

RESEARCH

Open Access



Assembly and analysis of *stephania japonica* mitochondrial genome provides new insights into its identification and energy metabolism

Ya Wu^{1,2†}, Zhihao Sun^{2,3†}, Zhaoyu Liu^{2,4}, Ting Qiu⁵, Xiaojing Li⁵, Liang Leng^{2*} and Shilin Chen^{2*}

Abstract

Stephania japonica, a popular indoor ornamental and medicinal plant widely found in southern China, contains many natural compounds with potential medicinal value. *S. japonica* is also favored by researchers for its ability to produce catharanthine. Energy metabolism functions in plant development, and the composition of mitochondrial genome is regarded as the foundation for understanding energy metabolism and getting insights into plant environmental adaptation. In present investigation, the whole mitochondrial genome of *S. japonica* was assembled from both second- and third-generation sequencing data. The mitochondrial genome size of *S. japonica* is 555,117 bp. It is depicted as a complex polycyclic structure. In addition, we conducted an in-depth study of the cytochrome c oxidase (*cox*) gene, of which expression levels in different tissues of *S. japonica* were measured by real-time quantification PCR. Two phylogenetic trees were established in the light of sequences concerning 19 conserved mitochondrial protein-coding genes and *cox* gene, respectively. Both phylogenetic trees show that *S. japonica* is more closely related to *Aconitum kusnezoffii*. The result showed that the *cox* genes were the most highly expressed in the roots. A high-quality mitochondrial genome exhibits potential application value for the progress of molecular markers, identification of species as super DNA barcoding, and resolve mitochondrial energy metabolism mechanisms in response to the environment using genomic information. With the recognition of the medicinal value of *Stephania* plants, the genomic information of *S. japonica* has been thoroughly studied and the comprehensive analysis of its mitochondrial genome in this investigation can offer valuable insights for the breeding of new plant varieties.

Keywords *Stephania Japonica*, Mitochondrial genome, Cytochrome c oxidase, Evolution, Polymerase chain reaction

[†]Ya Wu and Zhihao Sun contributed equally to this work.

*Correspondence:

Liang Leng

lling@cdutcm.edu.cn

Shilin Chen

slchen@cdutcm.edu.cn

¹School of Pharmacy, School of Modern Chinese Medicine Industry, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China

²Institute of Herbgonomics, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China

³College of Pharmacy, Hubei University of Chinese Medicine, Wuhan 430065, China

⁴School of Chinese Materia Medica, Tianjin University of Traditional Chinese Medicine, Tianjin 300193, China

⁵Wuhan Benagen Technology Co., Ltd., Wuhan, Hubei 430000, China



Introduction

During the last several years, Coronavirus Disease 2019 (COVID-19) was deemed a significant issue in people's lives. Research on the repurposing of clinically authorized medications for treating COVID-19 has shown that cepharanthine has significant therapeutic promise [1]. *Stephania japonica* was of interest to researchers because of its ability to produce catharanthine [1–3]. *S. japonica* is a widely distributed medicinal plant and widespread indoor decorative crop in southern China [4]. Medicinal plants contain numerous natural compounds with potential medicinal value. Taking inspiration from natural products is easier than designing specific small molecules based on the structure of proteins [5]. In recent studies, it has been found that *S. japonica* contains a variety of valuable alkaloid components, which can treat many types of diseases, including cancer [6] and leukopenia [3], which has garnered the attention of researchers in recent years, particularly in the development of drugs targeting receptors [7]. Therefore, *S. japonica* possesses extensive medicinal properties, and the recent high-quality nuclear genome sequencing of the *Stephania* genus has enhanced our understanding of the biosynthetic mechanisms of cepharanthine within *Stephania* species [8–10].

Chloroplast genomes exhibit simpler structural features compared to mitochondrial genomes, and consequently, chloroplast genome research in *Stephania* plants has also been extensively conducted [11, 12]. Despite the valuable data provided by chloroplast and nuclear genomes for various research purposes, the analysis of mitochondrial genome also offers indispensable and effective data for delving deeper into the genetic basis of its agronomic traits. Because of its maternal inheritance qualities, the mitochondrial genome is a valuable addition to the nuclear genome in many species' evolutionary studies [13, 14]. Mitochondria produce energy for each cell as their main function. Furthermore, they are vital for the multiplication, differentiation, and death of cells [15, 16]. Maguire's endosymbiosis theory holds that mitochondria develop into unique organelles during long-term plant symbiosis after emerging from the nucleogenetic archaea [17, 18]. For mitochondrial structure, early research suggested that the mitochondrial genome had a closed-loop structure and that all of the genetic information about mitochondria was included in the circular shape [19, 20]. However, recent research found that the mitochondrial genome may be more than a simple closed-loop structure [19]. Plant mitochondrial genomes differ in structure and sequence due to rapid invasion by short textual insertions and fragment migration of the chloroplast genome [14]. The complex structures of mitochondria have been found in many plants, such as *Abelmoschus esculentus* [19] and *Rhopalocnemis phalloides* [21].

Mitochondria are the primary sites of oxidative phosphorylation and ATP production in plant cells, influencing intracellular energy demand. They also act as pivotal organelles in the response of plants to environmental signals and play a vital role in energy balance [22, 23]. When faced with stress, mitochondria have the function of converting stress perception into signals of energy deficiency, which subsequently aid in restoring metabolic balance [24]. The mitochondrial genome also has potential for plant molecular identification; although the evolutionary rate is slower, it provides polymorphism and stability and can be used as a candidate sequence for multi-genome, multi-fragment barcoding [25].

Through this investigation, we achieved the assembly and annotation of mitochondrial genome of *S. japonica* for the first time. We also examined migration between mitochondria and chloroplast genomes, repetitive regions, as well as codon use bias. Clarifying the roles of plant mitochondria as well as their evolutionary and genetic links requires thorough investigations of plant mitochondrial genomes. We further investigated the evolutionary link between *S. japonica* and the degree of *cox* gene expression. The mitochondria are essential to plant growth and energy metabolism, and a comprehensive mitochondrial genome analysis of *S. japonica* will contribute to acknowledge the genetic information of *S. japonica* growth and development as well as evolution in the Ranunculales.

Materials and methods

DNA extraction and sequencing

Three healthy *S. japonica* plants were chosen from the Wuhan, China. Using the CTAB, DNA was extracted from roots, stems, leaves, and shoots. Wuhan Benagen Tech Solutions Company conducted both Oxford and Illumina sequencing (<http://en.benagen.com/>). HiSeq Xten PE150 Illumina, San Diego, CA, USA was deployed to obtain the Illumina sequencing data, and Oxford Nanopore GridION × 5 Oxford Nanopore Technologies, Oxford, UK, handled the Nanopore sequencing.

Acquirement of *S. Japonica* mitochondrial genome

The mitochondrial genome of *S. japonica* was acquired by a hybrid assembly strategy with second- combined with third-generation sequencing data. GetOrganelle (v 1.7.5) was utilized to assign the mitochondrial genome from about 10 G of sequenced second-generation data [26], and the graphical *S. japonica* mitochondrial genome was obtained. The long-read sequencing data was aligned to the assembled contigs by BWA software and the coverage depth was then determined via Samtools (v 0.9) [27]. Bandage [28] was applied for visualizing the mitochondrial genome. BWA (v 0.7.17) [29] was deployed to depict the Nanopore data to the base sequences and manually

remove the redundant segments of the chloroplast and nuclear. The mitochondrial genome of *S. japonica* with seven ring structures was obtained.

The Geseq software was conducted for the annotation of protein-coding genes (PCGs) from *S. japonica* mitochondrial genome [30], and *Aconitum kusnezoffii*, *Hepatica maxima*, and *A. thaliana* was selected as a reference. Annotation of mitochondrial genomic tRNAs utilizing tRNAscan-SE (v 2.0.2) [31] and the blastn of BLAST (v 2.12.0+) was used to annotate the genomic rRNA based on nucleic acid sequence similarity. Apollo (v. 6. 0) was also used for manually adjusting the position of genes [32].

Relative synonymous codon usage and Ka/Ks analyses

Phy-losuite was used for obtaining the PCGs from *S. japonica*'s mitochondrial genome [4]. The relative synonymous codon usage (RSCU) was calculated and analyzed with MEGA 11 [33]. The synonymous (Ks) and nonsynonymous (Ka) substitution rates of the PCGs in the *S. japonica* mitochondrial genome were examined by three species (*A. kusnezoffii*, *H. maxima*, and *A. thaliana*). TBtools [34] was utilized in this investigation to compute Ka/Ks.

Analysis of repeated sequences

Detection of simple sequence repeats (SSRs) using the Microsatellite Identification tool named as MISA (<https://webblast.ipk-gatersleben.de/misa/>) [35]. The repetition counts of ten, five, four, three, three, and three repeats for mono-, di-, tri-, tetra-, penta-, and hexameric bases were found. Tandem Repeats Finder (v 4.09) (<http://tandem.bu.edu/trf/trf.submit.options.html>) [36] was used with the default parameter to identify tandem repeats with > six bp repeat. The REPuter webserver (<https://bibiserv.cebitec.uni-bielefeld.de/reputer/>) [37] was utilized to determine the forward and reverse repeats, with the minimal repeat size set to 30 bp.

Migration of the mitochondrial genome to the chloroplast genome and RNA editing analysis

Based on the chloroplast genome of *S. japonica* already held by our research group, BLASTN software [38] was used for homologous fragment comparison with *S. japonica* mitochondrial genome to explore the evolution of mitochondrial fragments and the migration of the chloroplast fragments. Circos was used to visualize the comparison results [39]. Under a threshold of 0.2, RNA editing events were predicted depending on the PREP suit online website (<http://prep.unl.edu/>) [37].

Construction of phylogenetic tree and collinearity analysis

To establish the phylogenetic tree, the shared mitochondrial PCGs between *S. japonica* and 26 other species

were utilized. (Table S5 provides full species information). The mitochondrial genome information was downloaded from NCBI. PhyloSuite was utilized to extract the conserved PCGs (*atp1*, *atp4*, *atp6*, *atp8*, *matR*, *rps3*, *cox 2~3*, *ccmB*, *ccmC*, *ccmFC*, *ccmFN*, *nad 1~3*, *nad 5~7*, and *nad9*) [4]. We aligned sequences with the MAFFT algorithm [40]. The evolutionary relationship was analyzed via MRBAYES (v. 3.2.2) [41]. The results of the tree analysis were displayed by ITOL (<https://itol.embl.de/>) [42].

MCscanX was utilized for creating a multiple synteny plot of *S. japonica* with *H. maxima* and *A. kusnezoffii* [43]. The BLASTN results of pairwise comparisons of each mitochondrial genome were obtained via the BLAST program; homologous sequences larger than 500 bp were kept as conserved collinearity blocks. The Maximum-Likelihood technique was used in MEGA 11 (v 11.0.13) to create a phylogenetic tree of the *cox* gene as follows: General Time Reversible Model as adopted model; Gamma Distributed With Invariant Sites (G+I) as Rates among Sites. Table S5 provides full species information.

Real-time quantitative PCR

Using the total RNA extraction kit of plant (Foregene Biotech, Chengdu, China, CodeNo. RE-05011), the experiment was conducted on three more *S. japonica* that are presently being grown in our laboratory from the same source (2.1) (30°68'N, 103°81'E). Then the total RNA was transferred to cDNA utilizing the reagent kit with gDNA Eraser (Foregene Biotech, Chengdu, China, Code No. RT-01032). A quantitative real-time polymerase chain reaction (qRT-PCR) was deployed using the QuantStudio5 real-time PCR system. TB Green® Premix ExTaq™ II (Vazyme Biotech, Beijing, China, Code No. Q711-02) was used as the PCR reagent, during which three technical replications were carried out. The *ACT2* gene sequence (*AT3G18780.2*) [44] and the *GAPDH* gene sequence (*AT3G04120.1*) of *A. thaliana* was used as housekeeping genes [45]. Through the Blast result, the *SjapChr2G00058590.1* corresponds to the *ACT2* gene and *SjapChr10G00234090.1* corresponds to the *GAPDH* gene of *S. japonica*. These gene were used as internal reference for qPCR analysis [8]. The $2^{-\Delta\Delta CT}$ calculations were utilized to quantified gene expression. Table S6 contains a list of the primer sequences that were employed in this investigation.

Verifying recombination events mediated by repetitions

To verify the homologous recombination products supported by Nanopore long reads, we extracted sequences of 300–400 bp flanking the predicted repeat sequences to form a reference sequence and designed specific primers [46]. PCR amplification was performed on a 25 μ L

sample, which contained 12.5 μL $2 \times$ Taq Master Mix (Dye Plus) (Vazyme, Nanjing, China), 9.5 μL ddH₂O, 1 μL of each primer (10 $\mu\text{mol/L}$), and 1 μL DNA. The PCR reaction was carried out under the following conditions: denaturation at 95 $^{\circ}\text{C}$ for 3 min; followed by 35 cycles of 95 $^{\circ}\text{C}$ for 15 s, 58 $^{\circ}\text{C}$ for 15 s, and 72 $^{\circ}\text{C}$ for 2 min; and extension at 72 $^{\circ}\text{C}$ for 5 min. We observed the PCR amplicons by 1.0% agarose gel electrophoresis and sequenced the PCR amplicons using Sanger sequencing technology from Beijing Tsingke Biotech Co., Ltd. (Beijing, China).

Results

Characteristics of the mitochondrial genome of *S. Japonica*

In this investigation, the complete mitochondrial genome of *S. japonica* was obtained. The seven circular structures that make up the mitochondrial genome of *S. japonica* have a total length of 555,117 bp, 46.56% GC content and average depth over 90X (Table S1). The complete

mitochondrial genome of *S. japonica* consists of seven ring structures. In order to verify the accuracy of the assembled structure of *S. japonica*, we performed PCR amplification and Sanger sequencing on 300–400 bp specific primers designed for PCR amplification at both ends of the predicted repeat sequences in the assembled seven ring structures (Figure S1 A). PCR amplification showed that the band length was as expected (Figure S1C, D and E), and Sanger sequencing confirmed the existence of this complex conformation (Figure S2-10). The *S. japonica* contains three rRNA genes, 18 tRNA genes, and 40 PCGs (Fig. 1c). The annotated PCGs in the *S. japonica* mitochondrial genome are grouped into ten categories, as shown in Table 1.

A detailed analysis of the genes of the species can provide the necessary help for the identification of species and can also facilitate the in-depth study of species [47]. For the purpose of better analysis, the mitochondrial

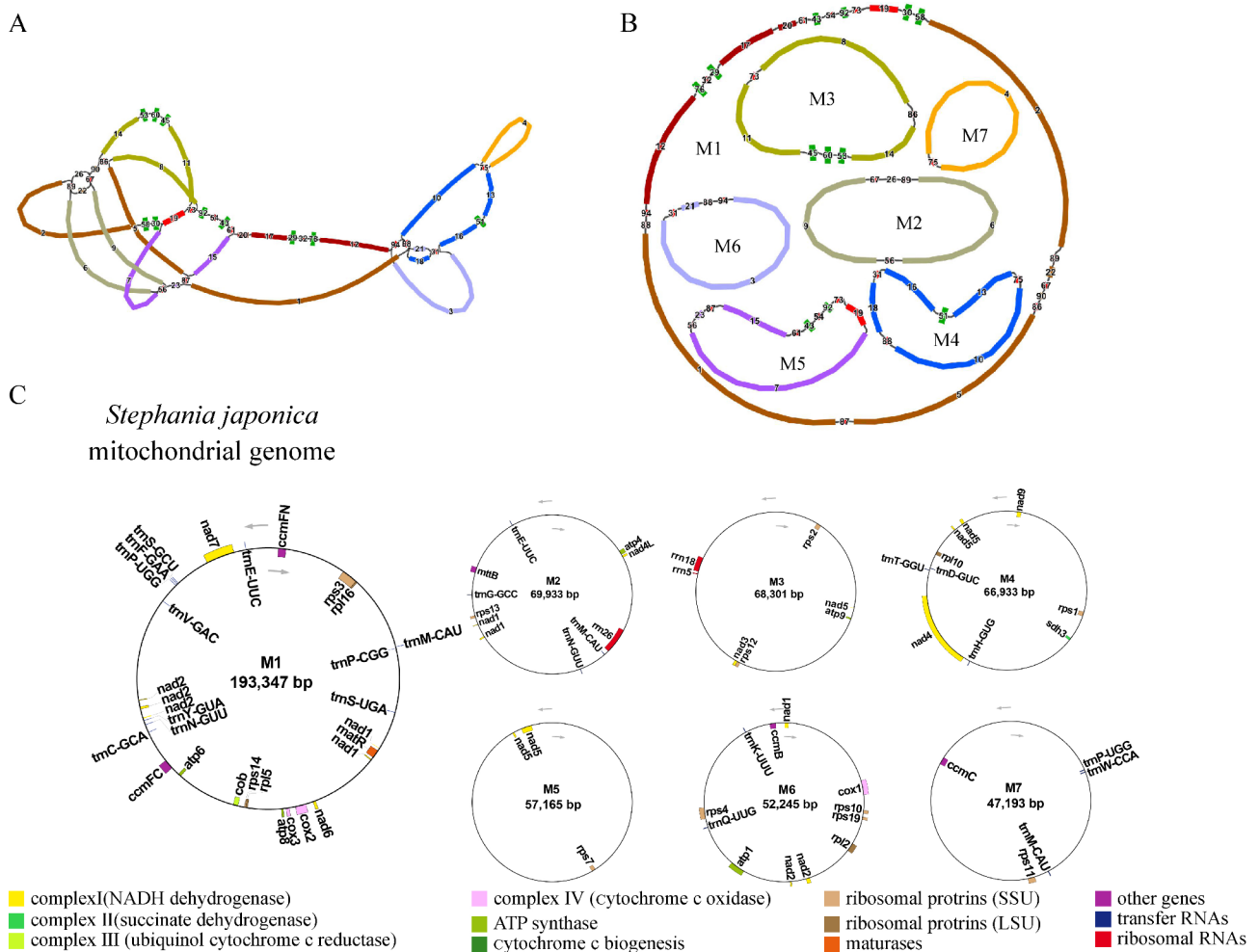


Fig. 1 *S. japonica's* mitochondrial genome structure and annotation. (a) GetOrganelle predicted seven circular contigs in *S. japonica's* mitochondrial genome. (b) The 2D structure of *S. japonica's* mitochondrial genome following the removal of nuclear and artificial chloroplast gene segments. (c) The mitochondrial genome annotations for *S. japonica*. The sequence information in (b) is accordingly marked on each ring. Different colors correspond to different roles of genes

Table 1 The encoding genes of *S. Japonica* mitochondrial genome

Groups	Gene Names
ATP synthase	<i>atp1, atp4, atp6, atp8, atp9</i>
NADH dehydrogenase	<i>nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7, nad9</i>
Cytochrome c biogenesis	<i>cob</i>
Ubiquinol cytochrome c reductase	<i>ccmB, ccmC, ccmFC, ccmFN</i>
Cytochrome c oxidase	<i>cox1, cox2, cox3</i>
Maturases	<i>matR</i>
Transport membrane protein	<i>mttB</i>
Large subunit of ribosome	<i>rpl2, rpl5, rpl10, rpl16</i>
Small subunit of ribosome	<i>rps1, rps2, rps3, rps4, rps7, rps10, rps11, rps12, rps13, rps14, rps19</i>
Succinate dehydrogenase	<i>sdh3</i>
Ribosome RNA	<i>rrn5, rrn18, rrn26</i>
Transfer RNA	<i>trnC-GCA, trnD-GUC, trnE-UUC (x2), trnF-GAA, trnG-GCC, trnH-GUG, trnK-UUU, trnM-CAU (x3), trnN-GUU (x2), trnP-CGG, trnP-UGG (x2), trnQ-UUG, trnS-GCU, trnS-UGA, trnT-GGU, trnV-GAC, trnW-CCA, trnY-GUA</i>

genes of *S. japonica*, *H. maxima*, and *A. kusnezoffii* were compared, and duplication and loss of some genes were found. In *H. maxima*, the mitochondrial genome was found for *atp1* replication and deletion of *rps14* and *rps10* in contrast to the other. In *S. japonica*, the mitochondrial genome was relatively complete with the deletion of only one gene (*sdh4*) compared with the other two

species, while four genes were found absent in *A. kusnezoffii* (*rpl2*, *rps2*, *rps11*, and *rps19*).

Codon usage analysis among protein-coding genes

The ratio of synonymous codon usage frequency to expected frequency is known as RSCU. The predicted frequency represents the average usage frequency of all codons that encode a specific amino acid. A codon's relative high use bias is shown when its RSCU value is more than one. As shown in Fig. 2, the codon in each amino acid was shown in different colors. For mitochondrial PCGs, there is a general bias in the codon usage, except for the initiation codons AUG and UGG, both of which have RSCU values of 1. For example, tyrosine (Tyr) has the greatest RSCU value of 1.52 and is highly biased for UAU, whereas histidine (His) is biased for CAU and has an RSCU value of 1.5. Additionally, all three show a high codon use bias, with the maximum RSCU of arginine (Arg), glutamine (Gln), and glycine (Gly) values being larger than 1.4.

Repeat sequence analysis

Based on the MISA online service platform, 187 SSRs were found in *S. japonica*, and they were distributed among seven mitochondrial circular molecules: circular molecule 1 (63 SSRs), circular molecule 2 (23 SSRs), circular molecule 3 (29 SSRs), circular molecule 4 (24 SSRs), circular molecule 5 (19 SSRs), circular molecule 6 (13 SSRs), and circular molecule 7 (16 SSRs). The overall distribution detail is shown in Figure S11. Adenine monomeric repeat sequences (27) accounted for 48.2% of

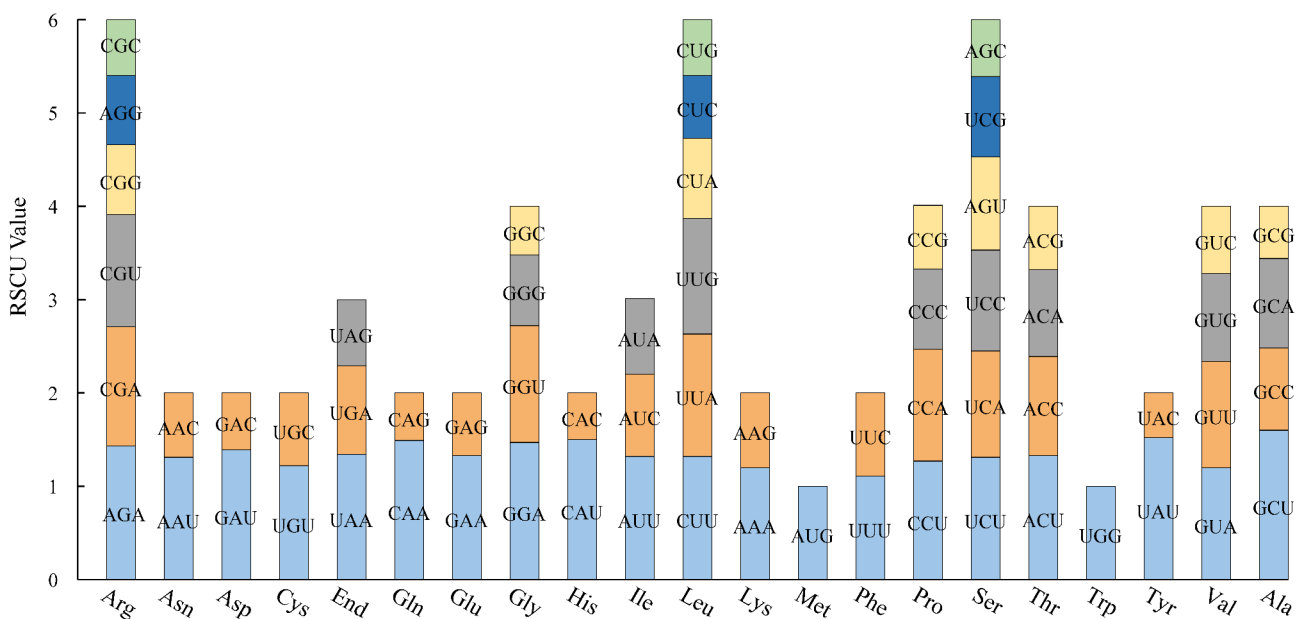


Fig. 2 RSCU in PCGs of *S. japonica*. The height of the bar means the total number of amino acid subtypes, and the fraction of each codon is shown by the different colors

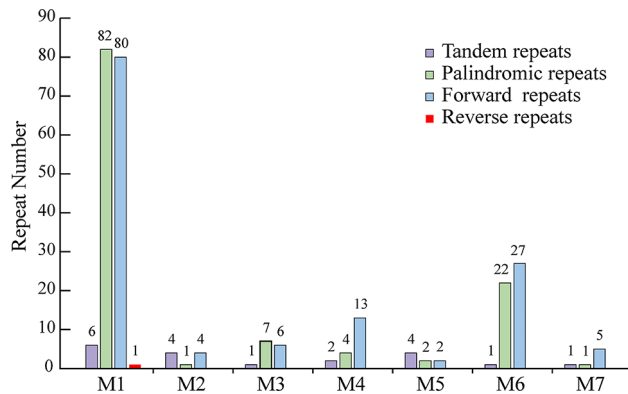


Fig. 3 The number of repetitive sequences in *S. japonica*. Blue denotes forward repetitions, red denotes reverse repeats, purple suggests palindromic repeats, and green represents tandem repeats

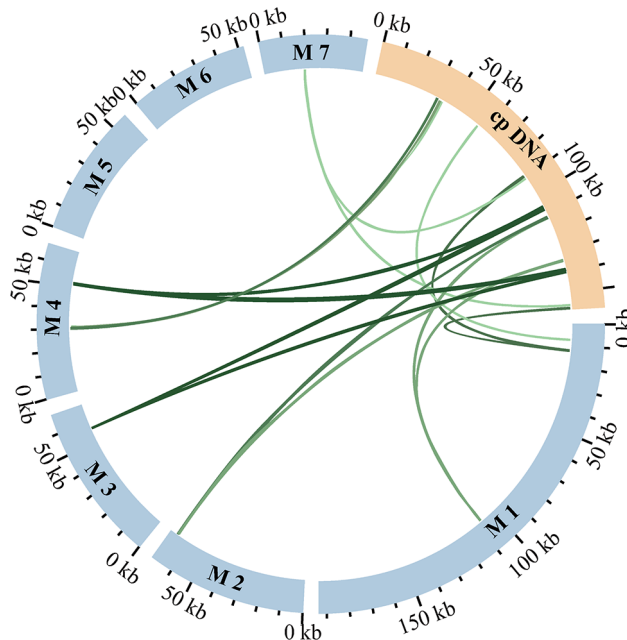


Fig. 4 Mitochondrial genome migration events in *S. japonica*. The genomes of the mitochondria and chloroplasts are represented by the blue and yellow arcs. The green lines between the arcs signify homologous genomic segments; the deeper color denotes a more similar section

the 59 monomeric SSRs, whereas 49.2% of all SSRs were composed of monomeric and dimeric repeats. Additionally, the proportion of tetrameric repeat sequences (61) was high and accounted for 32.6% of the total SSRs. However, no SSR sequences containing hexameric repeats were found within the assembled mitochondrial genome. Tandem repeats are widely observed in prokaryotic and eukaryotic genomes [48–50]. Totally 19 tandem repeats were determined (Table S3), 7 of which had a 100% match rate. Additionally, dispersed repeats in the mitochondrial genome were examined. Consequently, 257 duplicates with lengths higher than 30 were noted in total (shown in Fig. 3); the most scattered repeats were found

in circular molecule 1 (163 bp), while the least scattered repeats (four bp) were found in circular molecule 5.

DNA migration from chloroplast to mitochondria

A total of 16 homologous fragments totaling 4,831 bp in length were discovered to migrate from chloroplast to mitochondria in accordance with sequence similarity and the chloroplast genome. These fragments account for 0.87% of the complete mitochondrial genome. In particular, the two ultralong fragments were found in circular molecule 4, each 916 bp in length, and both contain parts of the tRNA gene (*tRNA-UGG*). A total of five intact tRNA genes (*trnM-CAU*, *trnN-GUU*, *trnI-CAU*, *trnD-GUC*, and *trnT-GGU*) were found in 16 homologous fragments. It may be evident from the data that two chloroplast PCGs migrated to the mitochondrial genome but lost their integrity during the migration (Table S4, Fig. 4).

Predicting RNA editing

RNA editing, the post-transcriptional bioprocess, routinely improves protein folding [17]. It is widespread and impressive for regulating mitochondrial gene expression among advanced plants. PREP was utilized to predict RNA editing loci in mitochondrial genes of *S. japonica*. The outcome revealed 684 RNA editing loci in total. The *nad4* gene, which includes 55 editing sites, has the most in *S. japonica*'s mitochondrial genome, as seen in Fig. 5. The *rps1* and *rps7* genes had the fewest editing sites in *S. japonica* mitochondrial genes, with only two RNA editing sites predicted, respectively. Among the 684 edited sites, 330 were found at the triplet code's first position, while 331 were found at the triplet code's second base. And 20 edited sites showed special cases where both the first position and second base were changed by editing, such as the substitution of phenylalanine (TTT / TTC) for the original proline (CCT / CCC). The properties of the original amino acids can also be changed after RNA editing, with 8.5% of amino acids performing hydrophobic to hydrophilic conversion, whereas 47.22% perform hydrophilic to hydrophobic conversion. Additionally, our findings demonstrated a leucine propensity in amino acids of projected editing codons following RNA editing, which is corroborated by the finding that 44.59% of amino acids underwent leucine conversion following RNA editing (Table 2). Additionally, we found that three genes (*ccmFC*, *atp6*, and *rps11*) have altered open reading frames as a result of RNA editing that produces a termination codon.

Phylogenetic analysis and collinearity analysis

Phylogenetic trees were constructed for 26 species under six orders of angiosperms (shown in Fig. 6), as well as *S. japonica*, according to the DNA sequences of 19 of the shared PCGs (*ccmB*, *ccmC*, *ccmFC*, *atp1*, *atp4*, *atp6*, *atp8*,

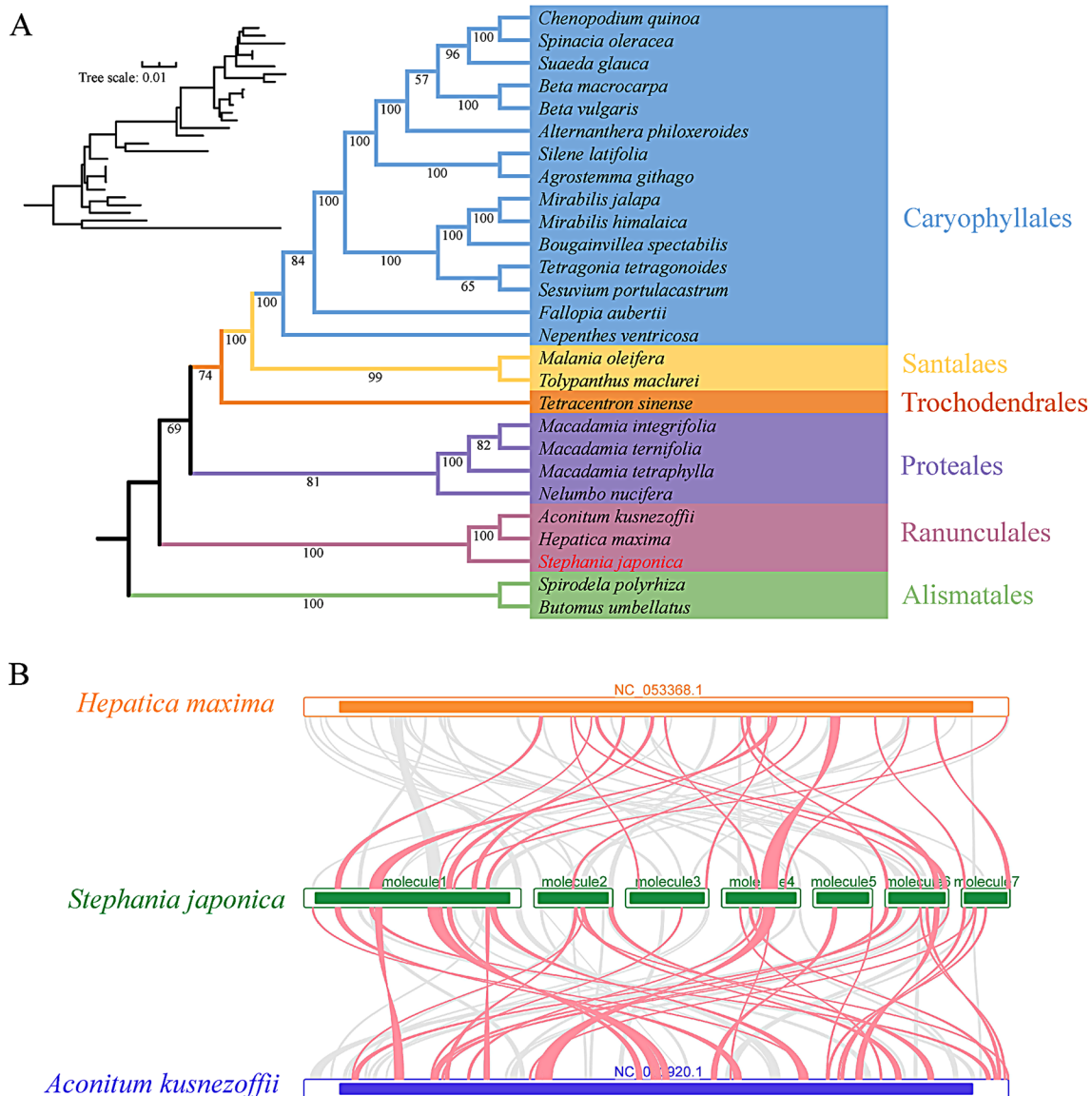


Fig. 6 The evolutionary analysis of *S. japonica*. **(A)** Phylogenetic tree analysis based on 19 conserved proteins encoded by *S. japonica*. **(B)** Analysis of colinearity between *S. japonica* and other two Ranunculales species

Among the 39 PCGs of *S. japonica*, six genes have Ka/Ks values greater than one compared to *H. maxima* and less than one (Fig. 7).

Evolution and expression analysis of the *cox* genes

The phylogenetic tree of the *cox* genes was consistent with that built by 19 conserved genes of the entire species (Figs. 6A and 8A). The *cox* gene expression levels in different tissues of *S. japonica* were also analyzed (root, stem, leaf, and bud). The relative expression levels of the *cox* genes verified tissue specificity. That is, they had superior expressions in the root than in other tissues, which may be connected to how the plant is growing [51]. The root of *S. japonica* also had higher metabolic activity during growth and required more energy, which

could explain the higher *cox* expression in these tissues [52].

Discussion

Decoding complete genome sequence information offers insights into genetic diversity, enabling the precise identification of gene variations that are crucial for understanding the genetic basis of traits, diseases, and evolutionary process. Plant cells possess three distinct genomes, containing the nuclear, chloroplast, as well as mitochondrial genomes, offers valuable resources for investigating plant evolution [53]. Plant mitochondrial genomes exhibit greater complexity compared to animal mitochondrial genomes due to many factors, including genome differences in size and repetitive sequences

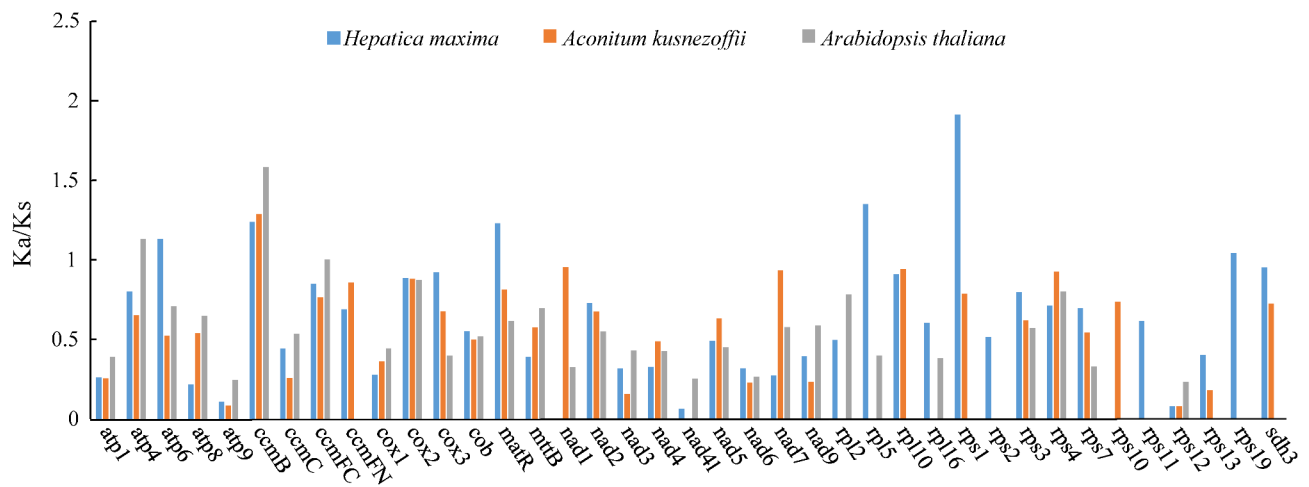


Fig. 7 The Ka/Ks values of 39 PCGs of *S. japonica* versus three species. Among them, *H. maxima* and *A. kusnezoffii* belong to Ranunculales

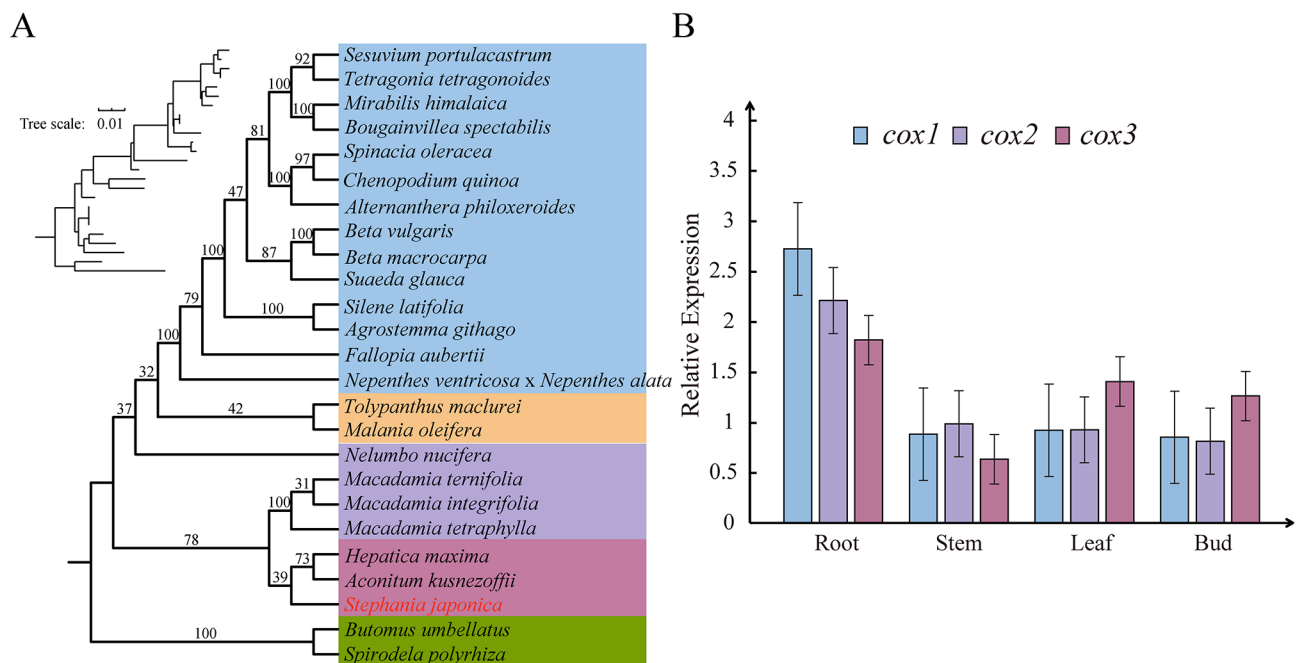


Fig. 8 Evolution and expression analysis of *cox* genes. **A.** Phylogenetic tree analysis based on *cox* genes with 24 species. The color of the box indicates categories of species: Caryophyllales in blue, Santalales in yellow, Proteales in perilla, Ranunculales in red, and Alismatales in green. **B.** Real-time quantification of *cox* genes expression in different tissues of *S. japonica*

[54]. In early research on plant mitochondrial genomes, they were reported to exhibit a single circular structure analogous to that of animal mitochondrial genomes [55]. However, numerous studies have shown that their actual structures can also possess a variety of branched, linear, or mixed forms of genomic organization [56–58]. A high-quality *S. japonica* genome has been assembled with a size of 643.4 Mb [8]. Through this investigation, we investigated the mitochondrial genome of *S. japonica*. *S. japonica* shows marked structural specificity compared to its relatives, being a complex structure with multiple branches, whereas the mitochondrial structure

of two Ranunculales species presents a classical one-loop structure [59, 60]. Multiring structures have also been discovered in the mitochondrial genomes of various species, including ferns, basal angiosperms, monocots, and dicots. For instance, the mitochondrial genome of *Fritillaria ussuriensis* Maxim consists of 12 circular structures, among which five rings contain only a single functional gene [56]. Similarly, the mitochondrial genome of *Angelica dahurica* also consists of 12 circular structures [61]. These results indicate that plant mitochondrial genomes are diverse and complex in terms of structure, size, and gene content.

RNA editing is critical for most events, such as generating initiation codons, termination codons, or changing codons to specify conserved amino acid positions. In plants, RNA editing of plant organ genomes frequently results in C-to-U/U-to-C transitions [62], and in the present study, it was all C to U editing in the mitochondrial genome of *S. japonica*, and there were 687 editing loci in the mitochondrial genes of *H. maxima* with the same orders as *S. japonica*, which were also for C to U editing [60]. Fewer editing sites (441) were revealed in *A. thaliana* compared to *S. japonica*, but the changes were consistent with the overall increase in overall hydrophobicity [62]. RNA editing is indispensable for regulating various physiological processes, such as plasmid progress and the response to stress [63]. Alterations in RNA editing are common in stress responses [64–66]. Furthermore, we found that three genes (*ccmFC*, *atp6*, and *rps11*) had altered open reading frames due to RNA editing to generate termination codons.

Moreover, only *sdh3* was found in the mitochondria of *S. japonica*. The *sdh4* gene was not found in *S. japonica*, in contrast to *H. maxima* and *A. kusnezoffii*. It has been demonstrated that *sdh* contributed to the production of ROS in mitochondria, regulating plant development and stress response in *A. thaliana* and Rice [67]. Whether it is related to the change in the response of *S. japonica* to its growth environment is worthy of further study. As the respiratory chain's final electron acceptor, *cox* is vital for oxidative phosphorylation and the conversion of O₂ to H₂O [68]. Inhibition of *cox17* gene expression in *A. thaliana* can reduce the response to salt stress [61]. The *cox* genes are usually thought to function in the evolution of species and the growth and progress of plants. For example, the *cox* deficient mutant of *A. thaliana* has problems in seed germination and root growth retardation [69]. And *cox* expression is specific and preferentially expressed in tissues taking high energy requirements, such as the root meristem [51]. The findings verify that the *cox* gene was found to be highly expressed in roots in *S. japonica*. To make species identification easier, some researchers have even suggested utilizing the organelle genome as a super DNA barcode [70]. The *cox* genes are relatively conserved in evolution and widely considered to be important enzymes involved in respiration and biological processes. It has served as a species identification DNA barcode and has been well applied for animal species identification [71]. In this investigation, the *cox* genes of *S. japonica* were utilized to build the phylogenetic tree, in agreement with one with 19 conserved genes. This result further shows the reliability of the evolutionary relationship and the conservation of *cox* genes, proving that the *cox* gene is also relatively conservative among plant species, although the mitochondrial genome is more complex. Although the application of the *cox* gene

for the determination of closely related plants has certain limitations, it is undeniable that the *cox* gene is still used to identify plant species. This study looked more closely at the *cox* gene expression pattern in *S. japonica*. The *cox* genes in *S. japonica* were expressed in the roots, stem, leaf, and bud. Among these tissues, *cox* gene expression was highest in roots. This might be related to the growth stage of the plant or the oxygen content in the air [72, 73].

Conclusions

This investigation successfully obtained the whole mitochondrial genome of *S. japonica*, providing a valuable resource for a better understanding of *Stephania* species. We conducted an interspecies study on the *cox* gene and believe that it can still be used as an alternative solution for species identification. We also further investigated the *cox* gene expression in different tissues, which may assist in research related to plant energy metabolism. This lays the foundation for mitochondrial-based species identification techniques, such as DNA barcoding, as well as research related to energy metabolism.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-11359-6>.

Supplementary Material 1

Supplementary Material 2

Author contributions

CSL and LL designed and oversaw the study. LL offered both experimental supplies and financial support. QT, LXJ, LZ, SZH, and WY will engage in data analysis. WY to experiment. WY, SZH, and LZ charts to create a first draft.

Funding

This work is supported by the talented person scientific research start funds subsidization project of Chengdu University of Traditional Chinese Medicine (030040016/030).

Data availability

All the mitochondrial genomes mentioned in this study can be available in NCBI (<https://www.ncbi.nlm.nih.gov/>). And the accession numbers can be found in the supplementary material. The mitochondrial genome of *S. japonica* can be available at Genbank under accession number OR500009. All raw sequencing data were deposited under the National Center for Biotechnology Information (NCBI) GenBank accession number PRJNA888087.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Collection of plant material

The collection of plant material complies with relevant institutional, national, and international guidelines and legislation.

Competing interests

The authors declare no competing interests.

Received: 15 November 2024 / Accepted: 11 February 2025

Published online: 24 February 2025

References

1. Fan H, He S-T, Han P, Hong B, Liu K, Li M, et al. Cepharanthine: a promising old drug against SARS-CoV-2. *Adv Biol (Weinh)*. 2022;6:e2200148.
2. Ohashi H, Watashi K, Saso W, Shionoya K, Iwanami S, Hirokawa T, et al. Potential anti-COVID-19 agents, cepharanthine and nelfinavir, and their usage for combination treatment. *iScience*. 2021;24:102367.
3. Dong J-W, Li X-J, Cai L, Shi J-Y, Li Y-F, Yang C, et al. Simultaneous determination of alkaloids dicentrine and sinomenine in *Stephania epigeae* by 1H NMR spectroscopy. *J Pharm Biomed Anal*. 2018;160:330–5.
4. Zhang D, Gao F, Jakovčić I, Zou H, Zhang J, Li WX, et al. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour*. 2020;20:348–55.
5. Gu X, Hao D, Xiao P. Research progress of Chinese herbal medicine compounds and their bioactivities: fruitful 2020. *Chin Herb Med*. 2022;14:171–86.
6. Kim DE, Min JS, Jang MS, Lee JY, Shin YS, Song JH, et al. Natural bis-benzylisoquinoline alkaloids-tetrandrine, fangchinoline, and cepharanthine, inhibit human coronavirus OC43 infection of MRC-5 human lung cells. *Biomolecules*. 2019;9:696.
7. Yang H, Wang Y, Liu W, He T, Liao J, Qian Z, et al. Genome-wide pan-GPCR cell libraries accelerate drug discovery. *Acta Pharm Sinica B*. 2024. <https://doi.org/10.1016/j.apsb.2024.06.023>.
8. Leng L, Xu Z, Hong B, Zhao B, Tian Y, Wang C, et al. Cepharanthine analogs mining and genomes of *Stephania* accelerate anti-coronavirus drug discovery. *Nat Commun*. 2024;15:1537.
9. Liu X, Shen S, Wang Y, Sun S, Yu T, Fu Y, et al. The genome of *Stephania Japonica* provides insights into the biosynthesis of cepharanthine. *Cell Rep*. 2024;43:113832.
10. Shang H, Lu Y, Xun L, Wang K, Li B, Liu Y et al. Genome assembly of *Stephania Longa* provides insight into cepharanthine biosynthesis. *Front Plant Sci*. 2024;15.
11. Wu L-L, Geng Y-M, Zheng L-P. Characterization and phylogenetic analysis of the Chloroplast genomes of *Stephania Japonica* var. *Timoriensis* and *Stephania Japonica* var. *Discolor*. *Genes*. 2024;15:877.
12. Wang J, Wang J, Shang M, Dai G, Liao B, Zheng J, et al. Comparatively analyzing of chloroplast genome and new insights into phylogenetic relationships regarding the genus *Stephania*. *Gene*. 2024;893:147931.
13. Kang L, Qian L, Zheng M, Chen L, Chen H, Yang L, et al. Genomic insights into the origin, domestication and diversification of *Brassica juncea*. *Nat Genet*. 2021;53:1392–402.
14. Liu H, Zhao W, Hua W, Liu J. A large-scale population based organelle pan-genomes construction and phylogeny analysis reveal the genetic diversity and the evolutionary origins of chloroplast and mitochondrion in *Brassica napus* L. *BMC Genomics*. 2022;23:339.
15. Bonora M, De Marchi E, Patergnani S, Suski JM, Celsi F, Bononi A, et al. Tumor necrosis factor- α impairs oligodendroglial differentiation through a mitochondria-dependent process. *Cell Death Differ*. 2014;21:1198–208.
16. Cavalier-Smith T. The origin of nuclei and of eukaryotic cells. *Nature*. 1975;256:463–8.
17. Cheng Y, He X, Priyadarshani SVGN, Wang Y, Ye L, Shi C, et al. Assembly and comparative analysis of the complete mitochondrial genome of *Suaeda glauca*. *BMC Genomics*. 2021;22:167.
18. Ogihara Y, Yamazaki Y, Murai K, Kanno A, Terachi T, Shiina T, et al. Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. *Nucleic Acids Res*. 2005;33:6235–50.
19. Li J, Li J, Ma Y, Kou L, Wei J, Wang W. The complete mitochondrial genome of okra (*Abelmoschus esculentus*): using nanopore long reads to investigate gene transfer from chloroplast genomes and rearrangements of mitochondrial DNA molecules. *BMC Genomics*. 2022;23:481.
20. Waltz F, Giegé P. Striking diversity of mitochondria-specific translation processes across eukaryotes. *Trends Biochem Sci*. 2020;45:149–62.
21. Yu R, Sun C, Zhong Y, Liu Y, Sanchez-Puerta MV, Mower JP, et al. The minicircular and extremely heteroplasmic mitogenome of the holoparasitic plant *Rhopalocnemis phalloides*. *Curr Biol*. 2022;32:470–e4795.
22. Møller IM. What is hot in plant mitochondria? *Physiol Plant*. 2016;157:256–63.
23. Zancani M, Braidot E, Filippi A, Lippe G. Structural and functional properties of plant mitochondrial F-ATP synthase. *Mitochondrion*. 2020;53:178–93.
24. Mazzucotelli E, Mastrangelo AM, Crosatti C, Guerra D, Stanca AM, Cattivelli L. Abiotic stress response in plants: when post-transcriptional and post-translational regulations control transcription. *Plant Sci*. 2008;174:420–31.
25. Ciarmiello LF, Pontecorvo G, Piccirillo P, De Luca A, Carillo P, Kafantaris I, et al. Use of Nuclear and mitochondrial single nucleotide polymorphisms to Characterize English Walnut (*Juglans regia* L.) Genotypes. *Plant Mol Biol Rep*. 2013;31:1116–30.
26. Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. 2020;21:241.
27. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *GigaScience*. 2021;10:giab008.
28. Wick RR, Schultz MB, Zobel J, Holt KE, Bandage. Interactive visualization of de novo genome assemblies. *Bioinformatics*. 2015;31:3350–2.
29. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*. 2009;25:1754–60.
30. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 2017;45:W6–11.
31. Lowe TM, Eddy SR. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 1997;25:955–64.
32. Lewis SE, Searle SMJ, Harris N, Gibson M, Lyer V, Richter J, et al. Apollo: A sequence annotation editor. *Genome Biol*. 2002;3:RESEARCH0082.
33. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol*. 2021;38:3022–7.
34. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant*. 2020;13:1194–202.
35. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics*. 2017;33:2583–5.
36. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999;27:573–80.
37. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R, REPuter. The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res*. 2001;29:4633–42.
38. Chen Y, Ye W, Zhang Y, Xu Y. High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic Acids Res*. 2015;43:7762–8.
39. Zhang H, Meltzer P, Davis S. RCircos: an R package for circos 2D track plots. *BMC Bioinformatics*. 2013;14:244.
40. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80.
41. F R, M T, van der P, DI M, A A. D, S H, MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61.
42. Letunic I, Bork P. Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*. 2019;47:W256–9.
43. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012;40:e49.
44. Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible W-R. Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol*. 2005;139:5–17.
45. Kram BW, Xu WW, Carter CJ. Uncovering the Arabidopsis thaliana nectary transcriptome: investigation of differential gene expression in floral nectariferous tissues. *BMC Plant Biol*. 2009;9:92.
46. Zhou Q, Ni Y, Li J, Huang L, Li H, Chen H, et al. Multiple configurations of the plastid and mitochondrial genomes of *Caragana spinosa*. *Planta*. 2023;258:98.
47. Chen S, Li Z, Zhang S, Zhou Y, Xiao X, Cui P, et al. Emerging biotechnology applications in natural product and synthetic pharmaceutical analyses. *Acta Pharm Sin B*. 2022;12:4075–97.
48. Tautz D, Renz M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res*. 1984;12:4127–38.
49. Katti MV, Ranjekar PK, Gupta VS. Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol Biol Evol*. 2001;18:1161–7.
50. Zhao Z, Guo C, Sutharzan S, Li P, Echt CS, Zhang J, et al. Genome-wide analysis of tandem repeats in plants and green algae. (Bethesda). 2014;G3:4:67–78.
51. Mansilla N, Racca S, Gras DE, Gonzalez DH, Welchen E. The complexity of mitochondrial complex IV: an update of cytochrome c oxidase biogenesis in plants. *Int J Mol Sci*. 2018;19:662.

52. Kubo N, Arimura S-I, Tsutsumi N, Kadowaki K-I, Hirai M. Isolation and characterization of the pea cytochrome c oxidase vb gene. *Genome*. 2006;49:1481–9.
53. Arimura S, Nakazato I. Genome Editing of Plant Mitochondrial and chloroplast genomes. *Plant Cell Physiol*. 2024;65:477–83.
54. Ma Q, Wang Y, Li S, Wen J, Zhu L, Yan K, et al. Assembly and comparative analysis of the first complete mitochondrial genome of *Acer Truncatum* Bunge: a woody oil-tree species producing nervonic acid. *BMC Plant Biol*. 2022;22:29.
55. Kitazaki K, Kubo T. Cost of having the largest mitochondrial genome: evolutionary mechanism of plant mitochondrial genome. *J Bot*. 2010;2010:620137.
56. Xie P, Wu J, Lu M, Tian T, Wang D, Luo Z, et al. Assembly and comparative analysis of the complete mitochondrial genome of *Fritillaria Ussuriensis* Maxim. (Liliales: Liliaceae), an endangered medicinal plant. *BMC Genomics*. 2024;25:773.
57. Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, et al. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol Genet Genomics*. 2002;268:434–45.
58. Han F, Bi C, Zhao Y, Gao M, Wang Y, Chen Y. Unraveling the complex evolutionary features of the *Cinnamomum camphora* mitochondrial genome. *Plant Cell Rep*. 2024;43:183.
59. Li S-N, Yang Y-Y, Xu L, Xing Y-P, Zhao R, Ao W-L, et al. The complete mitochondrial genome of *Aconitum Kusnezoffii* rchb. (ranales, ranunculaceae). *Mitochondrial DNA B Resour*. 2021;6:779–81.
60. Park S, Park S. Large-scale phylogenomics reveals ancient introgression in Asian hepatica and new insights into the origin of the insular endemic *Hepatica maxima*. *Sci Rep*. 2020;10:16288.
61. Li Y-Y, Liu Y-Y, Zeng X, Wu P, Li Q-M, Guo S-X, et al. Complete mitochondrial genome of *Angelica Dahurica* and its implications on evolutionary analysis of complex mitochondrial genome architecture in Apiaceae. *Front Plant Sci*. 2024;15:1367299.
62. Giegé P, Brennicke A. RNA editing in arabidopsis mitochondria effects 441 C to U changes in ORFs. *Proc Natl Acad Sci U S A*. 1999;96:15324–9.
63. Wang D, Meng S, Su W, Bao Y, Lu Y, Yin W, et al. Genome-wide analysis of multiple organellar RNA editing factor family in poplar reveals evolution and roles in drought stress. *Int J Mol Sci*. 2019;20:1425.
64. Ramadan A, Alnufaei AA, Fiaz S, Khan TK, Hassan SM. Effect of salinity on ccmfn gene RNA editing of mitochondria in wild barley and uncommon types of RNA editing. *Funct Integr Genomics*. 2023;23:50.
65. Ramadan AM, Said OAM, Abushady AM. Salinity stress reveals three types of RNA editing sites in mitochondrial nad7 gene of wild barley both in silico and in qRT-PCR experiments. *Theor Exp Plant Physiol*. 2022;34:13–22.
66. Shahid S, Sher MA, Ahmad F, Rehman S, ur, Farid B, Raza H, et al. Prediction of RNA editing sites and genome-wide characterization of PERK gene family in maize (*Zea mays* L.) in response to drought stress. *J King Saud Univ - Sci*. 2022;34:102293.
67. Jardim-Messeder D, Caverzan A, Rauber R, de Souza Ferreira E, Margis-Pinheiro M, Galina A. Succinate dehydrogenase (mitochondrial complex II) is a source of reactive oxygen species in plants and regulates development and stress responses. *New Phytol*. 2015;208:776–89.
68. Toffaletti DL, Del Poeta M, Rude TH, Dietrich F, Perfect JR. Regulation of cytochrome c oxidase subunit 1 (COX1) expression in cryptococcus neoformans by temperature and host environment. *Microbiol (Reading)*. 2003;149:1041–9.
69. Dahan J, Tcherkez G, Macherel D, Benamar A, Belcram K, Quadrado M, et al. Disruption of the CYTOCHROME C OXIDASE DEFICIENT1 gene leads to cytochrome c oxidase depletion and reorchestrated respiratory metabolism in *Arabidopsis*1[C][W]. *Plant Physiol*. 2014;166:1788–802.
70. Chen S, Yin X, Han J, Sun W, Yao H, Song J, et al. DNA barcoding in herbal medicine: retrospective and prospective. *Ywfxxb*. 2023;13:431–41.
71. Rodrigues MS, Morelli KA, Jansen AM. Cytochrome c oxidase subunit 1 gene as a DNA barcode for discriminating *Trypanosoma* *Cruzi* DTUs and closely related species. *Parasit Vectors*. 2017;10:488.
72. Zhu D, Li X, Tian Y. Mitochondrial-to-nuclear communication in aging: an epigenetic perspective. *Trends Biochem Sci*. 2022;47:645–59.
73. Mottis A, Herzig S, Auwerx J. Mitocellular communication: shaping health and disease. *Science*. 2019;366:827–32.

Publisher's note

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