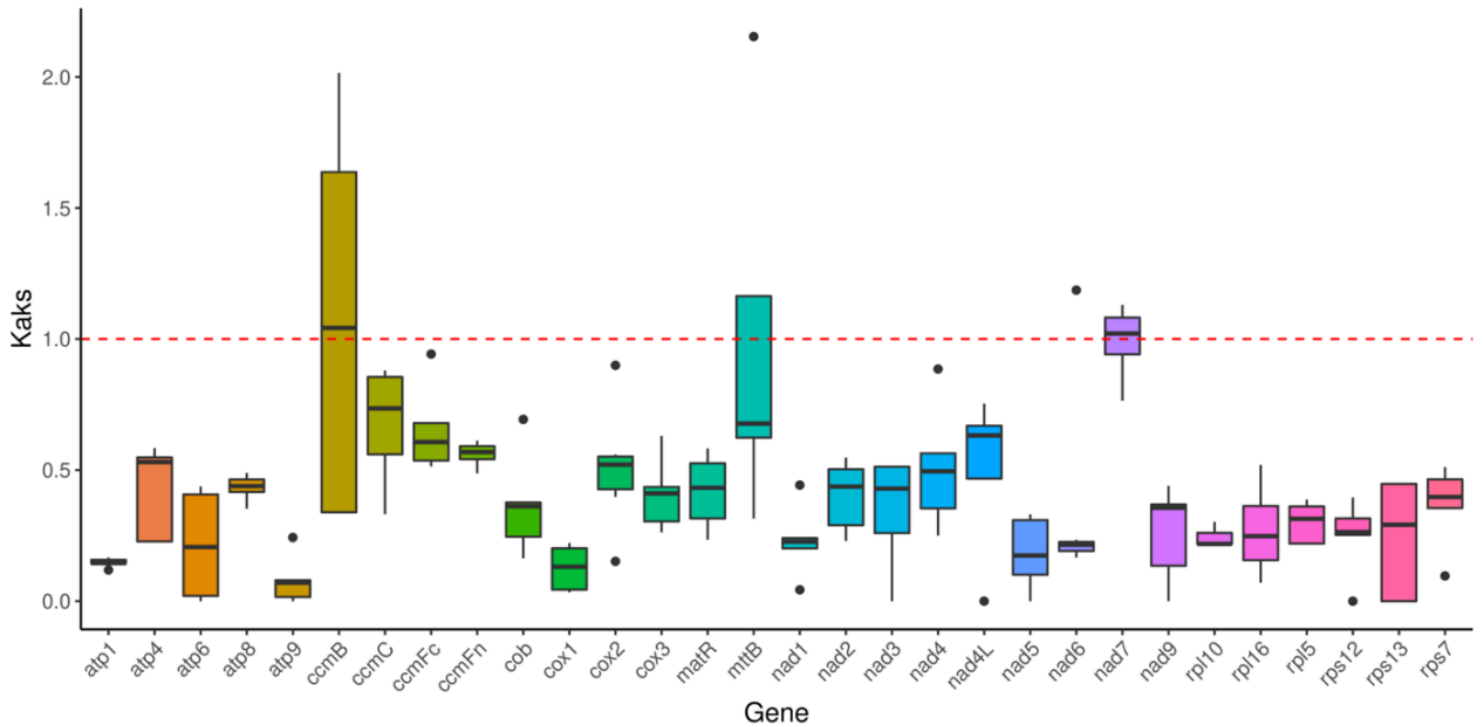


Figure 5

Species-based Ka/Ks bar line graph. The horizontal axis represents the gene name, and the vertical axis represents the ka/ks value. In the box plot, the upper and lower endpoints of the vertical lines above and below the rectangle respectively represent the upper and lower edges of the data, the thick line inside the rectangle represents the median, the upper and lower edges of the rectangle represent the upper and lower quartiles, and the dots outside the upper and lower edges of the data represent outliers.

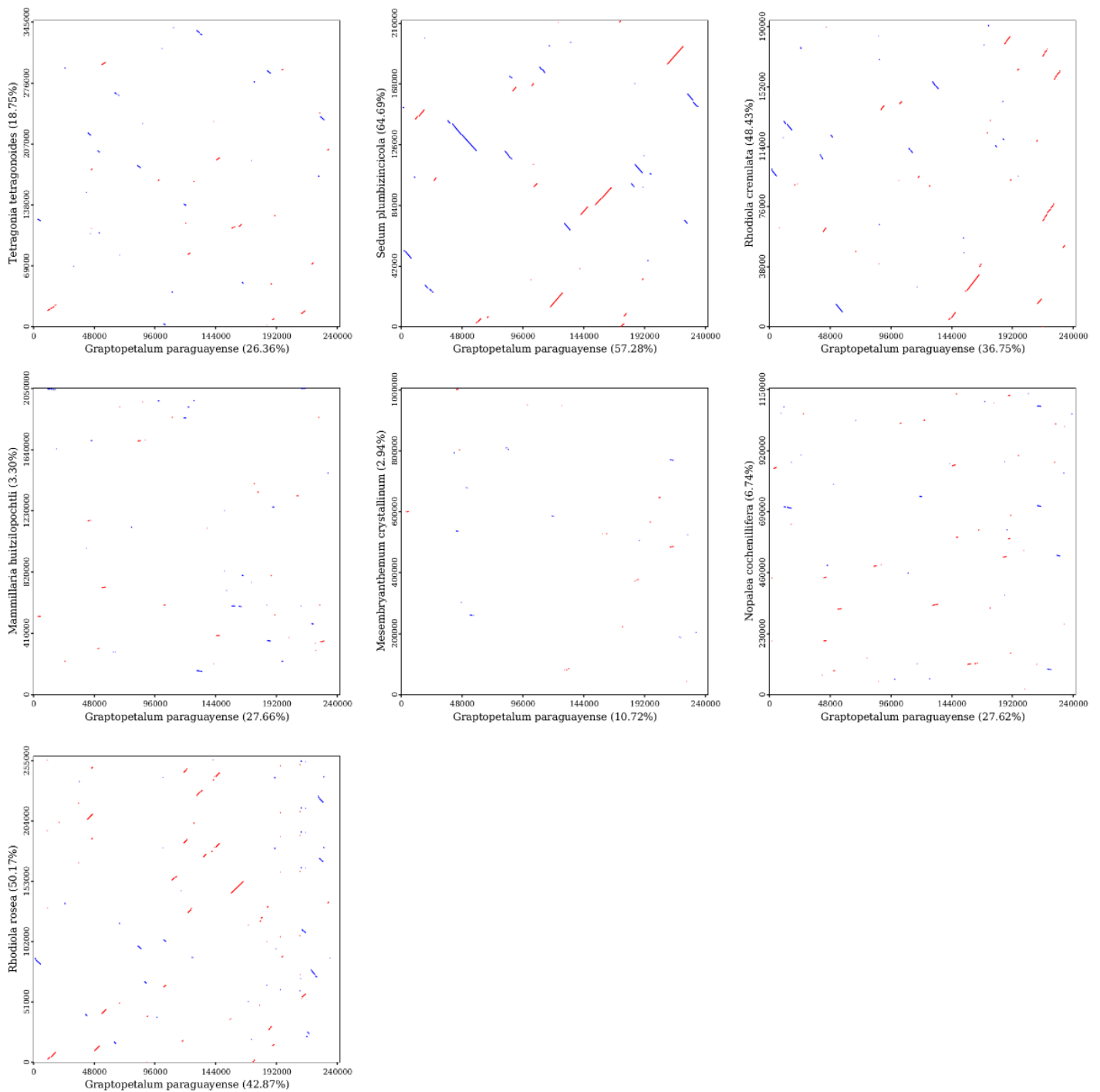


Figure 6

Collinear analysis of mitochondrial sequences. The horizontal axis in each box represents the assembled sequence, and the vertical axis represents other sequences. The values in parentheses indicate the proportion of homologous sequences in the total genome. The red lines in the boxes represent forward alignments, and the blue lines represent reverse complementary alignments.

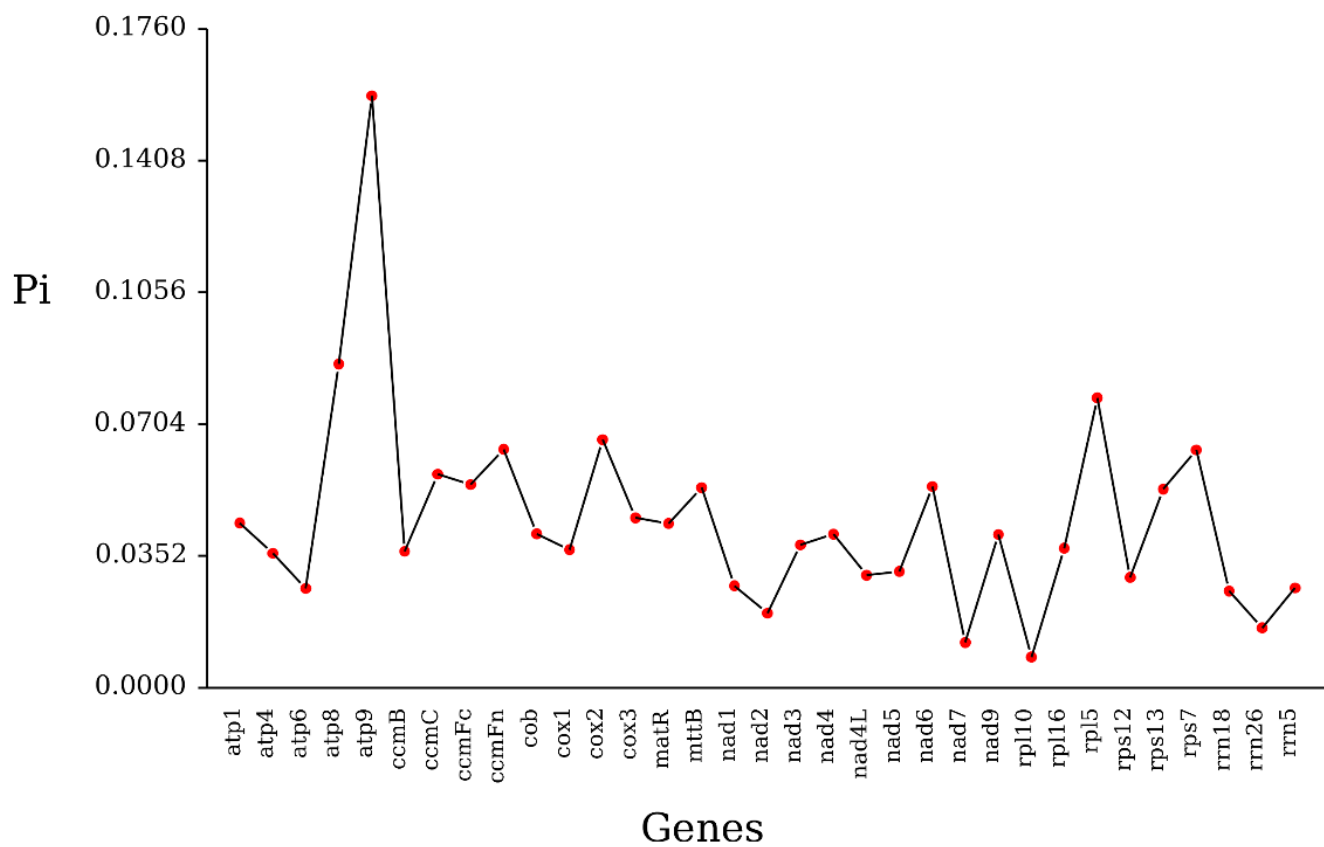


Figure 7

Gene Pi Value Line Chart. The x-axis represents gene names, and the y-axis indicates Pi values.

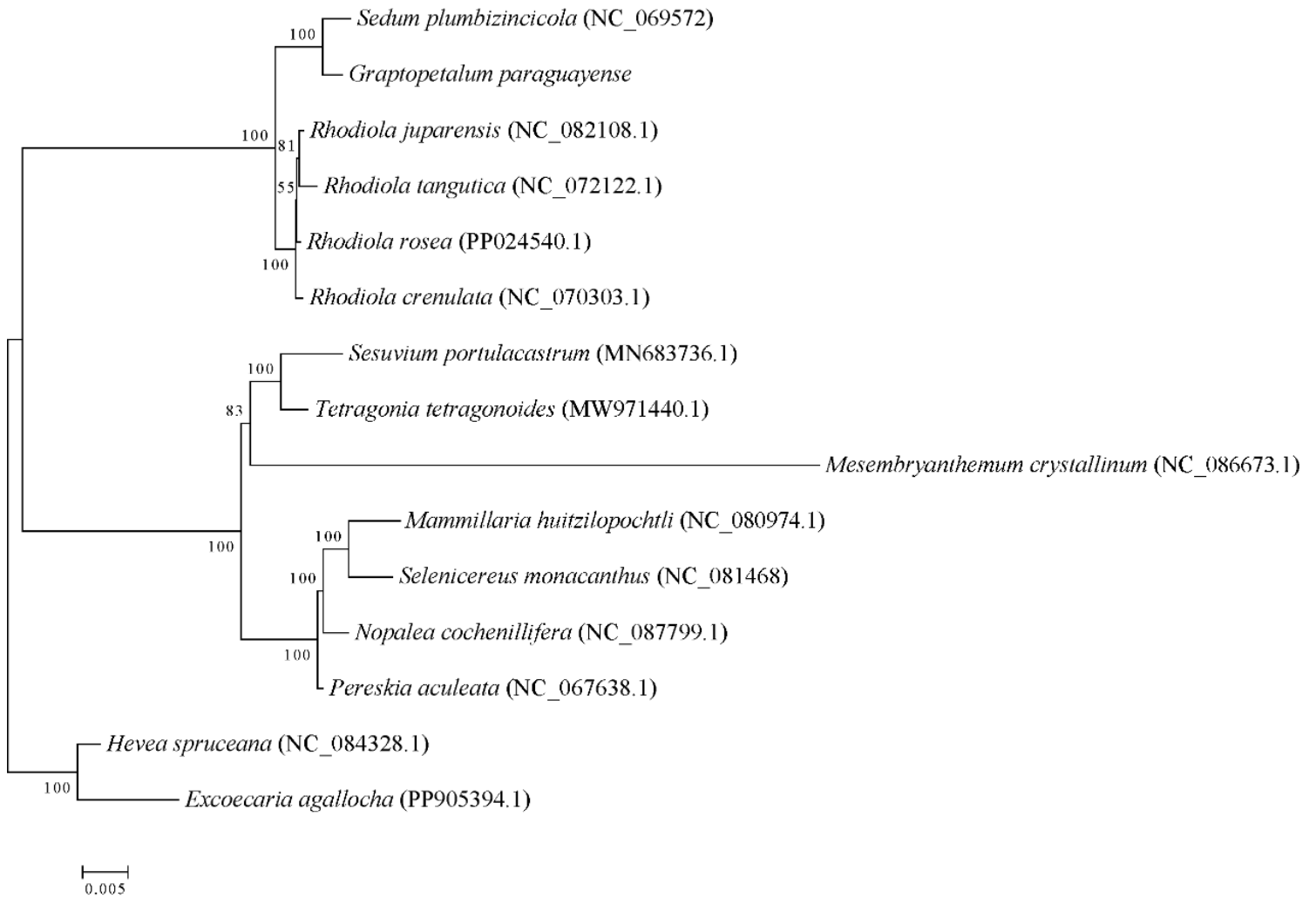


Figure 8

Analysis of mitochondrial system evolution. Evolutionary branch length: Also known as genetic variation or evolutionary distance. It represents the degree of change in evolutionary branches. The shorter it is, the smaller the difference and the closer the evolutionary distance. Distance scale: The unit length of the difference values between organisms or sequences, equivalent to the scale of an evolutionary tree. Bootstrap value: The bootstrap value is generally marked at the node position and is used to evaluate the credibility of the branch.

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable.docx](#)