



OPEN The characterisation of the complete mitochondrial genome of *Polygonatum Kingianum* reveals recombination mediated by repeats associated with DNA replication

Aien Tao^{1,2,4}, Jing Tao^{1,4}, Zhengkun Gan¹, Yue Zhang², Chenglong Yin³ & Guihua Jiang¹✉

Polygonatum kingianum Coll et Hemsl (*P. kingianum*) is recognized as one of the most valuable medicinal species, demonstrating a diverse range of bioactivities, including antidiabetic effects, reduction of blood lipids, anti-tumor properties, and hyperglycemic activity. However, the mitochondrial genome of *P. kingianum* has not yet been reported. Therefore, we sequenced and assembled the mitochondrial DNA (mtDNA) of *P. kingianum*. The results indicate that the mitochondrial genome possesses a multi-branched conformational structure. Further annotation of the mitochondrial genome reveals that it is 647,110 bp in length, containing a total of 35 distinct protein-coding genes, 22 tRNA genes, and 4 rRNA genes. The analysis of gene loss revealed that a total of seven genes were absent in the mitochondrial genome of *P. kingianum*. This includes six ribosomal protein genes (*rpl2*, *rpl6*, *rpl10*, *rps7*, *rps8*, *rps11*) and two succinate dehydrogenase genes (*sdh3* and *sdh4*), as well as one pseudogene. Furthermore, Analysis of mitochondrial genome coding sequences of ten closely related species revealed a total of 24 shared genes. The Ka/Ks values of all shared coding protein genes are less than 1, indicating that these genes have undergone purification selection during evolution and their protein functions are relatively stable. In addition, codon studies on *P. kingianum* mitochondria showed 29 high-frequency codons (RSCU > 1). The preference for codon usage is influenced by mutations and natural selection, but mainly determined by natural selection. Furthermore, we identified 31 homologous fragments spanning both chloroplast and mitochondrial genomes and 585 C-to-U RNA editing sites were predicted in mitochondrial PCGs. The phylogenetic tree established a close relationship between *P. kingianum* and the 27 closely related species. To sum up, this study will contribute to the application of population genetics and evolutionary research in the genus *Polygonatum* and other genera in the Asparagales family.

Keywords *Polygonatum Kingianum*, Mitochondrial genome, Homologous fragments

The *Polygonatum* Mill comprises approximately 40 species distributed throughout the Northern Hemisphere, with the highest concentration found in East Asia. In China alone, there are 31 species of *Polygonatum*^{1,2}. Many of these indigenous plants are economically valuable and have been used for millennia in ethnic and traditional Chinese medicine³. Remarkably, the Chinese Pharmacopoeia Committee (2020) of the People's Republic of China has recorded numerous species of *Polygonatum* in the Compendium of Materia Medica. These species include *P. kingianum*, *Polygonatum sibiricum* Redouté, *P. cyrtoneura* Hua, and *P. odoratum* (Mill.) Druce⁴.

Several ancient Chinese medical writings, including Shen Nong's Classic of the Materia Medica (also known as Shen Nong Ben Cao Jing in Chinese), are the source of *Polygonatum Rhizoma*'s therapeutic properties.

¹State Key Laboratory of Southwestern Chinese Medicine Resources, School of Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China. ²College of Medicine, Lijiang Culture and Tourism College, Lijiang 674100, China. ³College of Pharmacy, Dali University, Dali 671003, China. ⁴Aien Tao and Jing Tao contributed equally to this work. ✉email: 11469413@qq.com

According to *Traditional Chinese medicine* (TCM), *P. Rhizoma* is good for deficiency of spleen and stomach qi, weakness and exhaustion symptoms, low stomach yin, dry mouth and poor appetite, lung weakness with a dry cough, bloody cough from overexertion, back and knee soreness and weakness, early graying of hair and beard, and internal heat with excessive thirst⁵. Rich in flavonoids, polysaccharides, saponins, polyphenols, lignin, and several anthraquinone derivatives, *P. Rhizoma* is mainly known for its pharmacological properties, which include lowering blood cholesterol, promoting antioxidant benefits, delaying the onset of Alzheimer's disease, and preventing aging^{6–8}.

The classification and species identification of the *Polygonatum* genus is a very difficult problem. *Polygonatum Rhizoma* belongs to a multibasic original species, The traits of various groups are crossed, the geographical distribution areas are overlapped, and the morphology is transitional⁹. The relationship of the genus is complex, and the interspecific classification has been controversial, which also leads to confusion in medicinal use and seriously affects the safety and effectiveness of clinical medication^{10–12}. For example, in some parts of China, *P. verticillatum* (L.) All, *P. punctatum* Royle ex Kunth, *P. zanzlanscianense* Pamp, *P. filipes* Merr. ex C. Jeffrey & McEwan, *P. involucreatum* (Franch. & Sav.) Maxim., *P. desoulavyi* Kom, *P. macropodium* Turcz., *P. stenophyllum* Maxim., *P. roseum* (Ledeb.) Kunth, *P. cirrhifolium* (Wall.) Royle, etc. are also used as medicine³. Therefore, *Polygonatum*'s identification and classification are essential.

So far, the field of *Polygonatum*'s identification and classification study has grown significantly in the last few years, mainly focusing on chemical fingerprints, DNA molecular markers, DNA barcoding identification, and chloroplast genome phylogeny, etc^{10,11,13–18}. In contrast, the study of plant mitochondrial genomes is less in *Polygonatum* plants. Currently, only the mitochondrial genome of *P. kingianum* voucher SNX60 (PKV-SNX60) has been released to Genbank (GenBank: PP861176.1). However, no articles on the mitochondrial genome of *P. kingianum* have been published.

Mitochondria are important organelles in plant cells and participate in many metabolic processes related to the production of ATP storage molecules and cytoplasmic male sterility (CMS)^{19,20}. Compared with chloroplast genomes and animal mitochondrial genomes (mtDNA), plant mitochondrial genomes have some unique characteristics, including large differences in genome size and structure, extensive gene horizontal transfer and homologous recombination-mediated by repetitive sequences, increase or deletion of genes or whole chromosomes, high intron density, specific trans-splicing related to different intron types, RNA editing, etc²¹. Moreover, the mitochondrial genome structure is complex. In addition to the single-loop structure, the mitochondrial genome can also exist in linear, polycyclic, branched, and complex forms^{22–24}. Comprehending these characteristics will facilitate an understanding of how mitochondria function and evolve at various taxonomic levels, ranging from order to species.

In this study, we combined Nanopore ONT data with Illumina paired end reads to generate an extensive master graph of *P. kingianum*'s mtDNA. After that, we took a comprehensive picture of the structures and compared their mitochondrial DNA to those of other species. Our analysis of *P. kingianum*'s mtDNA will serve as a valuable resource for upcoming research on molecular diversity.

Materials and methods

Plant material, DNA extraction, and sequencing

In Lijiang, Yunnan Province, China (Longitude 100.243, Latitude 26.889), we collected young, healthy leaves from a farmed *P. kingianum* individual (Fig. 1). To extract complete genomic DNA and RNA at the same time, the leaves were combined and crushed. High-quality genomic DNA was extracted from the leaves using a modified Hexadecyl trimethyl ammonium Bromide (CTAB) technique²³. After building the libraries with SQK-LSK109, 27.14 GB of raw data with an average read size of 6069 bp were obtained from DNA sequencing on the Nanopore PromethION platform (Nanodrop Technologies, Wilmington, DE, US). After that, the data were refined and filtered using NanoFilt and NanoPlot from Nanopack²⁵. A total of 24.67 GB was produced by this procedure from 4.36 million reads, with an average read length of 6 kb. At the same time, libraries were created using the NexteraXT DNA Library Preparation Kit from premium DNA, with an average fragment length of 350 bp. Following that, sequencing was done on the Illumina Novaseq 6000 platform (Illumina, San Diego, CA, USA), yielding 19.76 gigabytes of raw sequence data. Following processing with the NGS QC Tool Kit v2.3.3²⁶, 64.7 million reads yielded 19.43 gigabytes of data.

Assembly and annotation of *P. kingianum*

To align third-generation sequencing data with the reference sequence (*P. kingianum*, PP861176-PP861177), we used minimap2 (v2.15-r905). The aligned sequences were then assembled using miniasm (v0.3-r179) to get the first assembly results. The first assembly results were then refined using nextPolish (v1.3.1, <https://github.com/Nextomics/NextPolish>). Next, the second-generation sequencing data was aligned with these revised results using Bowtie2 (v2.3.5.1). After that, the Illumina data was post-aligned into contigs using Unicycler (v0.4.8), and the visualization was done using Bandage²⁷. In the end, minimap2 was utilized once more to align the Nanopore data with the contig sequences that were generated, incorporate the connection details from Bandage, confirm the connections between contigs, and manually refine and complete the outcomes. Consequently, we were able to obtain *P. kingianum*'s whole mitochondrial genome. Using MFANNOT²⁸ and MITOFY²⁹, the entire mitogenome of *P. kingianum* was annotated. OGDRAW software was utilized to visualize the *P. kingianum* mitogenome map³⁰. *P. kingianum*'s whole assembled mitogenome sequence is available to the public through the NCBI GenBank with accession number PQ365714.

Comparison and analysis of ten Asparagales species MtDNA

Phylosuit software (version 1.2.2)³¹ with default parameters was used to extract the coding sequences (CDs) of ten Asparagales species, including *Dendrobium nobile* (OR413867.1-OR413891.1), *Paphiopedilum micranthum*



Fig. 1. Fresh leaves of *P. kingianum*.

(OP465200-OP465225), *P. kingianum* (voucher SNX60 PP861176-PP861177), *Dracaena cochinchinensis* (PP704687.1), *Gastodia javanica* (OP431959-OP431980.1), *Crocus sativus* (OL804177.1), *Chlorophytum comosum* (MW411187), *Asparagus officinalis* (MW411187), and *Allium fistulosum* (OL347690). The homologous sequences found by the BLAST tool served as the basis for the extraction. The MAFFT software³² was used to align the homologous sequences using its default settings. The homologous sequences that had aligned were cut and then combined. The 24 common genes from the mitochondrial genome were used to build a phylogenetic tree using MEGA software³³. Subsequently, the DnaSP v5 software³⁴ was utilized to haplotype-identify the aligned 24 common gene sequences, and the PopART software³⁵ was employed to construct the TCS haplotype network.

To evaluate the evolutionary selective pressure exerted by ten Asparagales species on particular genes that code for proteins, we categorized the ten Asparagales species pairwise, after homologous gene pairs were extracted, or homologous gene pairs were aligned using the MAFFT (v7.427) software³². The Ka, Ks values for each gene pair were then calculated using the KaKs Calculator (v2.0, <https://sourceforge.net/projects/kakscalculator2/>)³⁶, and the calculation technique chose the MLWL. After counting the K/Ks values for every pair of genes, the R ggplot2 package was used to plot boxplots and bar graphs.

Analysis of codon usage bias, repeat fragments prediction of RNA editing sites

Protein-coding sequences were obtained using the Phylosuit program with the default settings³¹. We used CodonW (v1.4.4, <http://codonw.sourceforge.net>) to calculate relative synonymous codon usage (RSCU) from the protein-coding genes of the mitochondrial genome and to analyze codon usage bias. To determine RSCU values, codon preference analysis of protein-coding genes in the mitochondrial genome is performed. The codon usage bias parameter of the CDS sequence was then obtained by using CodonW and CUSP, together with the effective number of codons (ENC), GC_1 , GC_2 , GC , GC_3 , GC_{all} , and GC_{3s} (the frequency of bases on the synonymous codon at the third position). By employing the GC_3 value as the abscissa and the ENC value as the ordinate in a scatter plot, the ENC plot was created. Formula (1) is used to calculate the theoretical ENC value. GC_3 was used as the abscissa and the theoretical ENC value as the ordinate in the drawing of a standard curve. Formula (2) is frequently used to determine the ratio of the two and obtain the frequency distribution table of the ENC ratio to observe the difference more clearly between theoretical ENC and actual ENC. The following is the formula:

$$ENC = 2 + GC_3 + \frac{29}{[GC_{32} + 2(1 - GC_3)]} \quad (1)$$

$$ENC_{ratio} = \frac{ENC_{theory} - ENC_{practice}}{ENC_{theory}} \quad (2)$$

Scatter plots were obtained with GC_3 as the abscissa and GC_{12} as the ordinate. The $y = x$ function curve was then plotted as a reference line for subsequent analysis.

The SSR analysis was conducted using the web-based application MISA³⁷. Using trf Software (version 409.linux64, parameters: 2778010502000-f-d-m)³⁸, tandem repeats were analyzed. Blastn Software (version 2.10.1, parameters: -word_size 7, evaluate le-5, eliminate redundancy and tandem duplication) was used to evaluate length repeats³⁹. The Circos software and Excel 2010 were used to visualize the map. Furthermore, PmtREP (<http://cloud.genepioneer.com:9929/#/tool/alltool/detail/336>) was used to forecast RNA editing events to obtain an additional understanding of RNA editing sites⁴⁰.

Analysis of SNP and indel of molecular markers in intron regions

Plant mitochondrial intron polymorphisms provide a useful tool for differentiating congeneric taxa. We collected the sequences of 15 intron areas-cox2-intron2, ccmFc-intron, nad2-intron1, nad4-intron1, nad5-intron1, nad5-intron2, nad7-intron2, and nad7-intron3 of ten species in the *Asparagales* to create the molecular markers in the intron regions. Utilizing the DNaSP v5⁴¹ and MEGA X software³³, sequence alignment was carried out, and a customized script was used to identify variations for each intron region. Tables containing all of the variants found inside each intron region can be utilized for downstream analysis and species authentication.

Identification of homologous fragment and collinear analysis

We evaluated homologous sequences between *P. kingianum*'s chloroplast (GenBank: OR995657.1) and mitochondrial organelle using the BLASTN program³⁹ with its default settings. We also linked the amino acid sequence of the plant's mitochondrial genome with that of the chloroplast genome. The Circos software was then used to visualize the obtained homologous segments⁴². We used Mauve (v2.3.1) to align reads against nine mitochondrial genomes and measured the lengths of collinearity blocks based on BLAST alignments to look at the evolution of species in more detail³⁹.

Phylogenetic analyses

We obtained 27 complete mitochondrial genome sequences from four distinct orders: Orchidaceae, Amaryllidaceae, Asparagaceae, and Iridaceae, sourced from the National Center for Biotechnology Information (NCBI) database. The species included *Dendrobium nobile* (OR413867.1-OR413891.1), *Dendrobium officinale* (OR413912-OR413932), *Dendrobium flexicaule* (OR413892.1-OR413911), *Dendrobium huoshanense* (OR413847-OR413866.1), *Paphiopedilum micranthum* (OP465200-OP465225), *Apostasia fujianica* (PP724664.1), *Phalaenopsis mannii* (PQ180326.1-PQ180344.1), *Phalaenopsis lobbii* (PQ180302.1-PQ180319.1), *P. kingianum* (voucher SNX60 PP861176-PP861177), *Dracaena cochinchinensis* (PP704687.1), *Cymbidium ensifolium* (OR754263.1-OR754281.1), *Cymbidium lancifolium* (isolate TEL OQ024442.1-OQ024450.1), *Gastrodia javanica* (OP431959-OP431980.1), *Gastrodia elata* (OP441091.1-OP441125.1), *Gastrodia flexistyla* (OP426364.1-OP426381.1), *Gastrodia crispa* (OP426348.1-OP426363.1), *Gastrodia longistyla* (OP441370.1-OP441338.1), *Cymbidium macrorhizon* (OQ029542.1-OQ029563.1), *Gastrodia peichatieniana* (OP404219.1-OP404234.1), *Gastrodia angusta* (OP414082.1-OP414118.1), *Crocus sativus* (OL804177.1), *Chlorophytum comosum* (MW_411187), *Allium cepa* (OQ_473923), *Asparagus officinalis* (MW411187), *Apostasia shenzhenica* (NC_077647), and *Allium fistulosum* (OL347690). The mitochondrial genomes of *Acorus tatarinowii* (OR066314.1-OR066322.1) were used as outgroups. A total of 28 species were identified, and the mtDNA and cpDNA common protein-coding genes (PCGs) were found to be 35 and 67, respectively. We used the MAFFT program to do a multiple sequence alignment to examine these common gene sequences. After trimming the sequences with trimAl (v1.4.rev15)⁴³ with the setting: -gt 0.7, we compared the sequence junctions. Following this trimming procedure, we identified that the model belonged to the GTR type using the jmodeltest-2.1.10 program⁴⁴. Then, with the GTRGAMMA model and a bootstrap value of 1000, we built a maximum-likelihood evolutionary tree using RAxML v8.2.10⁴⁵.

Results

The structure of *P. kingianum* mitochondrial genome

We sequenced the entire mitochondrial genome of *P. kingianum* and mitochondrial genome graph plots were visualized using the Bandage software²⁷. The results indicate that the mitochondrial DNA (mtDNA) of *P. kingianum* exhibits a complex multibranch conformation (Fig. 2). This structure comprises 20 nodes, each of which forms overlapping regions with adjacent nodes along the connecting lines. Subsequently, we utilized the depth information of the nodes to disentangle the repeat regions, from which we inferred a linear path. Finally, we determined the *P. kingianum* mitochondrial genome as a DNA molecule containing 30 contigs (Fig. 3). The mitochondrial genome was further annotated, revealing a total of 35 distinct protein-coding genes, 22 tRNA genes, and 4 rRNA genes (Table 1). The 36 core genes included five ATP synthase genes (*atp1*, *atp4*, *atp6*, *atp8*, *atp9*), nine NADH dehydrogenase genes (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*), four ubiquinol cytochrome c reductase genes (*ccmB*, *ccmC*, *ccmFc*, and *ccmFN*), three cytochrome c oxidase genes (*cox1*, *cox2*, *cox3*), a transport membrane protein (*mttB*), maturases (*matR*), and a cytochrome c biogenesis gene (*cob*). The non-core genes included two large subunits of the ribosome (*rpl5*, *rpl16*), nine small subunits of the ribosome (*rps1*, *rps10*, *rps12*, *rps13*, *rps14*, *rps19*, *rps2*, *rps3*, *rps4*), and one succinate dehydrogenase gene (*sdh4*).

Comparison and analysis of ten *Asparagales* species MtDNA

To further explore the evolutionary characteristics of the mitochondrial genome of *P. kingianum*. We compared it with nine other *Asparagales* species (*D. nobile*, *P. micranthum*, *P. kingianum*, *D. cochinchinensis*, *G. javanica*, *C. sativus*, *C. comosum*, *A. officinalis*, and *A. fistulosum*). The total gene content in the coding regions of the mitochondrial genomes of these ten *Asparagales* species ranged from 5.6 to 23.26%. AT, and GC content of these mitochondrial genomes fell within the ranges of 50.39–55.57%, and 42.13–49.61%. Compare the discovery that AT content was higher than GC content in ten *Asparagales* species mtDNA (Table 2). Results of the gene loss

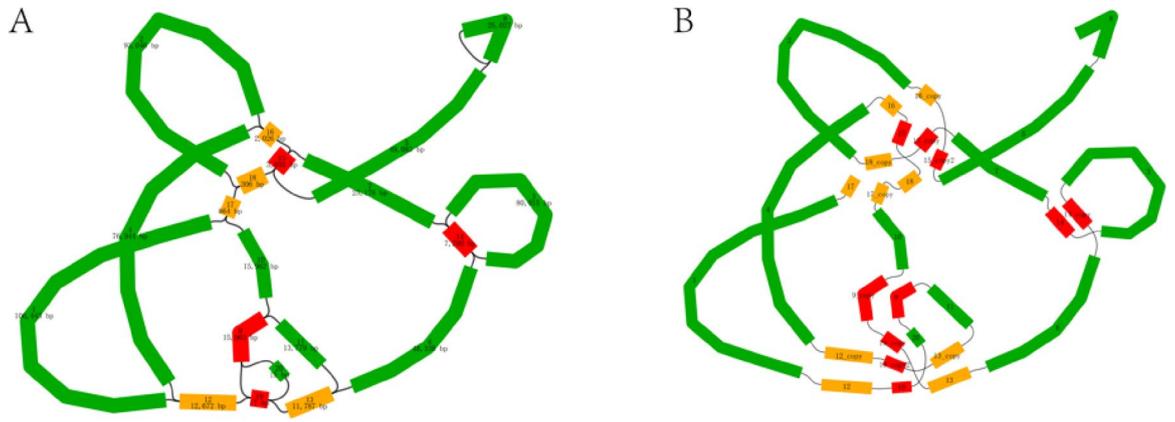


Fig. 2. The assembly result of the mitochondrial genome of *P. kingianum*. Note: (A) is a long-read, long-data-supported graph diagram presented by the Bandage software, where red and yellow nodes represent duplicated segments, with a total of 20 nodes; (B) shows a possible assembly path that we inferred, ultimately resulting in the complete mitochondrial genome of *P. kingianum*, assembled into a linear structure.

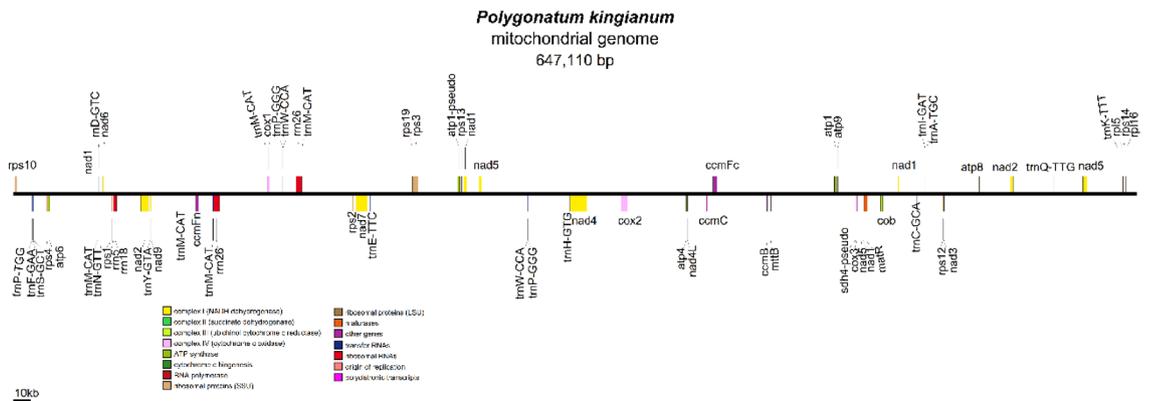


Fig. 3. The map of the mitochondrial genome of *P. kingianum*. Note: The upper genes in the horizontal line represent the transcription direction counterclockwise, while the lower genes are the opposite. Different functional genes with different faces Color logo.

Group of genes	Gene name
ATP synthase	#atp1 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 membrane protein	mttB
NADH dehydrogenase	nad1**** nad2**** nad3 nad4*** nad4L nad5**** nad6 nad7**** nad9
Ribosomal proteins (LSU)	rpl16 rpl5
Ribosomal proteins (SSU)	rps1 rps10* rps12 rps13 rps14 rps19 rps2 rps3* rps4
Succinate dehydrogenase	#sdh4
Ribosomal RNAs	rrn18 rrn26(2) rrn5
Transfer RNAs	trnA-TGC* trnC-GCA trnD-GTC trnE-TTC trnF-GAA trnH-GTG trnI-GAT* trnK-TTT trnM-CAT(5) trnN-GTT trnP-GGG(2) trnP-TGG trnQ-TTG trnS-GCT trnW-CCA(2) trnY-GTA

Table 1. The protein-coding genes classification of the mitochondrial genome of *P. Kingianum*. Notes: *:intron number;#Gene: Pseudo gene; Gene(2):Number of copies of multi-copy genes.

Organism	genome length	Total gene content%	No of PCGs	No of rRNAs	No of tRNAs	No of introns	AT%	GC%
<i>A.officinalis</i>	492,062	14.15	36	6	17	23	54.1	45.9
<i>C. comosum</i>	950,316	6.83	36	5	21	21	53.35	46.65
<i>A. fistulosum</i>	382,053	23.26	25	4	12	22	55.02	44.98
<i>C. sativus</i>	525,348	11.34	34	4	14	23	53.16	46.84
<i>G.javanica</i>	1,294,379	5.6	38	3	24	27	54.86	45.14
<i>Pmicranthum</i>	447,368	23.19	67	3	16	27	55.57	44.43
<i>D. nobile</i>	796,523	8.23	38	3	27	24	54.85	42.13
<i>Pkingianum</i>	647,110	10.94	35	4	22	26	53.8	46.2
<i>D.cochinchinensis</i>	525,335	12.24	35	3	24	23	50.39	49.61
<i>Pkingianum voucher SNX60</i>	639,831	10.5	35	3	17	26	53.54	46.46

Table 2. General characters of ten Asparagales species MtDNA.

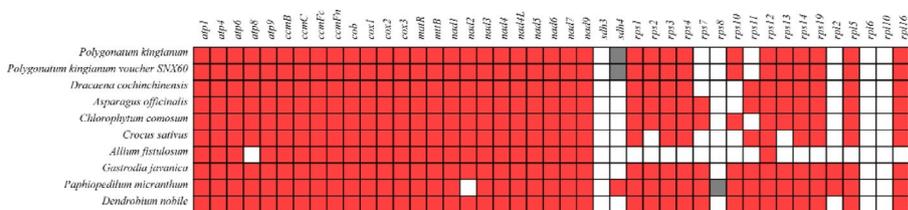


Fig. 4. Variable gene contents in the ten Asparagales species nine released Apiales mitochondrial genomes. The corresponding accession numbers and the related references are shown in Table 1. Note: Each column represents a variable gene, while each row represents a species. The mitochondrial genome described in this study is labeled with ‘.’.

analysis show most coding sequences (PCGs) of ten Asparagales species mtDNA are conserved, especially the respiratory chain-related genes, maturation enzymes, and methyltransferase genes. In contrast, the ribosomal proteins and succinate dehydrogenase genes were more variable. The genes of *rpl2*, *rpl6*, *rpl10*, *rps7*, *rps8*, *rps10*, *rps11*, *sdh3*, and *sdh4* have been lost in the mt genome of most species. Furthermore, compared to *P. kingianum*, ATP8 in *A. officinalis* and NAD2 in *G. javanica* has also been lost. Similarly, A total of seven genes were lost in the mitochondrial genome of *P. kingianum*. Including 6 ribosomal protein genes (*rpl2*, *rpl6*, *rpl10*, *rps7*, *rps8*, *rps11*) and 1 succinate dehydrogenase gene (*sdh3*) (Fig. 4). This result is consistent with the already published data (GenBank: PP861176.1).

To further analyze these mitochondrial genomes, we extracted the protein-coding genes (PCGs). We identified 24 shared genes (*atp1*, *atp4*, *atp6*, *atp9*, *ccmB*, *ccmC*, *ccmFc*, *ccmFn*, *cob*, *cox1*, *cox2*, *cox3*, *matR*, *mttB*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*, and *rps12*) and used them to construct a phylogenetic tree (Fig. 5). Simultaneously selecting *D. nobile* (OR413867.1), *G. javanica* (OP431959.1), and *P. micranthum* (OP465200.1) as outgroups, the clustering relationships among the seven outgroups, as shown in Fig. 16, are completely consistent, with the exception of variations in the support values of individual branches. The results from the two tree-building methods are consistent, indicating the reliability of the findings. Obviously, our results indicated that the mitochondrial genome of *P. kingianum*, along with previously published data for the same species, is clustered into a single clade. Following this, *D. cochinchinensis* was found to be closely related to *P. kingianum*.

To delve deeper into genetic variation, we analyzed the mutation sites of 24 genes from the mitochondrial DNA (mtDNA) of ten Asparagales species and constructed a haplotype network (Fig. 6). The results indicate that for the *atp1*, *matR*, *nad3*, and *nad4* genes, each of the ten species exhibited one unique haplotype. In contrast, for the *atp4*, *atp6*, *ccmB*, *ccmC*, *ccmFc*, *ccmFn*, *cob*, *cox1*, *cox2*, *cox3*, *mttB*, *nad1*, *nad2*, *nad5*, *nad6*, and *nad7* genes, *P. kingianum* had one haplotype, while each of the remaining eight species possessed one haplotype. Regarding the *atp9*, *nad9*, and *rps12* genes, both *P. kingianum* and *D. cochinchinensis* shared one haplotype, with the other seven species each exhibiting one haplotype. Finally, in the *nad4L* gene, *P. kingianum* and *A. officinalis* each had one haplotype, while the remaining seven species also had one haplotype each.

Our analysis aimed to assess the impact of environmental stress on mitochondrial genome evolution. Using the 24 shared protein-coding genes among the ten Asparagales species mtDNA, we calculated the nonsynonymous/synonymous mutation ratio (Ka/Ks) to observe the evolutionary effects of the mitochondrial genome under environmental (Fig. 7). Importantly, the Ka/Ks values of 24 PCGs < 1, indicating negative selection. This result indicates that all 24 protein-coding genes have undergone purifying selection during evolution and have relatively stable protein functions.

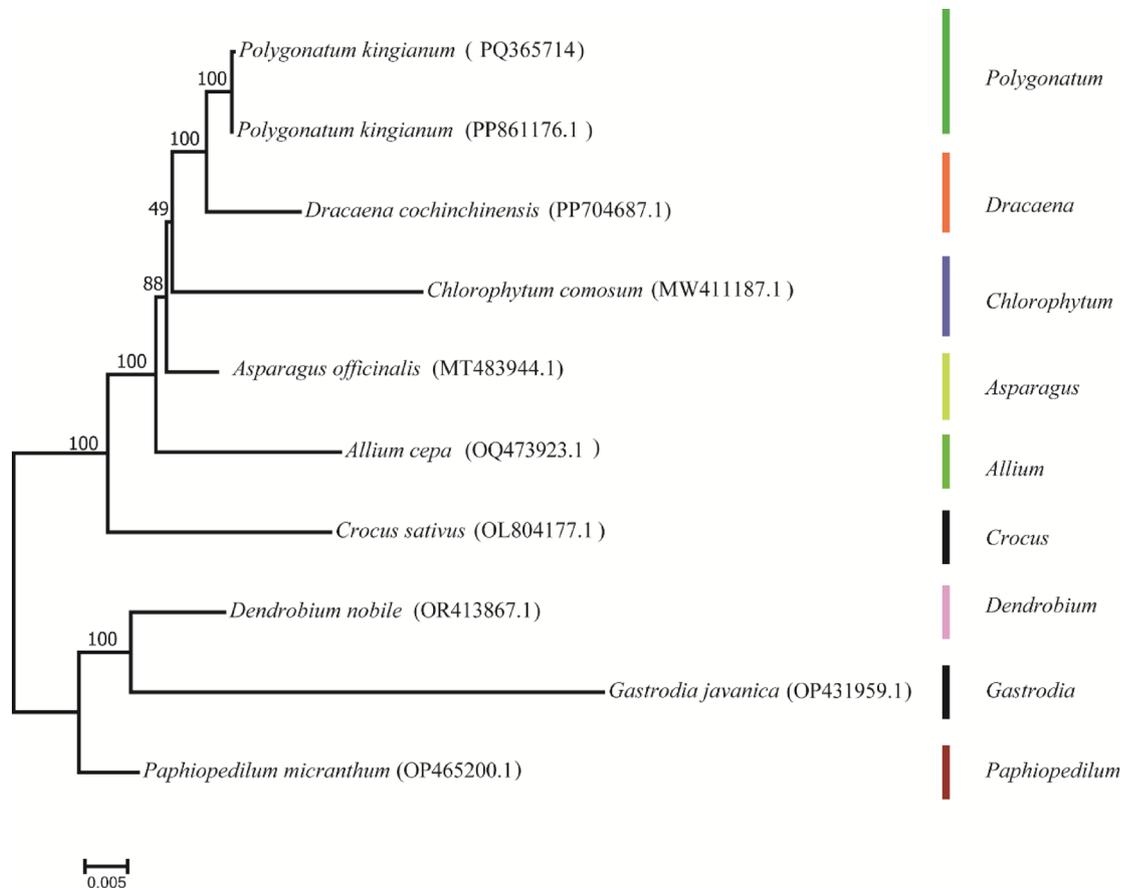


Fig. 5. Shows a phylogenetic tree based on 24 shared genes from 10 asparagine species.

Analysis of codon usage bias

In the complete mitochondrial genome of *P. kingianum*, we identified a total of 10,246 codons within the protein-coding genes (Table 3). *P. kingianum*'s mitochondrial DNA encodes all 21 amino acids, and we found 61 different types of amino acid codons. The major codon families are Leu (1,056 codons), Ser (979 codons), Pro (809 codons), Ile (797 codons), Arg (726 codons), Gly (716 codons), Ala (647 codons), Phe (644 codons), Val (627 codons), Thr (526 codons). Ter showed the highest usage bias. Ter (5 codons) and Ter (9 codons) showed the least number of codons. The codon preference of the mitogenomes is summarized in Fig. 8. We observed that 29 codons were used more frequently than expected, indicated by relative synonymous codon usage (RSCU) values > 1 , and 33 codons were used less frequently than expected (RSCU < 1). Interestingly, Met and Trp showed no codon preference (RSCU = 1). In terms of amino acids, aside from AUG (Met) and UGG (Trp), the majority of amino acids exhibited a bias in their codon usage pattern (Fig. 8). Most amino acids have at least two different codons, while Leu, Arg, and Ser have six codons each.

We extracted 35 PCGs (atp1, atp4, atp6, atp8, atp9, ccmB, ccmC, ccmFc, ccmFn, cob, cox1, cox2, cox3, matR, mttB, nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7, nad9, rpl16, rpl5, rps1, rps10, rps12, rps13, rps14, rps19, rps2, rps3, rps4) from the *P. kingianum* mitochondrial DNA. The GC content was calculated for the first (GC_1), second (GC_2), and third (GC_3) positions of these genes, resulting in values ranging from 38.06 to 57.45% for GC_1 , the mean value was 47.49%, 37.13–58.39% for GC_2 , the mean value was 43.63%, and 26.73–56.62% for GC_3 , the mean value was 37.4%. The average GC content of these different positions (GC_1 , GC_2 , and GC_3) was found to be $< 50\%$ (Table 3). This suggests a preference for A/T bases and A/T-ending codons in the *P. kingianum* mitochondrial DNA.

The main causes of codon preference are mutation pressure and natural selection. To explore the effects of mutation pressure and natural selection on codon preference, we also calculated the effective number of codons (ENC) for the 35 PCGs, which ranged from 38.9 to 59.85% (Table 4).

ENC was found by C- GC_3 analysis that the 35 PCGs of *P. kingianum* mitochondrial DNA are widely distributed, with most genes distributed below the standard curve and far away from the standard curve, indicating that the codon preference of most genes is mainly affected by natural selection. However, a few genes lie immediately above the standard curve, indicating that the codon preference of this part of genes is mainly affected by mutations (Fig. 9). Furthermore, in a neutrality plot analysis, we observed a correlation of 0.073 between GC_{12} and GC_3 in *P. kingianum* mitochondrial DNA. This indicates that the correlation between GC_3 and GC_{12} in the mitochondrial genome of *P. kingianum* plants is weak, and the codon usage preference is mainly influenced by natural selection (Fig. 10). In summary, natural selection pressure is the primary factor influencing

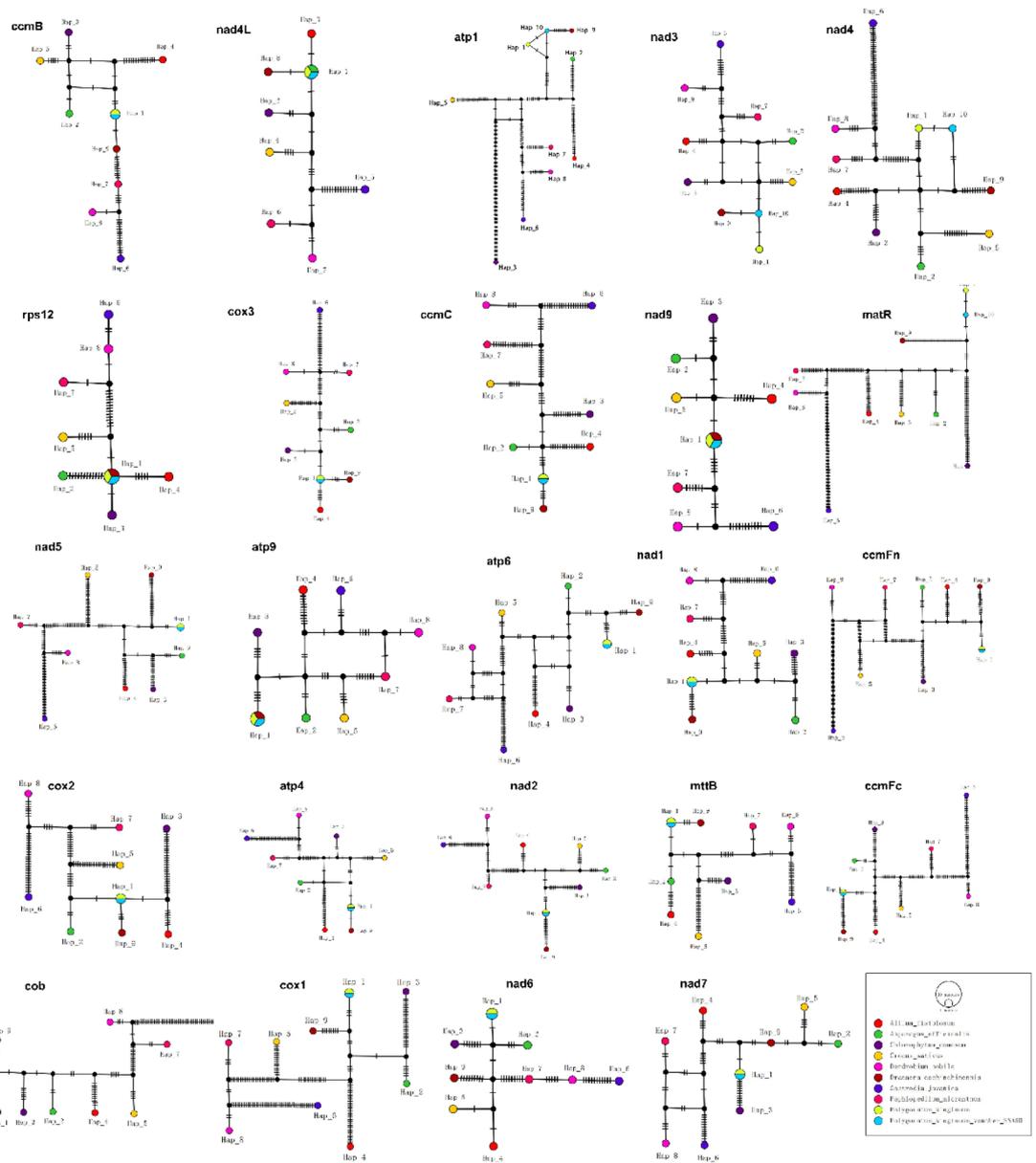


Fig. 6. Haplotype networks of the 10 species. Note: In the haplotype network diagram, each colored circle represents one haplotype, the size of the circle represents the total amount of samples contained in the haplotype, the connection between the two circles represents two haplotypes related to each other, and the short line on the line represents the number of base substitutions needed to change from one haplotype to another haplotype. In addition, each black dot indicates the inferred haplotype of the possible missing sample.

the codon usage bias in *P. kingianum* mitochondrial DNA. This conclusion is consistent with the study of Le Wang et al.²³.

Analysis of repeat fragments

Repetitive sequences play essential roles in the evolution of the plant mitochondrial genomes⁴⁶. The SSRs consist of a short sequence motif that is between 1 bp and 6 bp long and is arranged in tandem³⁷. A total of 199 SSRs were identified in the mitochondrial genome of *P. kingianum* (Table 5). The tetra-nucleotide repeat sequences were the most abundant with 64 loci (32.16%), compared with 50 (25.12%), 50 (25.12%), 27 (13.57%), 6 (3.02%), and 2 (1.00%) of mono, di, tri, penta, and hexa nucleotide repeats (Table 5), respectively. A total of 80 to 337 SSRs were found in the nine mitochondrial genomes of Asparagales species (*D. nobile*, *P. micranthum*, PKV-SNX60, *D. cochinchinensis*, *G. javanica*, *C. sativus*, *C. comosum*, *A. officinalis* and *A. fistulosum*) (Fig. 11; Table 5). Comparative analysis revealed significant differences between the SSRs of *P. kingianum* and the other nine species, especially comparing our mitochondrial genome and GenBank: PP861176.1, our mitochondrial genome over 10 SSRs.

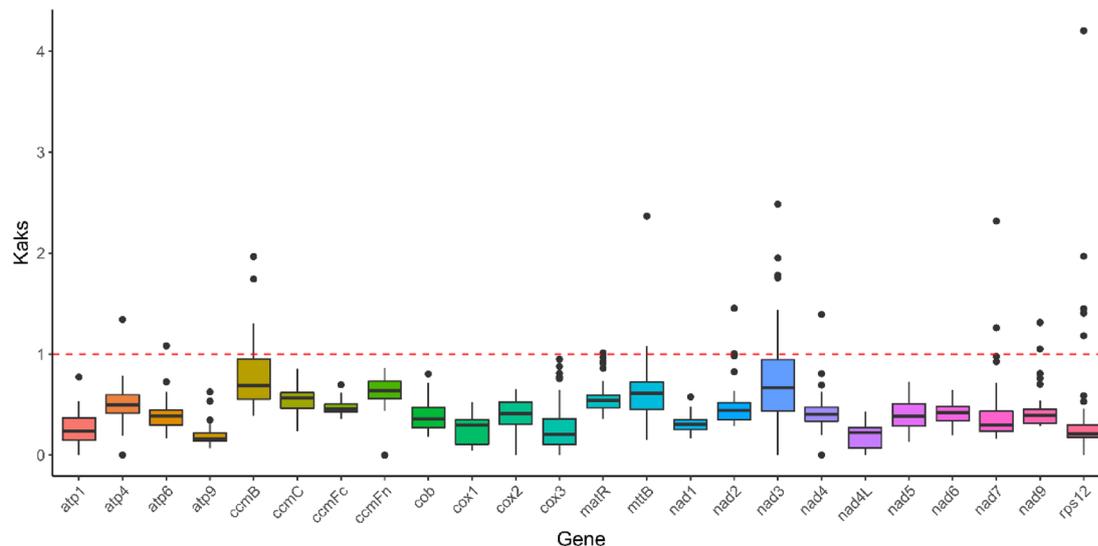


Fig. 7. Ka/Ks ratio calculated for the 24 shared genes from the mitochondrial genome of ten Asparagales species. Note: Different species were depicted using different colors.

AminoAcid	Symbol	Codon	No.	RSCU	AminoAcid	Symbol	Codon	No.	RSCU
*	Ter	UAA	19	1.7273	M	Met	AUG	278	1
*	Ter	UAG	5	0.4545	N	Asn	AAC	107	0.6751
*	Ter	UGA	9	0.8182	N	Asn	AAU	210	1.3249
A	Ala	GCA	156	0.9645	P	Pro	CCA	172	1.1563
A	Ala	GCC	155	0.9583	P	Pro	CCC	127	0.8538
A	Ala	GCG	86	0.5317	P	Pro	CCG	94	0.6319
A	Ala	GCU	250	1.5456	P	Pro	CCU	202	1.358
C	Cys	UGC	53	0.7626	Q	Gln	CAA	214	1.5507
C	Cys	UGU	86	1.2374	Q	Gln	CAG	62	0.4493
D	Asp	GAC	102	0.6182	R	Arg	AGA	164	1.3554
D	Asp	GAU	228	1.3818	R	Arg	AGG	83	0.686
E	Glu	GAA	281	1.3192	R	Arg	CGA	160	1.3223
E	Glu	GAG	145	0.6808	R	Arg	CGC	74	0.6116
F	Phe	UUC	277	0.8602	R	Arg	CGG	93	0.7686
F	Phe	UUU	367	1.1398	R	Arg	CGU	152	1.2562
G	Gly	GGA	257	1.4358	S	Ser	AGC	95	0.5822
G	Gly	GGC	101	0.5642	S	Ser	AGU	163	0.999
G	Gly	GGG	121	0.676	S	Ser	UCA	190	1.1645
G	Gly	GGU	237	1.324	S	Ser	UCC	163	0.999
H	His	CAC	61	0.4674	S	Ser	UCG	142	0.8703
H	His	CAU	200	1.5326	S	Ser	UCU	226	1.3851
I	Ile	AUA	223	0.8394	T	Thr	ACA	139	1.057
I	Ile	AUC	221	0.8319	T	Thr	ACC	138	1.0494
I	Ile	AUU	353	1.3287	T	Thr	ACG	80	0.6084
K	Lys	AAA	249	1.1422	T	Thr	ACU	169	1.2852
K	Lys	AAG	187	0.8578	V	Val	GUA	182	1.1611
L	Leu	CUA	152	0.8636	V	Val	GUC	127	0.8102
L	Leu	CUC	116	0.6591	V	Val	GUG	139	0.8868
L	Leu	CUG	103	0.5852	V	Val	GUU	179	1.1419
L	Leu	CUU	228	1.2955	W	Trp	UGG	138	1
L	Leu	UUA	256	1.4545	Y	Tyr	UAC	70	0.4682
L	Leu	UUG	201	1.142	Y	Tyr	UAU	229	1.5318

Table 3. Relative synonymous codon usage for each amino acid of the mitochondrial genome of *P. Kingianum*.

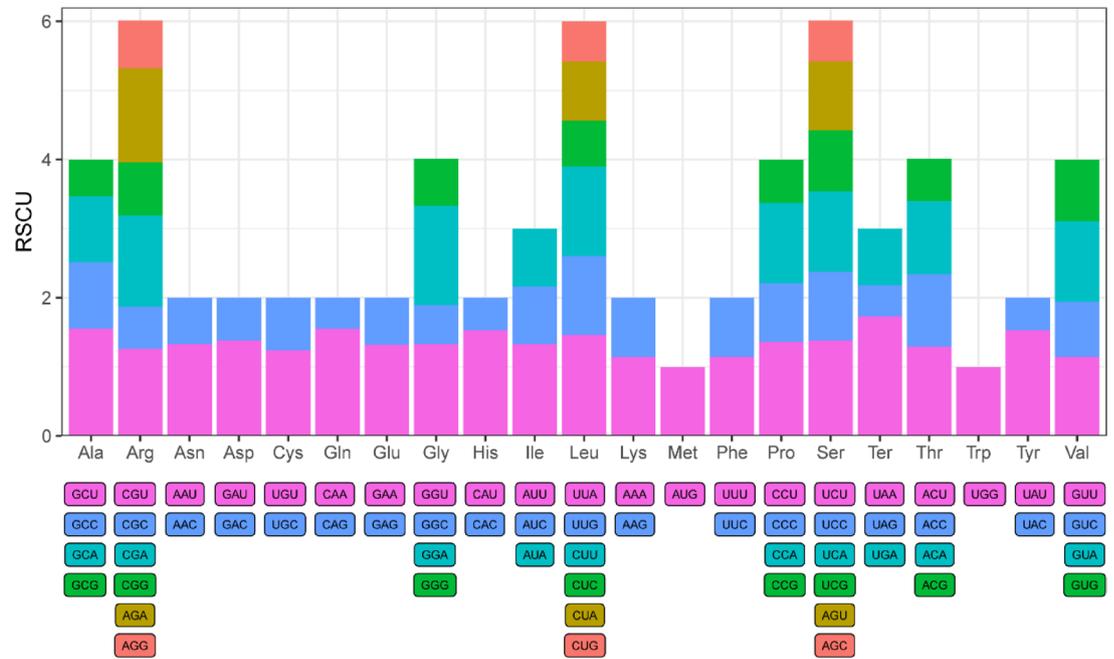


Fig. 8. Analysis of relative synonymous codon usage from *P. kingianum* mtDNA. Note: The codon content for the 21 amino acids and stop codons in all protein-coding genes present in the chloroplast genome of mitochondrial genome of *P. kingianum*.

Tandem repeats are composed of two or more contiguous, approximate copies of a pattern of nucleotides and are caused by the tandem duplication events in the genome sequences³⁸. We detected 31 tandem repeat sequences in the mitochondrial genome of *P. kingianum*. We identified tandem repeats in the nine mitochondrial genomes of Asparagales species. The details are shown in Table 6; Fig. 12. The number of tandem repeats was highly variable, ranging from 13 to 400. When comparing our mitochondrial genome to GenBank entry PP861176.1, we identified over 10 tandem repeat sequences in our mitochondrial genome.

In addition, we found 310 pairs of long repeats with a length greater than or equal to 30 in the mitochondrial genome of *P. kingianum*. The dispersed repeat sequence lengths 40–49 bp and 50–59 bp were the most abundant with 67 pairs (21.61%). In contrast, the dispersed repeat sequence lengths 30–39 bp with 47, pair, lengths 60–69 bp with 23 pair, lengths 70–79 bp with 33 pair, lengths 80–89 bp with 27 pair, lengths 90–99 bp with 10 pair, lengths 100–199 bp with 17 pair, lengths 200–999 bp with 12 pair, and lengths $\geq 1,000$ bp with 7 pair, respectively. Furthermore, we identified dispersed repeats in the nine mitochondrial genomes of Asparagales species. The details are shown in Fig. 13; Table 6. The numbers of dispersed repeats were highly variable and ranged from 146 to 43,220 pairs (Table 7), and there are significant differences. Compare our mitochondrial genome and GenBank: PP861176.1, our mitochondrial genome over 13 pair dispersed repeats sequences. These differential SSRs, tandem repeats sequences, and dispersed repeats sequences between *P. kingianum* and closely related species can be used as candidate sequences for species identification and molecular breeding.

Analysis of RNA editing sites

RNA editing is when RNA transcripts are processed by nucleotide insertions, deletions, or substitutions. It alters the nucleotides of mRNAs, tRNAs, rRNAs, and even miRNAs, ncRNAs, and retrotransposons, leading to high transcriptome diversities⁴⁷. In this research, we aimed to predict RNA editing sites in 35 distinct protein-coding genes (PCGs) within the mitochondrial genome of *P. kingianum* to gain a deeper understanding of gene expression within the mitochondrial genome. A total of 585 RNA editing events were identified at the editing sites, primarily involving the conversion of the nucleotide C to T (U) (Fig. 14). Notably, the *nad4* gene exhibited the highest number of RNA editing occurrences among all mitochondrial genes, with 52 RNA editing sites identified. The *ccmFn* gene had 42 RNA editing sites, and the *ccmB* and *nad2* gene each had 35 RNA editing sites. In contrast, a total of 501 to 613 RNA editing sites were found in the nine mitochondrial genomes of Asparagales species (PKV-SNX60 (GenBank: PP861176.1), *A. fistulosum*, *A. officinalis*, *C. comosum*, *C. sativus*, *D. nobile*, *D. cochinchinensis*, *G. javanica*, and *P. micranthum*). The details are shown in Table 8.

In addition, among the 585 sites, all sites were non-synonymous sites. Among the non-synonymous sites, there were 15 types of effects on the amino acid. The three most abundant amino acid alterations were Ser to Leu, Pro to Leu, and Ser to Phe corresponding to 126 (21.54%), 122 (20.85%), and 97 (16.58%) RNA editing sites, respectively. A total of 279 (47.69%) RNA editing sites changed hydrophilic amino acids to hydrophobic amino acids. And we found two RNA editing sites located in the CDS of *atp6*, *atp9*, *ccmFc*, and *rps10* that created a stop codon. *ccmFc*-1321 created a stop codon for the *ccmFc* that was the final amino acids downstream from the start codon of *ccmFc* (Table 9).

Genes	codon No.	GC1	GC2	GC3	GC_all	ENC	GC3 s
atp1	510	57.45	42.75	35.1	45.1	51.86	35.4
atp4	196	44.9	42.86	38.27	42.01	59.85	38.8
atp6	307	49.19	37.13	36.16	40.83	50.81	36.3
atp8	155	38.06	37.42	43.23	39.57	50.23	44.1
atp9	75	50.67	49.33	38.67	46.22	39.61	38.7
ccmB	207	45.89	43.48	36.23	41.87	49.82	36.4
ccmC	249	46.59	48.19	36.95	43.91	50.59	37.1
ccmFc	451	49	44.12	42.79	45.31	55.17	43.1
ccmFn	596	50.84	50	41.95	47.6	54.8	42
cob	394	50.51	41.62	34.26	42.13	53.54	34.1
cox1	528	47.73	42.8	36.74	42.42	52.92	36.9
cox2	261	52.11	39.08	32.95	41.38	47.01	32.7
cox3	266	52.63	45.11	34.96	44.24	52.16	34.7
matR	657	54.64	45.81	56.62	52.36	53.52	56.8
mttB	116	44.83	40.52	46.55	43.97	51.25	45.5
nad1	326	53.99	43.25	30.98	42.74	51.53	31.2
nad2	489	42.74	44.38	35.38	40.83	50.92	35.2
nad3	119	43.7	47.9	37.82	43.14	45.42	37.9
nad4	496	46.37	43.75	36.29	42.14	53.81	37
nad4L	101	39.6	40.59	26.73	35.64	40.04	27.3
nad5	671	44.11	44.71	35.92	41.58	51.19	36.1
nad6	260	47.31	43.08	45.38	45.26	57.91	45.2
nad7	417	55.4	46.28	32.37	44.68	47.02	32.5
nad9	191	51.83	42.41	33.51	42.58	52.99	34.2
rpl16	137	48.91	58.39	29.93	45.74	53.86	30.7
rpl5	189	48.15	38.1	43.39	43.21	55.85	44.3
rps1	203	44.83	42.36	40.89	42.69	47.92	40.6
rps10	128	42.19	38.28	39.84	40.1	59.32	40.3
rps12	126	55.56	48.41	31.75	45.24	54.36	33.6
rps13	117	50.43	41.03	28.21	39.89	38.9	27.4
rps14	101	44.55	46.53	38.61	43.23	59.18	38.5
rps19	91	38.46	46.15	32.97	39.19	42.91	34.1
rps2	222	42.34	38.74	35.59	38.89	53.96	35.6
rps3	549	45.36	40.8	44.81	43.66	58.09	45
rps4	345	41.45	41.74	37.1	40.1	54.2	37.9

Table 4. Codon usage bias parameter information for CDS sequence of *P. kingianum* MtDNA.

Intracellular gene transfer from Chloroplast to mitochondrial organelles

In the evolutionary process of higher plants, the transfer of genetic material within cells is a commonly observed phenomenon in mitochondrial genomes. However, it is worth noting that these sequence fragments originating from the chloroplast organelle exhibit relatively lower conservation⁴⁸. Therefore, we analyzed to investigate the migration of sequences from the chloroplast to the mitochondrial organelles in *P. kingianum* (Fig. 15). Through sequence similarity analysis, 31 homologous fragments shared were identified between the chloroplast and mitochondrial genomes, with alignment lengths ranging from 38 to 4,363 bp (Table 10). Collectively, these homologous fragments totaled 9,853 bp, accounting for approximately 1.52% of the *P. kingianum* mitochondrial DNA. Including seven tRNA genes (trnA-UGC, trnI-CAU, trnM-CAU, trnN-GUU, trnP-UGG, trnH-GUG, trnW-CCA) and one rRNA gene (rrn16). Alternatively, some small fragment sequences from chloroplasts are subsets of larger fragment sequences or multiple repeatedly in mitochondrial DNA, suggesting that these fragments may have undergone multiple independent transfer integration, replication, and recombination within the mitochondrial DNA after transfer integration.

Analysis of collinearity among *P. kingianum* and other nine species

To assess the relationship between the mitochondrial genome of *P. kingianum* and nine other Asparagales species. We applied BLAST to obtain the results of pairwise comparisons of individual mitochondrial genomes and map Multiple Synteny Plot using homologous sequences as conserved collinear blocks (Fig. 16). The results show that *P. kingianum* and other species possess many orthologous syntenic genome blocks. Compare our results and the NCBI has already published data of PKV-SNX60 (GenBank: PP861176.1), the collinear blocks totaled 579,524 bp, representing 89.56% of the length of our mitochondrial genome. In contrast, the

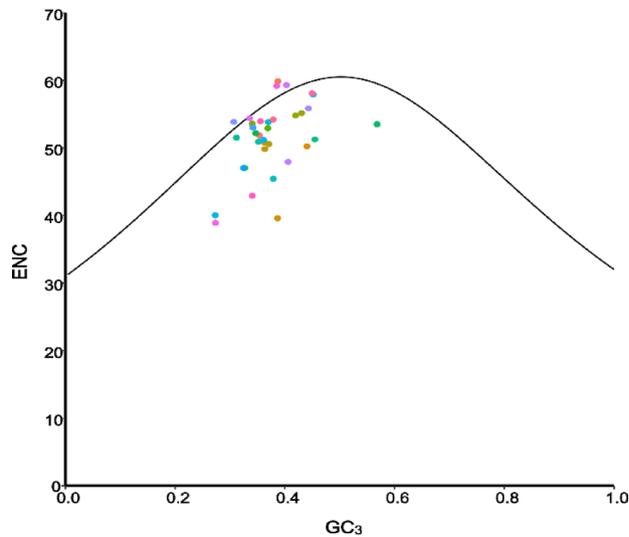


Fig. 9. ENC-plot analysis. Note: The closer the ENC value is to the theoretical curve, the more GC₃ s influence in determining codon usage, and otherwise natural selection affects the codon usage bias. The closer the ENC ratio is to zero, the more the gene codons are affected by base mutations.

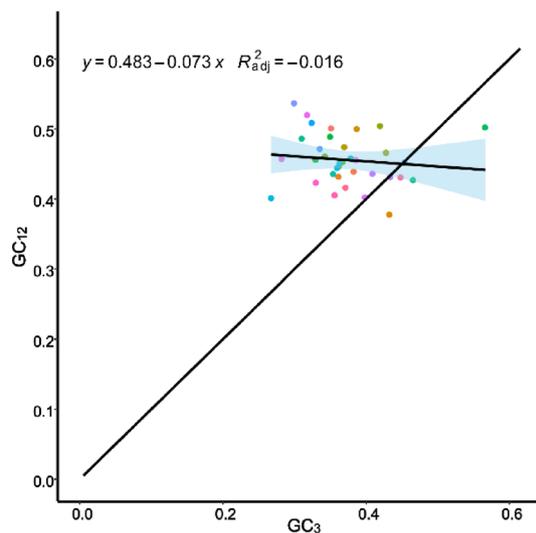


Fig. 10. Analysis of neutrality plot. Using GC₃ as the abscissa and GC₁₂ as the ordinate. The $y = x$ function curve was then plotted as a reference line for subsequent analysis. The stronger the correlation between GC₁₂ and GC₃, the more codon usage is affected by base mutations, and otherwise, the more affected by selection pressure.

mitochondrial genomes of *P. kingianum* showed poor collinearity with other eight species, and the order of the collinear blocks was inconsistent among the individual mitochondrial genomes, implying that between *P. kingianum* and other species (*D. nobile*, *P. micranthum*, *D. cochinchinensis*, *G. javanica*, *C. sativus*, *C. comosum*, *A. officinalis* and *A. fistulosum*), evolutionarily may have undergone substantial genome rearrangements leading to extreme structural unconservation.

Phylogenetic analysis

To determine the phylogenetic placement of *P. kingianum*, we meticulously chose species that are closely related and retrieved mitochondrial DNA (mtDNA) and chloroplast DNA (cpDNA) sequences for 28 species from NCBI. Through comparative analysis, we discovered 35 common protein-coding genes (PCGs) in the mtDNA sequences and 67 in the cpDNA sequences. Using *A. tatarinowii* as the outgroup, we clarified the evolutionary connections among these 28 closely related species, emphasizing the genes they share in their mitochondria and chloroplasts. We developed a phylogenetic tree based on the 35 mitochondrial PCGs that are prevalent in these 28 plant species as shown in Fig. 17. Phylogenetic tree analysis found the same results for evolutionary trees using 35 mitochondria shared CDS and 67 chloroplasts shared CDS. By further comparative analysis, our results and

Species	Repeat sequence type	SSRs number	Species	Repeat sequence type	SSRs number
<i>A.officinalis</i>	mono	50	<i>P.micranthum</i>	mono	34
<i>A. officinalis</i>	di	24	<i>P.micranthum</i>	di	30
<i>A.officinalis</i>	tri	20	<i>P.micranthum</i>	tri	22
<i>A.officinalis</i>	tetra	28	<i>P.micranthum</i>	tetra	33
<i>A. officinalis</i>	penta	3	<i>P.micranthum</i>	penta	11
<i>A.officinalis</i>	hexa	0	<i>P.micranthum</i>	hexa	0
<i>C.comosum</i>	mono	22	<i>D.nobile</i>	mono	51
<i>C.comosum</i>	di	45	<i>D.nobile</i>	di	41
<i>C.comosum</i>	tri	39	<i>D.nobile</i>	tri	28
<i>C.comosum</i>	tetra	59	<i>D.nobile</i>	tetra	65
<i>C.comosum</i>	penta	10	<i>D.nobile</i>	penta	4
<i>C.comosum</i>	hexa	3	<i>D.nobile</i>	hexa	2
<i>A.fistulosum</i>	mono	16	<i>D.cochinchinensis</i>	mono	37
<i>A.fistulosum</i>	di	19	<i>D.cochinchinensis</i>	di	35
<i>A.fistulosum</i>	tri	11	<i>D.cochinchinensis</i>	tri	18
<i>A.fistulosum</i>	tetra	25	<i>D.cochinchinensis</i>	tetra	40
<i>A.fistulosum</i>	penta	7	<i>D.cochinchinensis</i>	penta	7
<i>A.fistulosum</i>	hexa	2	<i>D.cochinchinensis</i>	hexa	2
<i>C.sativus</i>	mono	29	PKV-SNX60	mono	51
<i>C.sativus</i>	di	25	PKV-SNX60	di	43
<i>C.sativus</i>	tri	20	PKV-SNX60	tri	24
<i>C.sativus</i>	tetra	60	PKV-SNX60	tetra	63
<i>C.sativus</i>	penta	2	PKV-SNX60	penta	6
<i>C.sativus</i>	hexa	6	PKV-SNX60	hexa	2
<i>G.javanica</i>	mono	11	<i>P.kingianum</i>	mono	50
<i>G.javanica</i>	di	135	<i>P.kingianum</i>	di	50
<i>G.javanica</i>	tri	47	<i>P.kingianum</i>	tri	27
<i>G.javanica</i>	tetra	97	<i>P.kingianum</i>	tetra	64
<i>G.javanica</i>	penta	46	<i>P.kingianum</i>	penta	6
<i>G.javanica</i>	hexa	1	<i>P.kingianum</i>	hexa	2

Table 5. SSRs sequence characters of ten Asparagales species MtDNA.

GenBank: PP861176.1 form a separate branch, relationships are recent. Moreover, among the 28 species under *Asparagales*, all species of *Orchidaceae* clustered into large groups, while *Asparidaceae* cluster with *Lycoridaceae* and *Iris*, indicating that *Orchidaceae* are genetically distant from *Asparidaceae* in the phylogeny. However, in the *Asparagidae*, the ones closer to *Polygonatum* are *Dracaena* and *Chlorophytum*, indicating some kinship in genetic evolution. This study can provide reference and evidence for the genetic evolution and classification of *Asparagales*.

Intron SNP and indel analysis of *P. kingianum* MtDNA

The utility of mitochondrial genome data allows us to develop novel molecular markers for species identification. In the current study, we aligned and compared the sequences of the 8 introns regions (cox2-intron2, ccmFc-intron, nad2-intron1, nad4-intron1, nad5-intron1, nad5-intron2, nad7-intron2, nad7-intron3) of mitochondrial genomes of *P. kingianum* mtDNA. A total of 1451 SNPs and 275 Indels were detected in 8 intron regions. Each intron region exhibited a varying number of SNPs and Indels, ranging from 30 to 224 and 27 to 46, respectively (Please refer to supplementary material table S1 for details). The results are consistent for our mitochondrial genome and GenBank: PP861176.1 for the SNP and InDel sites in eight introns of six genes.

Discussion

Both chloroplasts(cp.) and mitochondria(mt) are important organelles of plants, among which chloroplasts can be regarded as the cornerstone of ecosystems and are an important place for plant photosynthesis, while mitochondria play a key role in the energy metabolism of plant cells⁴⁹. Moreover, studies have revealed that plant mitochondria have important functions in disease immunity and defense⁵⁰. Nevertheless, in contrast to the stable nature of plant chloroplasts, plant mitochondrial genomes have undergone substantial evolutionary transformations. It is a great challenge for the right assembly of plants' mitochondrial genomes. The structures of plant mitochondrial genomes are complex. These encompass polycyclic sequences, linear extensions, and radial configurations. This situation may be caused by repeat-mediated recombination⁵¹. Many researchers have investigated the complex structural variations in plant mitochondrial genomes, such as *Angelica dahurica*⁵², *Panax notoginseng*²⁴, *Angelica biserrata*²³, *Rhododendron delavayi*⁵³, *Flammulina filiformis*⁵⁴, *Santalum*

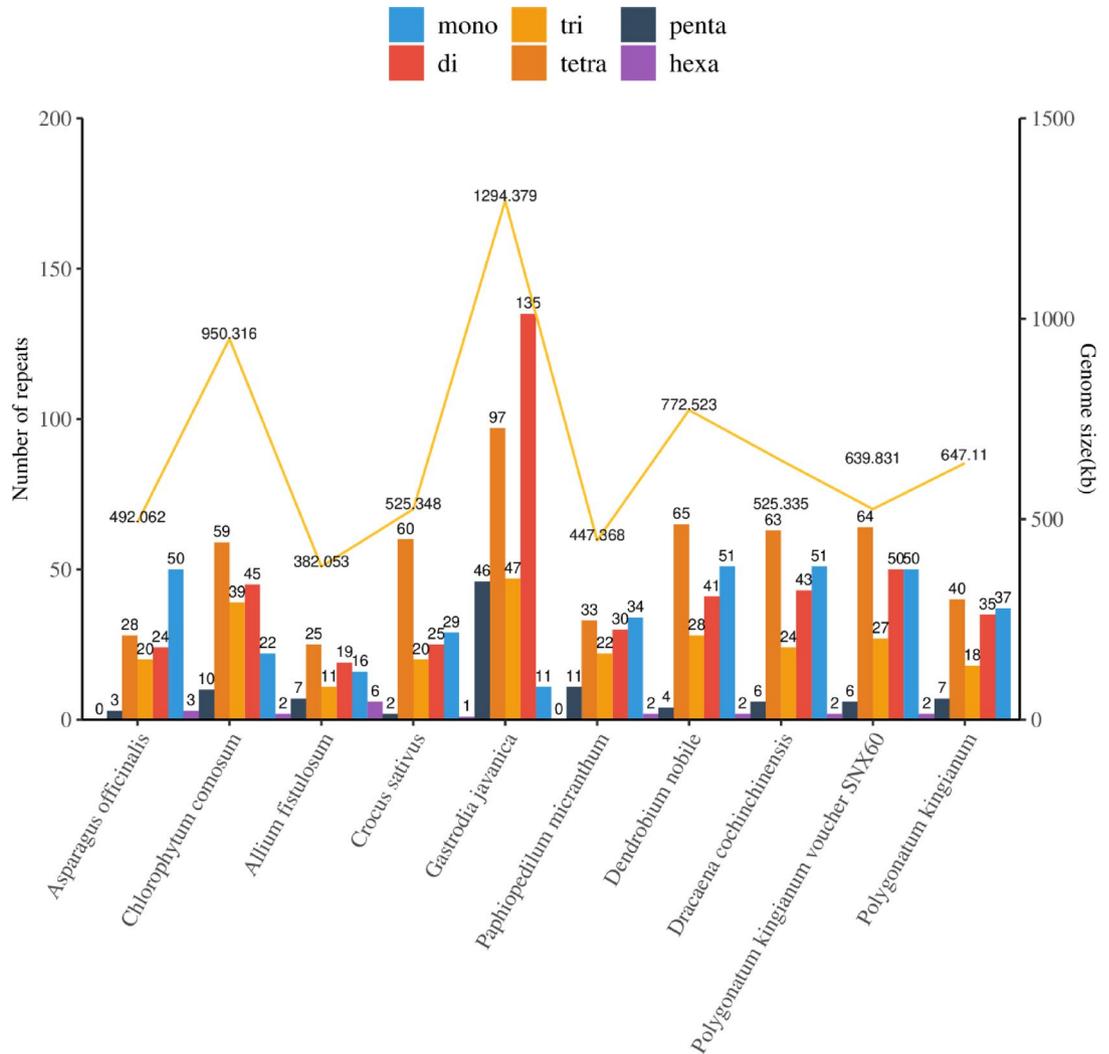


Fig. 11. Comparison of the numbers of simple sequence repeats and the sizes of mitochondrial genomes of ten Asparagales species. Note: The X-axis shows different species. Columns with different colors represent repeats having different types of repeat units. The number of a particular type of repeat is shown above the corresponding column. The sizes of the mitochondrial genomes are connected with a line.

Species information	Number
<i>A.officinalis</i>	10
<i>C.comosum</i>	101
<i>A.fistulosum</i>	36
<i>C.sativus</i>	16
<i>G.javanica</i>	400
<i>Pmicranthum</i>	27
<i>D.nobile</i>	62
<i>Pkingianum</i>	31
<i>D.cochinchinensis</i>	13
<i>Pkingianum voucher SNX60</i>	21

Table 6. Number of tandem repeats of ten Asparagales species MtDNA.

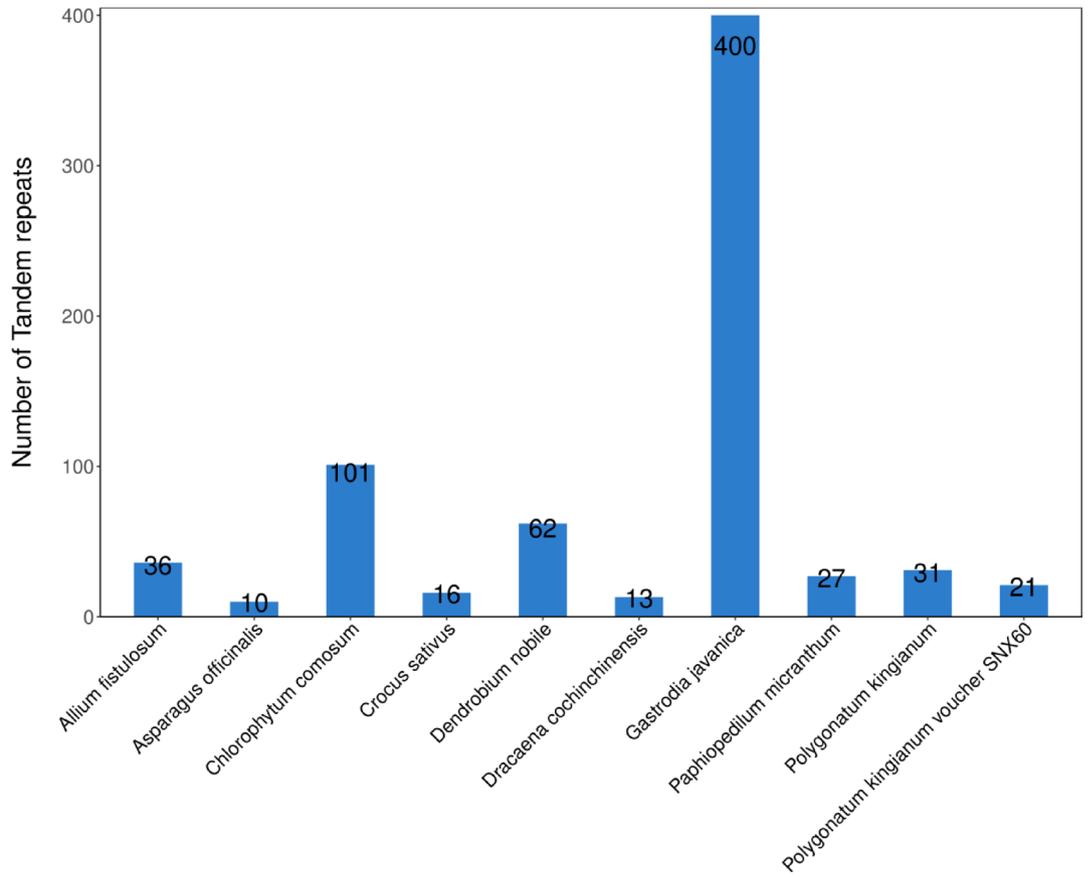


Fig. 12. Comparison of the numbers of Tandem repeats of mitochondrial genomes of the ten Asparagales species.

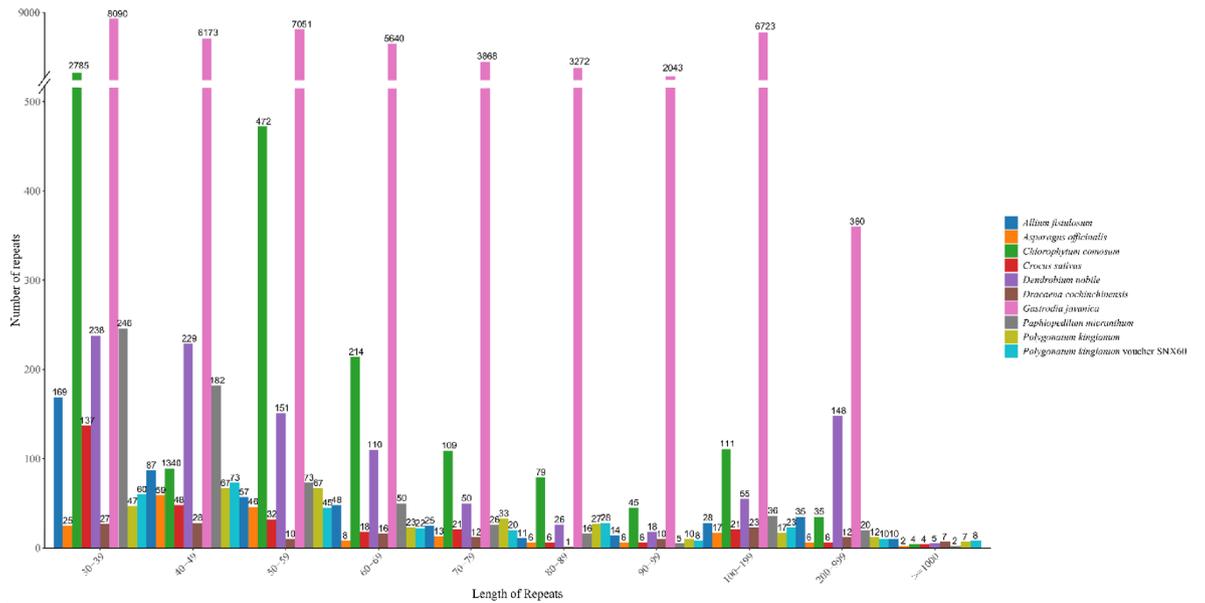


Fig. 13. Comparison of the numbers of length repeats of mitochondrial genomes of the ten Asparagales species.

Gene length	Species	Repeat the numbe	Gene length	Species	Repeat the numbe
30–39	<i>A.officinalis</i>	25	80–89	<i>A.officinalis</i>	6
30–39	<i>C.comosum</i>	2785	80–89	<i>C.comosum</i>	79
30–39	<i>A.fistulosum</i>	169	80–89	<i>A.fistulosum</i>	11
30–39	<i>Crocus sativus</i>	137	80–89	<i>Crocus sativus</i>	6
30–39	<i>G.javanica</i>	8090	80–89	<i>G.javanica</i>	3272
30–39	<i>Pmicranthum</i>	246	80–89	<i>Pmicranthum</i>	16
30–39	<i>D. nobile</i>	238	80–89	<i>D. nobile</i>	26
30–39	<i>D. cochinchinensis</i>	27	80–89	<i>D. cochinchinensis</i>	1
30–39	<i>Pkingianum voucher SNX60</i>	60	80–89	<i>Pkingianum voucher SNX60</i>	28
30–39	<i>P. kingianum</i>	47	80–89	<i>P. kingianum</i>	27
40–49	<i>A.officinalis</i>	59	90–99	<i>A.officinalis</i>	6
40–49	<i>C.comosum</i>	1340	90–99	<i>C.comosum</i>	45
40–49	<i>A.fistulosum</i>	87	90–99	<i>A.fistulosum</i>	14
40–49	<i>Crocus sativus</i>	48	90–99	<i>Crocus sativus</i>	6
40–49	<i>G.javanica</i>	6173	90–99	<i>G.javanica</i>	2043
40–49	<i>Pmicranthum</i>	182	90–99	<i>Pmicranthum</i>	5
40–49	<i>D. nobile</i>	229	90–99	<i>D. nobile</i>	18
40–49	<i>D. cochinchinensis</i>	28	90–99	<i>D. cochinchinensis</i>	10
40–49	<i>Pkingianum voucher SNX60</i>	73	90–99	<i>Pkingianum voucher SNX60</i>	8
40–49	<i>P. kingianum</i>	67	90–99	<i>P. kingianum</i>	10
50–59	<i>A.officinalis</i>	46	100–199	<i>A.officinalis</i>	17
50–59	<i>C.comosum</i>	472	100–199	<i>C.comosum</i>	111
50–59	<i>A.fistulosum</i>	57	100–199	<i>A.fistulosum</i>	28
50–59	<i>Crocus sativus</i>	32	100–199	<i>Crocus sativus</i>	21
50–59	<i>G.javanica</i>	7051	100–199	<i>G.javanica</i>	6723
50–59	<i>Pmicranthum</i>	73	100–199	<i>Pmicranthum</i>	36
50–59	<i>D. nobile</i>	151	100–199	<i>D. nobile</i>	55
50–59	<i>D. cochinchinensis</i>	10	100–199	<i>D. cochinchinensis</i>	23
50–59	<i>Pkingianum voucher SNX60</i>	45	100–199	<i>Pkingianum voucher SNX60</i>	23
50–59	<i>P. kingianum</i>	67	100–199	<i>P. kingianum</i>	17
60–69	<i>A.officinalis</i>	8	200–999	<i>A.officinalis</i>	6
60–69	<i>C.comosum</i>	214	200–999	<i>C.comosum</i>	35
60–69	<i>A.fistulosum</i>	48	200–999	<i>A.fistulosum</i>	35
60–69	<i>Crocus sativus</i>	18	200–999	<i>Crocus sativus</i>	6
60–69	<i>G.javanica</i>	5640	200–999	<i>G.javanica</i>	360
60–69	<i>Pmicranthum</i>	50	200–999	<i>Pmicranthum</i>	20
60–69	<i>D. nobile</i>	110	200–999	<i>D. nobile</i>	148
60–69	<i>D. cochinchinensis</i>	16	200–999	<i>D. cochinchinensis</i>	12
60–69	<i>Pkingianum voucher SNX60</i>	22	200–999	<i>Pkingianum voucher SNX60</i>	10
60–69	<i>P. kingianum</i>	23	200–999	<i>P. kingianum</i>	12
70–79	<i>A.officinalis</i>	13	>=1000	<i>A.officinalis</i>	2
70–79	<i>C.comosum</i>	109	>=1000	<i>C.comosum</i>	4
70–79	<i>A.fistulosum</i>	25	>=1000	<i>A.fistulosum</i>	10
70–79	<i>Crocus sativus</i>	21	>=1000	<i>Crocus sativus</i>	4
70–79	<i>G.javanica</i>	3868	>=1000	<i>Pmicranthum</i>	2
70–79	<i>Pmicranthum</i>	26	>=1000	<i>D. nobile</i>	5
70–79	<i>D. nobile</i>	50	>=1000	<i>D. cochinchinensis</i>	7
70–79	<i>D. cochinchinensis</i>	12	>=1000	<i>Pkingianum voucher SNX60</i>	8
70–79	<i>Pkingianum voucher SNX60</i>	20	>=1000	<i>P. kingianum</i>	7
70–79	<i>P. kingianum</i>	33			

Table 7. Dispersed repeat sequence characters of ten Asparagales species MtDNA.

*album*⁵⁵, *Myricaria laxiflora*⁵⁶ and so on, and multiple tools are developed to decode the dynamically evolving genomes^{57,58}. In this study, we sequenced and analyzed the mitochondrial genome of *P. kingianum* and found that the *P. kingianum* mitochondrial genome is a very complex multi-branched structure containing 20 nodes, with each node forming overlapping regions with each other along the connecting lines. Then, according to the

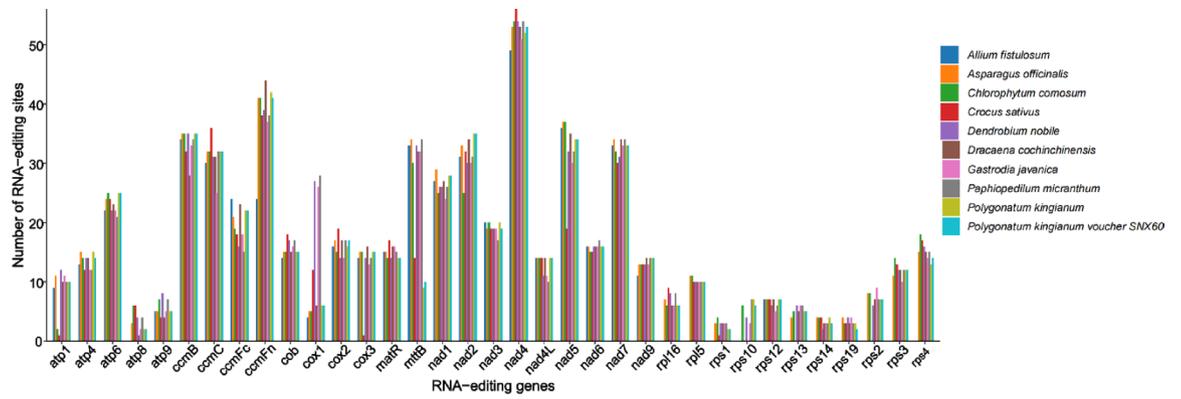


Fig. 14. Prediction of RNA editing sites based on the PCGs.

Sample	A.fistulosum	A.officinalis	C.comosum	C. sativus	D.nobile	D. cochinchinensis	G.javanica	P. micranthum	P.kingianum	P. kingianum voucher SNX60
atp1	9	11	2	1	12	10	11	10	10	10
atp4	13	15	14	12	14	14	12	12	15	14
atp6	22	24	25	24	22	23	22	21	25	25
atp8	0	3	6	6	4	1	2	4	2	2
atp9	5	5	7	4	8	4	5	7	5	5
ccmB	34	35	35	32	35	28	33	34	35	35
ccmC	30	32	32	36	31	31	25	32	32	32
ccmFc	24	21	19	18	16	23	18	15	22	22
ccmFn	24	41	41	38	39	44	37	38	42	41
cob	14	15	15	18	17	15	16	17	15	15
cox1	4	5	5	12	27	6	26	28	6	6
cox2	16	17	15	19	14	17	14	17	16	17
cox3	14	15	15	1	14	16	13	14	15	15
matR	15	15	14	17	14	16	16	15	14	14
mtbB	33	34	30	14	33	32	32	34	9	10
nad1	27	29	25	26	26	27	24	26	28	28
nad2	31	33	25	32	30	34	30	31	35	35
nad3	20	19	20	19	19	19	19	17	20	19
nad4	49	53	54	56	54	53	51	54	52	53
nad4L	14	14	14	14	11	14	11	10	14	14
nad5	36	37	37	19	32	35	30	32	34	34
nad6	16	16	15	15	16	16	16	17	16	16
nad7	33	34	32	30	31	34	33	34	33	33
nad9	11	13	13	13	13	14	13	14	14	14
rpl16	0	7	6	9	8	6	6	8	6	6
rpl5	0	11	11	10	10	10	10	10	10	10
rps1	0	3	4	1	3	3	3	3	2	2
rps10	0	0	6	0	4	0	3	7	7	6
rps12	7	7	7	7	6	7	5	6	7	7
rps13	0	4	5	0	6	5	6	6	5	5
rps14	0	4	4	4	2	3	3	3	4	3
rps19	0	4	3	3	4	3	4	3	3	2
rps2	0	8	8	0	6	7	9	7	7	7
rps3	0	11	14	13	12	12	10	12	12	12
rps4	0	15	18	17	16	15	14	15	13	14

Table 8. RNA-editing difference of ten Asparagales species MtDNA.

Type	RNA-editing	Number	Percentage
Hydrophilic-hydrophilic	CAC(H) => TAC(Y)	9	
	CAT(H) => TAT(Y)	21	
	CGC(R) => TGC(C)	12	
	CGT(R) => TGT(C)	34	
	total	76	12.99%
Hydrophilic-hydrophobic	ACA(T) => ATA(I)	5	
	ACC(T) => ATC(I)	2	
	ACG(T) => ATG(M)	7	
	ACT(T) => ATT(I)	4	
	CGG(R) => TGG(W)	38	
	TCA(S) => TTA(L)	74	
	TCC(S) => TTC(F)	42	
	TCG(S) => TTG(L)	52	
	TCT(S) => TTT(F)	55	
	total	279	47.69%
Hydrophilic-stop	CAA(Q) => TAA(X)	1	
	CGA(R) => TGA(X)	3	
	total	4	0.68%
Hydrophobic-hydrophilic	CCA(P) => TCA(S)	8	
	CCC(P) => TCC(S)	15	
	CCG(P) => TCG(S)	6	
	CCT(P) => TCT(S)	22	
	total	51	8.72%
Hydrophobic-hydrophobic	CCA(P) => CTA(L)	52	
	CCC(P) => CTC(L)	8	
	CCC(P) => TTC(F)	4	
	CCG(P) => CTG(L)	34	
	CCT(P) => CTT(L)	28	
	CCT(P) => TTT(F)	12	
	CTC(L) => TTC(F)	6	
	CTT(L) => TTT(F)	20	
	GCC(A) => GTC(V)	1	
	GCG(A) => GTG(V)	6	
	GCT(A) => GTT(V)	4	
	total	175	29.91%
All		585	100%

Table 9. Type of RNA-editing of *P. kingianum* MtDNA.

depth information on the note, the respective repeat region, which is a linear genome of 647,110 bp, GC content of 46.2%, revealing a total of 35 distinct protein-coding genes, 22 tRNA genes, and 4 rRNA genes. In comparison to the published mitochondrial genome of *P. kingianum* (GenBank: PP861176.1), the mitochondrial genome we determined for *P. kingianum* is larger, featuring a greater number of genes, as well as an increased quantity of rRNAs and tRNAs.

To further explore the evolutionary characteristics of the mitochondrial genome of *P. kingianum*, we compared it with nine other Asparagales species (*D. nobile*, *P. micranthum*, *P. kingianum*, *D. cochinchinensis*, *G. javanica*, *C. sativus*, *C. comosum*, *A. officinalis*, *A. fistulosum*) genome length, total gene content%, No of PCGs, No of rRNAs, No of tRNAs, AT%, and GC% in the mitochondrial of 10 plant species show significant differences. During the evolution of the angiosperm mitochondrial genomes, the loss of PCG has occurred frequently⁵⁹. Collinearity analysis revealed inconsistent order of collinear blocks between mitochondrial genomes across species, with *P. kingianum* mitomes and closely related species experiencing substantial genome rearrangements and extremely unconserved in structure.

In addition, our analysis revealed that the majority of PCGs are conserved in 10 plant mitochondria, especially the mitochondrial respiratory chain-related genes, maturase, and methyltransferase genes. In contrast, the ribosomal protein and succinate dehydrogenase genes are highly variable. A total of seven genes were lost in the *P. kingianum* mitochondrial genome, including six ribosomal protein genes (rpl2, rpl6, rpl10, rps7, rps8, rps11) and one succinate dehydrogenase gene (sdh3). There is also a pseudogene for the succinate dehydrogenase gene (sdh 4) (Table 1). The Ka/Ks ratio is a valuable metric for understanding how genes evolve in response to environmental stresses in flowering plants. When studying genetic alterations and their impact on phenotypes in

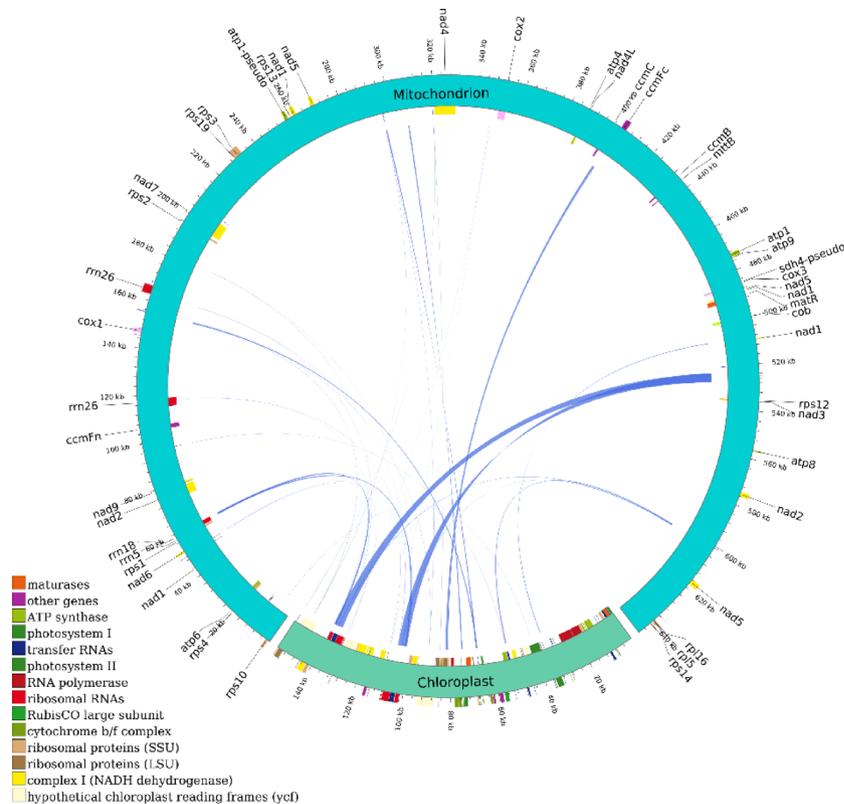


Fig. 15. Homologous analysis based on the different organelles. Note: The regions connected by arcs indicate regions of good homology, where the red arc represents the reverse sequence, and the gray region represents the sequence forward direction. Furthermore, some regions that failed to have collinear blocks represent those unique to this species.

different individuals or populations of seed plants, the K_a/K_s ratio can reveal the evolutionary path of genes and the underlying genetic mechanism^{23,54}. The results showed that the average K_a/K_s values of all ten species PCGs were < 1 , showing negative selection, indicating that these species PCGs had all undergone purifying selection during evolution, were highly conserved during evolution, and had relatively stable protein functions. This result is consistent with the gene-loss analysis.

Codons, which carry important recognition and transitional information within seed plants, are of great significance in the context of genetic mutations. The use of certain synonymous codons during protein translation is notably influenced by species-specific variations, leading to what is known as codon usage bias. This preference for specific synonymous codons plays a crucial role in shaping the genetic characteristics of these organisms^{60,61}. In this study, we assessed the relative synonymous codon usage in *P. kingianum*'s mitochondrial DNA. The results indicated *P. kingianum* mitochondria have 29 codons with $RSCU > 1$ and all end with A/U, suggesting a greater bias to use codons ending with A/U. GC_1 and GC_2 are significantly different from GC_3 , indicating that the composition of the first and second bases and the third base can differ somewhat. Meanwhile, GC_3 and ENC are significantly correlated, indicating that they influence the bias of codon usage. Neutral mapping and ENC-plot analysis found that natural selection may play an important role in synonymous codon usage bias in *P. kingianum* chloroplast genomes, including other factors such as mutation pressure and base composition. This conclusion is consistent with the study of Jia-hao Liu et al.⁶².

In biological cells, the DNA sequence contains many repeated sequences, which can be divided into two main categories: Tandem repeat and Interspersed repeat. The difference between tandem repeats and scattered repeats is whether the part of the repeat is distributed next to each other, adjacent is tandem repeats, not adjacent is scattered repeats. Among tandem repeats, Simple Sequence Repeat (SSR) is a special sequence of serial repeats, generally no more than 6 bp. Because SSR has characteristics such as dominant inheritance, it can be used to develop molecular marker development^{24,63,64}. We found that there are 199 SSRs, 31 tandem repeats, and 310 pairs of length greater than or equal to 30 repeats in the mitochondrial genome. Which were quite different by comparison with closely related species. Compared to the already published data (GenBank: PP861176.1), SSRs, tandem repeats, and lengths greater than or equal to 30 repeats over 10, 10, and 13, respectively. This study is useful for interspecies identification and molecular-assisted breeding.

During mitochondrial evolution, some chloroplast fragments will migrate into the mitochondrial genome^{49,65}. We found that there are 31 homologous fragments in *P. kingianum* mitochondria, ranging from 38 to 4,363 bp, with a total length of 9,853 bp. By annotation, eight complete genes were identified on the homologous fragments, including seven tRNA genes (trnA-UGC, trnI-CAU, trnM-CAU, trnN-GUU, trnP-UGG, trnH-GUG, trnW-

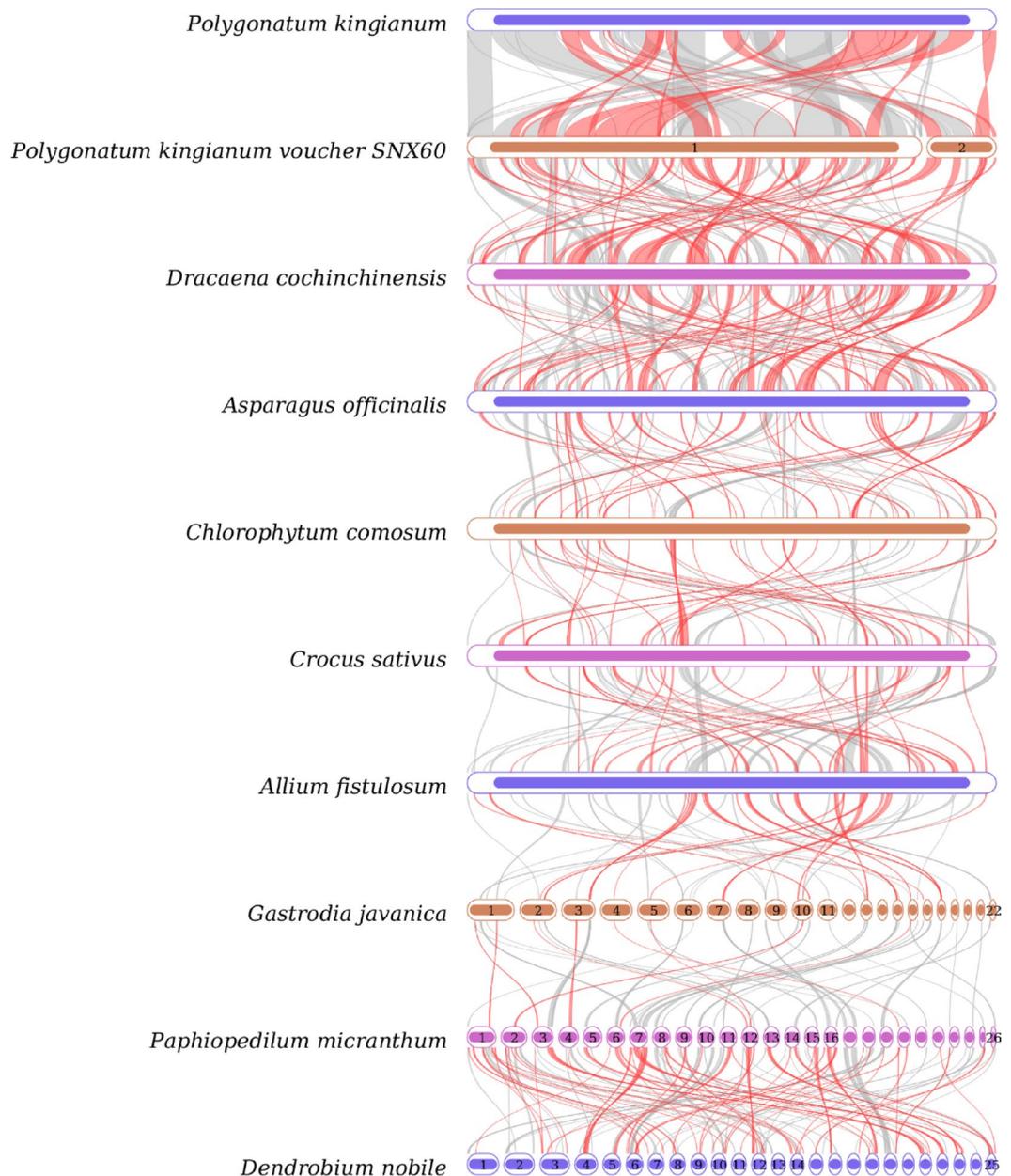


Fig. 16. Collinear analysis of ten Asparagales species. The red arcs indicate inverted regions, while the gray arcs indicate better homologous regions. The red arcs indicated inverted regions. The gray arcs indicated better homologous regions.

CCA) and one rRNA gene (*rrn16*), suggesting that these homologous segments may have undergone multiple independent transfer integration, replication, and recombination within the mitochondrial DNA. Similar results exist for the transfer of related genes in other species^{23,65}.

RNA editing is widely found in plant organelles (mitochondria and chloroplasts) and is essential for plant growth and development⁶⁶. RNA editing can improve the mitochondrial protein sequence homology between different species, enabling proteins to be better folded to function^{67,68}. Research has shown that RNA editing mainly consists of base replacements mediated by deaminase enzymes in organelles, such as cytosine to uracil (C-to-U), uracil to cytosine (U-to-C), and adenine to hypoxanthine (A-to-I) conversions⁶⁹. Our analysis revealed that among the 35 protein-coding genes (PCGs) in the mitochondrial genome of *P. kingianum*, a total of 585 RNA editing sites were predicted, all involving C to U editing. Notably, the *nad4* gene exhibited the highest frequency of editing events among all mitochondrial genes, followed by the *ccmFn* gene. There is a significant difference in RNA editing times among closely related species. Notably, when comparing our mitochondrial genome to GenBank: PP861176.1, our data reveal more than two RNA editing events (Fig. 14). The primary differences between the two are illustrated in the RNA editing occurrences of the *ccmFn*, *mttB*, *nad4*, *rps10*, *rps14*, *rps19*, and *rps4* genes (Table 8).

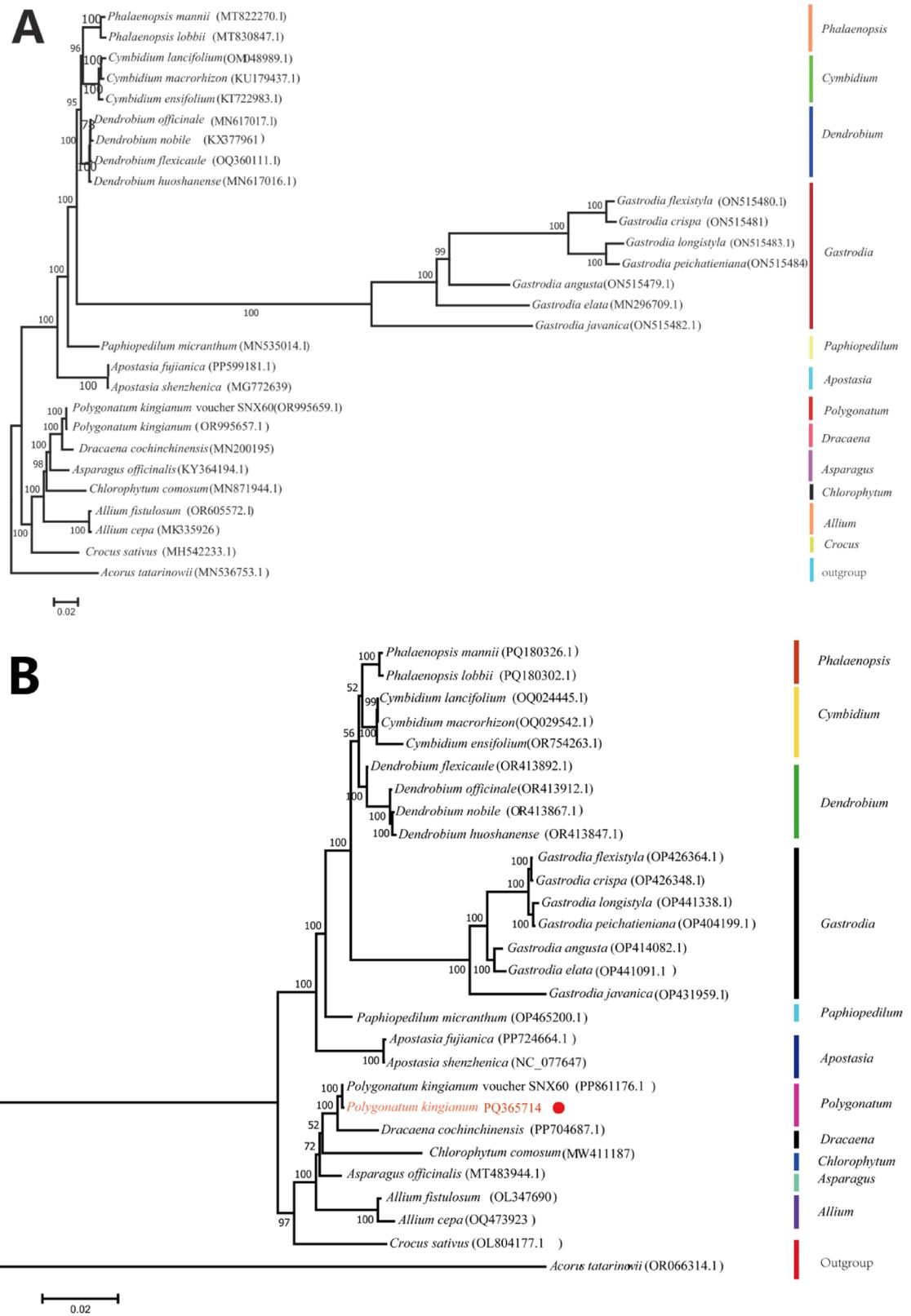


Fig. 17. Based on the maximum likelihood tree of 28 species. (A) plant chloroplasts use 67 common CDS to construct an evolutionary tree; (B) plant mitochondria use 35 common CDS to construct an evolutionary tree.

To determine the phylogenetic status of *P. kingianum*, we downloaded the mtDNA and cpDNA sequences of 27 closely related species from NCBI and identified 35 and 67 common PCGs from mtDNA and cpDNA by comparative analysis. The results show that our mitochondrial genome and (GenBank: PP861176.1) are clustered into one clade, belonging to the same species type. Among the 28 species under Asparagales, all species of Orchidaceae clustered into large groups, while Asparidaceae cluster with Lycoridaceae and Iris, indicating that Orchidaceae are genetically distant from Asparidaceae in the phylogeny. Furthermore, in the Asparagidae, the ones closer to *Polygonatum* are *Dracaena*, indicating some kinship in genetic evolution. The results further support Meng Ran's research view⁷⁰. In that study, the results of the mitochondrial and chloroplast phylogenetic analyses were consistent. However, it is worth noting that since NCBI currently reported few mitochondrial genome data of *Polygonatum* and related plants, *Polygonatum* only published mitochondrial genome data of *P. kingianum*, which has some limitations in the phylogenetic analysis of this genus. Therefore, it is necessary to continuously carry out the sequencing research of *Polygonatum* species.

Conclusions

In this study, we successfully sequenced the mitochondrial genome of *P. kingianum*, revealing a unique multibranch conformation in its mtDNA structure. Furthermore, we conducted a comprehensive analysis of codon preference, repeat sequences, homologous sequences, RNA editing, and SNP/InDel variations in *P. kingianum*, comparing these findings with closely related species. This research offers valuable insights into the breeding, conservation, and classification of *P. kingianum*, providing a foundation for understanding its genetic diversity and adaptive evolution. Future analyses may explore key mitochondrial genes and regulatory mechanisms, thereby facilitating the selection of superior germplasm and accelerating the development of improved cultivars.

Data availability

Publicly available datasets were analyzed in this study. Tese data are accessible at the NCBI-SRA (<https://www.ncbi.nlm.nih.gov/Genbank/update.html>) under the accession number PQ365714. If someone wants to request the data from this study, please contact Tao Ai-en (E-mail addresses: 2515073996@qq.com; Tel: 18987245571).

Received: 20 October 2024; Accepted: 8 May 2025

Published online: 06 June 2025

References

1. Sinicae, C. A. O. S. *Flora of China Volume 15 Liliaceae 2* (Science, 1990).
2. Jia-Jian, W. et al. The biogeographic South-North divide of polygonatum (Asparagaceae tribe polygonateae) within Eastern Asia and its recent dispersals in the Northern hemisphere. *PLoS One*. **11** (11), e166134 (2016).
3. Jia, Z. Y. & Minru *Chinese Ethnic Medicine Dictionary* (China Medical Science, 2016).
4. Commission, C. P. *Pharmacopoeia of the people's republic of china* (People's Medical Publishing House, 2020).
5. Puwei, W. *Shemong's herbal classic* (Shanxi Science and Technology Publishing House, 2010).
6. Cui, X. et al. A Review: The Bioactivities and Pharmacological Applications of Polygonatum sibiricum polysaccharides, *Molecules* **23** (5) 1170. (2018).
7. Zhao, P. et al. The genus polygonatum: A review of ethnopharmacology, phytochemistry and Pharmacology. *J. Ethnopharmacol.* **214**, 274–291 (2018).
8. Huan, G. et al. Review on the genus Polygonatum polysaccharides: extraction, purification, structural characteristics and bioactivities. *Int. J. Biol. Macromol.* **229** (2023).
9. Mingying, Z. et al. Molecular authentication of medicinal Polygonatum species utilizing the universal DNA barcode sequences. *Chin. Herb. Med.* **54** (01), 235–244 (2023).
10. Xia, M. et al. Out of the Himalaya-Hengduan mountains: phylogenomics, biogeography and diversification of Polygonatum mill. (Asparagaceae) in the Northern hemisphere. *Mol. Phylogenet Evol.* **169**, 107431 (2022).
11. Zhang, L. Y. C. W. & Mingying Molecular authentication of medicinal Polygonatum species utilizing the universal DNA barcode sequences. *Chin. Herb. Med.* **54** (01), 235–244 (2023).
12. Qijian Tian, Z. Z. Current situation of polygonatum plant in categories identification and resources distribution. *Res. Pract. Chin. Med.* **01**, 18–21 (2007).
13. Zhao, P. et al. Characterisation and saccharide mapping of polysaccharides from four common Polygonatum spp. *Carbohydr Polym.* **233**, 115836 (2020).
14. Yang Pei, Z. H. X. T. Identification study of DNA barcode sequences in the medicinal plants of polygonatum. *World Chin. Med.* **10** (08), 1173–1176 (2015).
15. Shen Xiaoxia, S. J. W. Z. Phylogenetic relationship of Polygonatum medicinal plants based on CpDNA sequences analyses of MatK gene and rps16 gene. *Mod. Chin. Med.* **23** (02), 275–279 (2021).
16. Missouri Botanical Garden, S. B. S. L. Department of ecology and evolutionary biology, using phylogenomics to reconstruct phylogenetic relationships within tribe polygonateae (Asparagaceae), with a special focus on Polygonatum. *Mol. Phylogenet Evol.* **129**, 202–213 (2018).
17. Yih, L. S. et al. Plastid genome sequencing, identification of nuclear SNP markers, and quality assessment of medicinal rhizomatous herb Polygonatum odoratum (Asparagaceae) cultivars, ecol. *Evol* **11** (12), 7660–7676 (2021).
18. Jiao, J. et al. Species identification of polygonati rhizoma in China by both morphological and molecular marker methods. *Comptes rendus - Biologies.* **341** (2), 102–110 (2018).
19. Wang, L. L. X. R. & Zhongni Advances in mitochondrial proteins responding to stresses in plants. *Plant. Physiol. J.* **54** (02), 221–231 (2018).
20. Danfeng, T., Fan, W., Aziz, K., Fazal, M. & Ruiyang, Z. Degradation of mitochondrial structure and deficiency of complex I were associated with the Transgenic CMS of rice. *Biol. Res.* **54** (1), 6 (2021).
21. Zhang, Z. R. H. S. & Xiao Research progress on mitochondrial genome of higher plant. *J. Agr Sci. Tech. -Iran.* **13** (04), 23–31 (2011).
22. Yanlei, F. et al. Assembly and phylogenomic analysis of cotton mitochondrial genomes provide insights into the history of cotton evolution. *Crop J.* **11** (6), 1782–1792 (2023).
23. Wang, L. et al. Assembly and comparative analysis of the first complete mitochondrial genome of a traditional Chinese medicine Angelica biserrata (Shan et Yuan) Yuan et Shan. *Int. J. Biol. Macromol.* **257** (Pt 1), 128571 (2024).

24. Heyu, Y. et al. The mitochondrial genomes of *Panax Notoginseng* reveal recombination mediated by repeats associated with DNA replication. *Int. J. Biol. Macromol.* **252**, 126359 (2023).
25. Neurodegenerative Brain Diseases, V. U. O. A., Group, N. S. F. C. & Bioinformatics, U. O. C. S. Department Of Biomolecular Engineering And Bioinformatics, V.U.O.A. Neurodegenerative Brain Diseases Group, V.U.O.A. Neurodegenerative Brain Diseases Group, NanoPack: Visualizing and processing long-read sequencing data., *Bioinformatics (Oxford, England)* **34** (15) 2666–2669. (2018).
26. Patel, R. K., Jain, M. & Toolkit, N. G. S. Q. C. A toolkit for quality control of next generation sequencing data. *PLoS One.* **7** (2), e30619 (2017).
27. Wick, R. R., Schultz, M. B., Zobel, J. & Holt, K. E. Bandage: interactive visualization of de Novo genome assemblies. *Bioinformatics* **31** (20), 3350–3352 (2015).
28. Direct, G. D. L. A. RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *J. Mol. Biol.* **313** (5), 1003–1011 (2001).
29. J, A. A. et al. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae)., *mol. Biol. Evol.* **27** (6), 1436–1448 (2010).
30. Greiner, S., Lehwark, P. & Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **47** (W1), W59–W64 (2019).
31. Dong, Z. et al. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.* **20** (1), 348–355 (2020).
32. Kazutaka, K. & M, S. D. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30** (4), 772–780 (2013).
33. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11, *Mol. Biol. Evol.* **38** (7) 3022–3027. (2021).
34. P, L. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinf. (Oxford England).* **25** (11), 1451–1452 (2009).
35. Leigh, J. W. S. S., Bryant, D. S. S. & Nakagawa, S. S. S. *Popart: Full-feature Software for Haplotype Network Construction* 1110–1116 (Methods in Ecology & Evolution, 2015). (No.9).
36. Zhang, Z. et al. KaKsCalculator: calculating Ka and Ks through model selection and model averaging. *Genom. Proteom. Bioinform.* **4** (4), 259–263 (2006).
37. Beier, S., Thiel, T., Münch, T., Scholz, U. & Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **33** (16), 2583–2585 (2017).
38. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27** (2), 573–580 (1999).
39. Chen, Y., Ye, W., Zhang, Y. & Xu, Y. High speed BLASTN: an accelerated megablast search tool. *Nucleic Acids Res.* **43** (16), 7762–7768 (2015).
40. Li, L., Fu, H., Altaf, M. A., Wang, Z. & Lu, X. The complete mitochondrial genome assembly of *Capsicum pubescens* reveals key evolutionary characteristics of mitochondrial genes of two *Capsicum* subspecies. *BMC Genom.* **25** (1), 1064 (2024).
41. Librado, P. & Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25** (11), 1451–1452 (2009).
42. Zhang, H., Meltzer, P. & Davis, S. RCircos: an R package for circos 2D track plots. *BMC Bioinform.* **14**, 244 (2013).
43. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. TrimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25** (15), 1972–1973 (2009).
44. Posada, D. JModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* **25** (7), 1253–1256 (2008).
45. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30** (9), 1312–1313 (2014).
46. Cole, L. W., Guo, W., Mower, J. P. & Palmer, J. D. High and variable rates of Repeat-Mediated mitochondrial genome rearrangement in a genus of plants. *Mol. Biol. Evol.* **35** (11), 2773–2785 (2018).
47. Lukeš, J., Kaur, B. & Spejler, D. RNA editing in mitochondria and plastids: Weird and widespread. *Trends Genet.* **37** (2), 99–102 (2021).
48. Wanqi, X. The mitochondrial genome of *cistanche* genus in china. <https://10.27648/d.cnki.gzxhu.2020.000890>. (2020).
49. Wang, H., Wu, Z., Li, T. & Zhao, J. Highly active repeat-mediated recombination in the mitogenome of the aquatic grass *Hygroryza aristata*. *BMC Plant. Biol.* **24** (1), 644 (2024).
50. Jiyang, W., Guojuan, X., Yuese, N., Xuli, W. & Guoliang, W. Mitochondrial functions in plant immunity. *Trends Plant. Sci.* **27** (10), 1063–1076 (2022).
51. Zhang, X. et al. Assembly of the complete mitochondrial genome of *pereskia aculeata* revealed that two pairs of repetitive elements mediated the recombination of the genome. *Int. J. Mol. Sci.* **24** (9) (2023).
52. Li, Y. Y. et al. Complete mitochondrial genome of *Angelica Dahurica* and its implications on evolutionary analysis of complex mitochondrial genome architecture in apiaceae. *front. Plant. Sci.* **15**, 1367299 (2024).
53. Zuo, W. & Li, H. Assemble and comparative analysis of the mitochondrial genomes of *Rhododendron Delavayi*: insights into phylogenetic relationships and genomic variations. *Gene* **927**, 148741 (2024).
54. Hao, T. et al. Comparative analyses of *flammulina filiformis* mitochondrial genomes reveal high length polymorphism in intergenic regions and multiple intron gain/loss in *cox1*. *Int. J. Biol. Macromol.* **221**, 1593–1605 (2022).
55. Liu, G. H. et al. Structural analysis of the mitochondrial genome of *Santalum album* reveals a complex branched configuration. *Genomics* **116** (5), 110935 (2024).
56. Sup, C. W. S. et al. Mitochondrial genome analysis of *myricaria laxiflora*, a protected endangered plant. *Phyton-Int J. Exp. Bot.* **8**, 1981–1993 (2024).
57. Fischer, A., Dotzek, J., Walther, D. & Greiner, S. Graph-based models of the *Oenothera* mitochondrial genome capture the enormous complexity of higher plant mitochondrial DNA organization. *NAR Genom Bioinform.* **4** (2), c27 (2022).
58. He, W., Xiang, K., Chen, C., Wang, J. & Wu, Z. Master graph: an essential integrated assembly model for the plant mitogenome based on a graph-based framework. *Brief. Bioinform* **24** (1) (2023).
59. Zhong, F. et al. Comprehensive analysis of the complete mitochondrial genomes of three *Coptis* species (*C. Chinensis*, *C. Deltoidea* and *C. Omeiensis*): the important medicinal plants in China. *front. Plant. Sci.* **14**, 1166420 (2023).
60. Hershsberg, R. & Petrov, D. A. Selection on codon bias. *Annu. Rev. Genet.* **42**, 287–299 (2008).
61. Yuaning, Z. Construction of multifunctional codon analysis and optimization platform and its application study. <https://10.27461/d.cnki.gzjdx.2020.001018>. (2020).
62. Liu Jiahao, W. J. G. Y. Analysis of Codon Usage Patterns and Bias in the Chloroplast Genome of *Polygonatum kingianum* Coll. Et Hemsl., *Molecular Plant Breeding* 1–16.
63. Hao, Z. et al. Complete mitochondrial genome of *Melia azedarach* L., reveals two conformations generated by the repeat sequence mediated recombination. *BMC Plant. Biol.* **24** (1), 645 (2024).
64. Zhang, L. C. F. Y. & Lanlan Progress on the application of mitochondrial SSR molecular markers in plants. *Curr. Biotechnol.* **13** (06), 821–826 (2023).
65. Wentao, S., Jianlan, D., Chao, W. & Quan, K. The garden asparagus (*Asparagus officinalis* L.) mitochondrial genome revealed rich sequence variation throughout whole sequencing data., *front. Plant. Sci.* **14**, 1140043 (2023).

66. Yong, W. et al. Maize PPR-E proteins mediate RNA C-to-U editing in mitochondria by recruiting the trans deaminase PCW1. *Plant. Cell.* **35** (1) (2022).
67. Kaiyun, X. et al. Experimental evidence for the functional importance and adaptive advantage of A-to-I RNA editing in fungi. *P Natl. Acad. Sci. Usa.* **120** (12), e2075938176 (2023).
68. Haijun, X. et al. Rice PPS1 encodes a DYW motif-containing pentatricopeptide repeat protein required for five consecutive RNA-editing sites of nad3 in mitochondria. *New Phytol.* **220** (3), 878–892 (2018).
69. Zhu, X. F. Z. Q. & Lin Research progress of RNA editing. *Biotechnol. Bull.* **38** (01), 1–14 (2022).
70. Ran, M. & A Phylogenomic Study on Nolinoideae of Asparagaceae. <https://10.27750/d.cnki.gjdx.2020.000069>. (2020).

Acknowledgements

We thank the Editor and the anonymous reviewers for their insightful comments and suggestions on the manuscript. The authors thank Shenzhen Huitong Biotechnology Co., Ltd. for help in genome sequencing and analysis.

Author contributions

Aien Tao: Conceptualization, Writing-original draft. Jing Tao: Writing - original draft. Zhengkun Gan: Writing - original draft, Writing - review & editing. Yue Zhang: Writing - review & editing, Visualization. Chenglong Yin: Writing - original draft. Guihua Jiang: Supervision.

Funding

This work was supported by Yunnan Fundamental Research Projects (grant NO. 202301 AT070898, NO. 202101 AT070001), National Natural Science Foundation of China (No. 82173928), Sichuan Provincial Natural Science Foundation Project (No. 2024NSFSC0704).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-01756-y>.

Correspondence and requests for materials should be addressed to G.J.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025