Comparative analysis of the chloroplast and mitochondrial genomes of *Saposhnikovia divaricata* revealed the possible transfer of plastome repeat regions into the mitogenome

Yang Ni¹, Jingling Li¹, Haimei Chen¹, Jingwen Yue², Pinghua Chen²* and Chang Liu¹*

Abstract

**Background:** *Saposhnikovia divaricata* (Turcz.) Schischk. is a perennial herb whose dried roots are commonly used as a source of traditional medicines. To elucidate the organelle-genome-based phylogeny of *Saposhnikovia* species and the transfer of DNA between organelle genomes, we sequenced and characterised the mitochondrial genome (mitogenome) of *S. divaricata*.

**Results:** The mitogenome of *S. divaricata* is a circular molecule of 293,897 bp. The nucleotide composition of the mitogenome is as follows: A, 27.73%; T, 27.03%; C, 22.39%; and G, 22.85. The entire gene content is 45.24%. A total of 31 protein-coding genes, 20 tRNAs and 4 rRNAs, including one pseudogene (*rpl16*), were annotated in the mitogenome. Phylogenetic analysis of the organelle genomes from *S. divaricata* and 10 related species produced congruent phylogenetic trees. Selection pressure analysis revealed that most of the mitochondrial genes of related species are highly conserved. Moreover, 2 and 46 RNA-editing sites were found in the chloroplast genome (cpgenome) and mitogenome protein-coding regions, respectively. Finally, a comparison of the cpgenome and the mitogenome assembled from the same dataset revealed 10 mitochondrial DNA fragments with sequences similar to those in the repeat regions of the cpgenome, suggesting that the repeat regions might be transferred into the mitogenome.

**Conclusions:** In this study, we assembled and annotated the mitogenome of *S. divaricata*. This study provides valuable information on the taxonomic classification and molecular evolution of members of the family Apiaceae.

**Keywords:** De novo assembly, Organelle genomes, Phylogenetic analysis, DNA transfer, Selective pressure analysis

© The Author(s) 2022. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.
and Shandong provinces [3]. *S. divaricata* is resistant to salinity, cold and drought, and it is often grown as a sand-fixing plant in the dry areas of northwest China [4]. Moreover, *S. divaricata* is usually used as a medicine to treat colds, arthralgia, headaches and other diseases. Thus far, over 100 compounds have been isolated from *S. divaricata*, including abundant chromones, coumarins, acid esters and polyacetylenes, which are potential active components for the treatment of diseases of the immune, nervous and respiratory systems [5]. Exploring the organelle genomes of *S. divaricata* will help us classify *Saposhnikovia* species and provide a genetic resource for further study.

Organelle genomes are critical in sustaining an organism's growth and development. Like the nuclear genome, the plant organelle genome has various strategies to repair DNA damage and maintain the integrity of the genetic material to withstand the damage caused by genotoxic stresses [6]. Organelle genomes have been extensively analysed to understand a taxon’s classification and evolution. To date, 6,804 complete chloroplast genomes (cpgenomes) and 433 plant mitochondrial genomes (mitogenomes) have been released in the GenBank Organelle Genome database (https://www.ncbi.nlm.nih.gov/genome/browse/; last accessed on February 25, 2022) [7]. The number of sequenced mitogenomes is fewer than that of the cpgenomes probably because of the complex structures of the former, which resulted from the violent redox reactions that accompanied the rearrangement of some DNA fragments [8, 9]. Previous studies have found that the evolution of the mitogenome affects cytoplasmic male sterility (CMS), a phenomenon with important implications to plant breeding genetics [10].

The cpgenome of *S. divaricata* has been reported in a previous study [11]. However, no study has described its mitogenome and the exchange of DNA between the cpgenome and the mitogenome. In this study, we de novo assembled the cpgenome and the mitogenome of *S. divaricata*. We report the mitogenome of this species for the first time and compared the differences between the cpgenome assembled herein and the one published in a previous study. Moreover, we systematically analysed the gene content, repeat sequences, selective pressure and RNA-editing sites. Finally, we explored the phylogenetic relationships among *S. divaricata* and 10 related species. This study provides valuable information on the taxonomic classification, molecular evolution and breeding of *Saposhnikovia* species.

**Results**

**General features of the organelle genomes of *S. divaricata***

The cpgenome (MZ089852) is 147,832 bp and has a typical quadripartite structure consisting of a pair of inverted repeats (IR) regions of 18,653 bp, a large single-copy (LSC) region of 93,202 bp and a small single-copy (SSC) region of 17,324 bp (Fig. 1A). The gene contents (GC) of the IR, LSC and SSC regions are 44.58, 35.94 and 30.85%, respectively. The cpgenome encodes 114 unigenes, including 80 protein-coding genes, 30 tRNAs and 8 rRNA genes (Table S1). Eighteen genes have one intron (trnK-UUU, rps16, trnG-UCC, atpF, rpoC1, trnL-UAA, trnV-UAC, petB, petD, rpl16, rpl2, ndhB, trnL-GAU, trnA-UGC, ndhA, trnT-GGU and ndhB), and two genes (ycf3 and clpP) contain two introns. The GCs of the coding sequences, tRNAs and rRNA sequences are 37.95, 53.55 and 55.22%, respectively. We then compared the cpgenome and the one published before using dotplot. The results showed that the two cpgenomes were highly collinear (Figure S3), and differed in two one-base indels (Figure S4).

Several mitogenomes of Apiaceae have been previously reported, which are from *Daucus carota* subsp. *sativus* (281,132 bp) and *Bupleurum falcatum* (463,792 bp) [12, 13]. The *S. divaricata* mitogenome (MZ128146) is a circular molecule of 293,897 bp (Fig. 1B). The nucleotide composition of the whole mitogenome is as follows: A, 27.73%; T, 27.03%; C, 22.39%; and G, 22.85. The entire GC is 45.24%, similar to that of *D. carota* subsp. *sativus* (45.41%). A total of 31 protein-coding genes, 20 tRNAs and 4 rRNA genes, including one pseudogene (rpl16), were annotated in the mitogenome (Table 1). There were eight genes contain introns and the composition of introns were shown in the Table S3.

**Repeat analysis**

Microsatellites are also known as simple sequence repeats (SSRs). They are mono-, di-, tri-, tetra- or pentanucleotide DNA units and mostly appear in eukaryotes [14]. A total of 76 and 41 SSRs were detected in the cpgenome and the mitogenome, respectively (Fig. 2A, Tables S4 and S5). In the cpgenome, the most abundant SSRs have a single-nucleotide repeat unit, particularly A/T. The number of A/T repeat units accounts for 88.4% of all identified SSR repeats. However, the SSRs are evenly distributed among the various types in the mitogenome. A total of 12, 4, 6, 16 and 3 SSRs have mono-, di-, tri-, tetra- and pentanucleotide repeat units, respectively. The most abundant SSRs in the mitogenome have a tetranucleotide repeat unit, representing 39.0% of all the repeat number. These SSRs could be potential identification markers for *S. divaricata*.

Tandemly repeated DNA sequences, which have a unit length longer than 6 bp, are highly dynamic components of genomes [15]. Most repeats are found in intergenic regions, but some are in coding sequences or pseudogenes [16] (Fig. 2B, Tables S6 and S7). A total of 25 and
Fig. 1 The circular maps of the organelle genomes of *S. divaricata*. A The circular map of the cpgenome. B The circular map of the mitogenome. Genomic features transcribed clockwise and counter-clockwise are drawn on the inside and outside of the circle, respectively. Genes are color-coded based on their functional groups. GC content is represented on the inner circle by the dark gray plot.
26 tandem repeats were identified in the cpgenome and the mitogenome, respectively, and these repeats were evaluated further for their potential application in DNA fingerprinting.

Dispersed repeats are essential in generating genetic diversity, and they make valuable contributions to the evolution of plant genomes [17]. There are four kinds of dispersed repeats, namely, forward repeats, reverse repeats, complement repeats and palindromic repeats. In the cpgenome, all four types of dispersed repeats were found. In both genomes, the most abundant and the longest repeats are forward repeats, with the longest fragment being 22,397 bp in the mitogenome. Its number accounts for 34.7% of the total repeats in the mitogenome (Fig. 2, Tables S8 and S9). By contrast, only 33 forward repeats and 24 palindromic repeats were found in the cpgenomes, and most of them are 30–50 bp long.

### Table 1 Gene composition in the mitogenome of *S. divaricata*

| Group of genes                        | Name of genes                        |
|---------------------------------------|---------------------------------------|
| ATP synthase                          | *atp1*, *atp4*, *atp6*, *atp8*, *atp9* |
| Cytochrome c biogenesis                | *ccmB*, *ccmC*, *ccmF*<sup>1</sup>, *ccmF*<sup>n</sup> |
| Ubichinol cytochrome c reductase       | Cob                                   |
| Cytochrome c oxidase                   | *coxt*, *coxt2*, *cox3*               |
| Maturases                             | *matR*                                |
| Transport membrane protein             | *mttB*                                |
| NADH dehydrogenase                    | *nad1*<sup>b</sup>, *nad2*<sup>b</sup>, *nad3*, *nad4*<sup>b</sup>, *nad4L*, *nad5*<sup>b</sup>, *nad6*, *nad7*<sup>b</sup>, *nad9* |
| Large subunit of ribosomal proteins   | *rpl5*, *rpl10*, *rpl16*<sup>a</sup>   |
| Small subunit of ribosomal proteins   | *rps3*, *rps4*, *rps12*, *rps13*, *rps14* |
| Ribosomal RNAs                        | *rrn4.5*, *rrn5*, *rrn18*, *rrn26*    |
| Transfer RNAs                         | *trnY-GUA*, *trnW-CCA*, *trnS-UGA*, *trnQ-UUG*, *trnP-UGG*, *trnN-GUU*, *trnM-CAU*, *trnK-UUU*, *trnL-CAU*, *trnH-GUG*, *trnG-GCC*, *trnF-GAA*, *trnE-UUC*, *trnD-GUC*, *trnC-GCA*, *trnM-CAU*, *trnS-UGA*, *trnR-ACG*<sup>(x 2)</sup> |

<sup>a</sup>Labeled the pseudogenes, <sup>b</sup>Labeled the genes that contain introns

---

**Fig. 2** The Repeat analysis of the *S. divaricata* organelle genomes. **A** The repeat sequences identified in the cpgenome. **B** The repeat sequences identified in the mitogenome. The C1 circle shows the dispersed repeats connected with yellow, blue, purple, and pink arcs from the center going outward. The C2 circle shows the tandem repeats as short bars. The C3 circle shows the microsatellite sequences identified using MISA. The scale is shown on the C3 circle, with intervals of 10 kb.
Sequence similarity between the mitogenome and the cpgenome

A total of 10 groups of mitogenome fragments were identified to likely be derived from the cpgenome according to sequence similarity (Fig. 3, Table S10). These fragments add up to 17,921 bp in length and occupy 6.1% of the mitogenome. We numbered the group from 'I' to 'X'. Group I contains two repetitive sequences of 6813 bp long (GI-a-m: 119569–112,758; GI-b-m: 150,885–144,074). Their sequences are similar to those in the IR regions of the cpgenome (Gl-a-c: 101,995–108,807; Gl-b-c: 139,040–132,228). Group II contains three repeat sequences in the mitogenome (GII-a-m: 71,079–70,221; GII-b-m: 110,682–109,824; GII-c-m: 141,998–141,140). Their sequences are also similar to those in the IR regions of the cpgenome. Groups III and IV contain unique sequences of 104 and 82 bp respectively (GIII-m: 73,257–73,155; GIV-m: 34,749–34,668), similar to the sequences in the IR regions of the cpgenome (Table S10). The six other groups contain only single-copy sequences in the mitogenome and the cpgenome, and they represent 8.063% of the entire homologous DNA sequences between the two genomes. For the repeat direction, if the repeat sequences were in the protein-coding region, we used the sequences in the sense strand. However, if the repeat sequences were in the noncoding regions, we did not specify the direction (Table S10). These similar sequences might have resulted from the transfer of plastome sequences into the mitogenome during evolution.

Phylogenetic analysis

To study the evolution of the organelle genomes of S. divaricata, we conducted a phylogenetic analysis of the organelle genomes of S. divaricata and 10 related species. Two Solanum species were selected as the outgroups. In total, we used the nucleotide sequences of 71 common genes (atpA, atpB, atpE, atpF, atpH, ccsA, cemA, matK, ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, petA, petD, petG, petL, petN, psaA, psaB, psaC, psaL, psaI, psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, rbcL, rpl14, rpl16, rpl20, rpl22, rpl2, rpl32, rpl33, rpl36,

Fig. 3 Comparison of the cpgenome and mitogenome of S. divaricata. The blue and green outer arcs represent the mitogenome (mtDNA) and cpgenome (cpDNA), respectively, and the inner green arcs show the homologous DNA fragments. The scale is shown on the outer arcs, with intervals of 20 kb. The repeat sequences in groups I and II are shown. The sequence name ‘Gl-a-m’ indicates that this sequence belongs to Group I, repeat sequence ‘a’ in the mitogenome (m). GI represent the group I. The letter in the middle represent the sequence identifier in the same group. The letter at the end indicates whether the sequence is a segment of the cpgenome or the mitogenome.
Substitution rates of protein-coding genes
To explore the evolutionary rate of mitochondrial genes, we calculated the nonsynonymous substitution rate (dN) and the synonymous substitution rate (dS) for the 14 shared protein-coding genes. According to the criterion dN/dS > 1, there was likely a positive selection on the atp9 gene has a low dN/dS radio with the smallest variations, suggesting that it is a super-conserved gene that plays a crucial role in the mitogenomes’ functioning (Table S11).

Prediction of RNA-editing sites
The phenomenon of RNA editing has been observed in the chloroplasts of several angiosperm plants [18]. By mapping the transcriptome data to the reference cpgenome and mitogenome, we identified 2 and 75 RNA-editing sites, respectively (Fig. 6, Table S12). The two RNA-editing sites from the cpgenome are located in the protein-coding regions of the rps16 and cdpP genes. For the 75 RNA-editing sites in the mitogenome, 29 and 46 RNA-editing sites are located in the intergenic spacer regions and the protein-coding regions, respectively. These genes include the genes nad4, nad5, nad6, nad8, nad7, cox1, cox3, rpl5, rpl10, rps3, rps7, atp1, atp6, atp8, atp9 and rrm18. In the future, these predicted RNA-editing sites must be experimentally validated. We filtered about 27 Mb RNA reads on the basis of the organelle genomes. In total, we assembled 1220 transcripts via de novo assembly by using the Trinity software. The length of the largest transcript is 4658 bp. By comparing the

Fig. 4 The phylogenetic relationships between S. divaricata and other 10 related plants. a phylogenetic analysis of cpgenomes based on the nucleotide sequences of 71 protein-coding genes from the cpgenome (atpA, atpB, atpE, atpF, atpH, ccsA, cemA, matK, ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhJ, ndhK, petA, petD, petG, petL, petN, psaA, psaB, psaC, psaI, psaJ, psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbK, psbL, psbM, psbN, psbT, rbcL, rpl14, rpl16, rpl20, rpl22, rpl23, rps3, rps16, rpoA, rpoB, rpoC1, rpoC2, rpoJ, rps11, rps12, rps14, rps15, rps16, rps18, rps19, rps2, rps3, rps4, rps7, rps8, ycf2, ycf3, ycf4). b phylogenetic analysis based on the nucleotide sequences of 14 protein-coding genes from the mitogenome (atp1, atp4, atp6, atp9, cob, ccmB, ccmC, cob, matA, nad3, nad4L, nad6, nad9, rps12, rps4). The sequence obtained from this study was highlighted in Bold. Phylogenetic analysis was conducted with the best evolutionary model *TVM + F + I + G4* and *GTR + F + G4* based on Bayesian Information Criterion (BIC) scores for the cpgenomes and mitogenomes, respectively.

rpoA, rpoB, rpoC1, rpoC2, rps11, rps12, rps14, rps15, rps16, rps18, rps19, rps2, rps3, rps4, rps7, rps8, ycf2, ycf3 and ycf4) for cpgenome-based phylogenetic analysis (Fig. 4). By contrast, we utilised 14 common genes (atp1, atp4, atp6, atp9, ccmB, ccmC, cob, matR, nad3, nad4L, nad6, nad9, rps12 and rps4) for the mitogenome-based phylogenetic analysis. The trees built with the cpgenome and the mitogenome clustered S. divaricata and D. carota together. The overall structures of the two trees are identical (Fig. 4).
**Fig. 5** The boxplots of dN/dS values among each mitochondrial gene in the 10 related plants. The "X" axis shows the name of protein-coding genes, and the "Y" axis shows the dN/dS values.

**Fig. 6** The distribution of RNA editing sites across different genes. The "X" axis shows the name of protein-coding genes, and the "Y" axis shows the number of predicted RNA editing sites.
protein-coding genes of the organelle genomes, we found that rpl5, rps14, cox1 and rps7 may have co-transcribed in the 150–180 kb co-linear block and had been retained during evolution (Table S13).

Discussion

In this study, we reported the mitogenome of *S. divaricata* for the first time. The cpgenome assembled from the same sequencing data is 2 bp shorter than the published one [11]. The two indels are located in the introns of trnA-UGC (Figure S4). No similar studies have established yet whether this indel, which is in the IR region, affects the function of the genome. Comparison of the two cpgenomes using dotplot and BLASTn found no rearrangements (Figures S3–4). Phylogenetic analysis using the mitogenome and our newly assembled cpgenome showed congruent results.

We compared the collinearity between the published mitogenomes of related species to obtain genome rearrangement (Fig. 7). Dot plot analysis revealed that *S. divaricata* has a large number of co-linear regions with *D. carota*. The largest co-linear region is approximately 30 kb. We further predicted the possible polycistronic transcript units to determine the possible evolutionary relationships of these co-linear fragments. The results were consistent with those of the phylogenetic analysis. The co-linear blocks of the more distantly related species are small possibly because the structure of plant mitogenomes is extremely dynamic.

Comparison of the cpgenome and mitogenome sequences suggested that a DNA transfer event occurred in the cpgenome IR region (Figure S5). Mitochondrial genomes are often riddled with plastid DNA-derived sequences, called mitochondrial plastid DNAs (MTPTs) [19–21]. We counted the MTPTs in the 10 related species used in our phylogenetic analysis and obtained three conclusions. Firstly, the mitogenome of *S. divaricata* has the largest MTPT sequence (6813 bp), far exceeding the second-largest MTPT sequence in *C. lanceolata* (2995 bp) of the order Apiales (Table S14). Secondly, an MTPT of 888 bp in length is shared among the related species (Figure S5). However, the similarity (74%) is relatively low in the BLAST results. We speculate that it might represent a fragment of the cpgenome that migrated early into the mitogenome (Table S15). Lastly, this 888 bp MTPT is mostly found in the 11 Apiales species as a single-copy sequence. However, it has two copies in *P. grandiflorus* and three in *S. divaricata*.

dN/dS analysis is commonly used to identify potential selection on genes. In general, most genes in mitogenome are conserved and in neutral evolution and under purifying selection. However, two proteins, namely, *ccmB*, and *rps4*, had dN/dS ratios of > 1. Cytochrome c biogenesis protein B (*ccmB*) is a member of the *ccm* gene family crucial for cytochrome c biosynthesis [22]. The plant mitogenome acquired this biosynthesis process from early prokaryote cells [23, 24]. Ribosomal protein S4 (*rps4*) is one of the proteins from the small ribosomal subunit S4 that directly binds to 165 ribosomal RNA [25]. In a previous study, the *ccmB* gene was found to have undergone positive selection in Lamiales plants [26]. The biological relevance of this observation remains to be illustrated.

The ATP synthase subunit 9 (*atp9*) gene can be found in mitochondrial and nuclear DNA. Its migration is often a potential driving force for mitogenome evolution and is frequently used in CMS breeding [27–29]. The *atp9* gene is strongly negatively selected in related plants, similar to those previously reported [30]. The purifying selection of the *atp9* gene indicates that it could be used in CMS breeding of related plants.

Both mitochondria and chloroplasts had been once independent prokaryotes. Over time, cpgenomes became progressively smaller, whereas mitogenomes gradually expanded because of frequent exchanges with nuclear and chloroplast DNA [31]. In plants, the mitogenome is considerably larger than the cpgenome [30, 32, 33]. In the present study, the mitogenome (293,897 bp) is nearly twice the size of the cpgenome (147,832 bp), consistent with previous research findings. A large part of the mitogenome is similar to the cpgenome [34, 35]. Previous research has shown that MTPT regions are mutational hotspots [36]. Herein, we found 10 groups of sequences in the mitogenome of *S. divaricata*, representing 6.1% of the mitogenome, similar to cpgenome sequences. Four of these are similar to sequences in the IR regions. Thus, the sequences from the IR regions of the cpgenome can be reasonably speculated to have contributed to the expansion of the mitogenome [37].

Several efficient and accurate bioinformatics analysis software tools were used in this study to enhance the quality of the analysis results. Automatic annotation usually results in errors, such as missing 5′ and 3′ end sequences. Apollos is widely used to correct errors in automatically predicted results [38]. The standard bootstrap method is extensively used to evaluate the robustness of the phylogenetic analysis results. However, it can consume very large amounts of computing resources. UFBoot2 has improved its resampling strategies for phylogenomic data and performed better than UFBoot [39]. The REPuter software is widely employed for organelle genome repeat analysis [40]. Compared with the vmatch software, it can identify two more types of repeat sequences, namely, complement and reverse repeats. These software programs were used in this study for annotation error correction, phylogenetic analysis and repetitive sequence analysis.
Fig. 7 The dotplot graphs reveal collinear regions between mitogenomes in related species compared to S. divaricata. The red line segment represents forward direction, and the blue line segment represents reverse direction.
Plant mitogenomes are difficult to assemble for two reasons. Firstly, there are no efficient methods for enriching plant mitochondria before DNA extraction. Secondly, unlike animal mitogenomes, plant mitogenomes are highly diverse, particularly those of angiosperms. For example, the size of plant mitogenomes ranges from 66 kb to 2 MB, making the use of a reference-based method for genome assembly challenging [41–43]. The structure of plant mitogenomes can be complex. Mitogenomes can have multiple chromosomes [44]. The presence of long repeat sequences can further complicate the assembly process. In this study, we successfully assembled the mitogenome of S. divaricata, thanks to the scarcity of repeat elements in the mitogenome. Nevertheless, long reads produced by third-generation sequencing technologies are needed to validate the correctness of the mitogenome [45].

Conclusions
In this study, we reported the mitogenome of S. divaricata for the first time and assembled its cpgenome from the same sequencing data set. Phylogenomic analysis with the mitogenome and the cpgenome assembled showed congruent trees. We identified 10 mitochondrial DNA fragments homologous to those in the cpgenome by comparing the mitogenome and cpgenome sequences. DNA fragments from the cpgenome IR region might have transferred into the mitogenome and contributed to its length expansion. This study provides valuable information to understand the coordinated evolution of the cpgenomes and the mitogenomes of plants belonging to the family Apiaceae.

Methods
Plant materials, DNA extraction and sequencing
Fresh young leaves of S. divaricata were collected from the Institute of Medicinal Plant Development (IMPLAD), Beijing, China. Total DNA was extracted using a DNA extraction kit (Tiangen Biotech, Beijing, China) and stored at the herbarium of IMPLAD with the accession number Implad 20,170,491. DNA library was constructed from 1 μg genomic DNA, and the library was sequenced using the next-generation sequencing platform (Illumina, San Diego, CA, USA).

Genome assembly and annotation
The organelle genomes were assembled with GetOrganelle (v.1.6.4) [46]. In particular, the cpgenome was assembled with the parameters ‘-R -15 -k 21,45,65,85,105 -F embplant_pt’. By comparison, the mitogenome was assembled with the parameters ‘-R 50 -k 21,45,65,85,105 -P 1000000 -F embplant_mt’. The bandage software (v.0.8.1) tool was used to visualise the connections among contigs [47]. The cpgenome and the mitogenome were annotated using GeSeq and CPGAVAS2, respectively [48, 49]. The annotation results were manually improved by using Apollo (v.1.11.8) [38]. Lastly, the structures of the cpgenome and the mitogenome were plotted using CPView-RSG (http://www.therbalgenomics.org/cpgview/) and OGdraw [50], respectively. The cpgenome and the mitogenome had been submitted to GenBank with the accession numbers MZ089852 and MZ128146, respectively.

DNA transfer between the chloroplast and the mitochondrion
Sequence similarity between the cpgenome (MZ089852) and the mitogenome (MZ128146) were analysed to identify transferred DNA fragments by using BLASTN with an e-value cut-off of 1e-5 [51]. The results were visualised using the Circos package implemented in TBtools [52, 53].

Analysis of repeat elements
Microsatellite sequence repeats were identified using MISA with the parameters ‘1-10 2-5 3-4 4-3 5-3 6-3’ [54]. Tandem repeats were identified using TRF with the parameters ‘2 7 8 10 50 500 -f -d -m’ [55]. Dispersed repeats were identified using REPter web server (https://bibiserv.cbi.unc.edu/repeter/, 2001) with the parameters ‘Hamming Distance 3, Maximum Computed Repeats 5000, Minimal Repeat Size 30’ and filtered with an e-value cut-off of 1e-5 [40].

Sequence alignment and phylogenetic inference
Differences in the sequences of the published cpgenome of S. divaricata and the cpgenomes assembled herein were compared using BLASTN with an e-value cut-off of 1e-5 [51]. For phylogenetic analysis, the sequences of shared genes were extracted and concatenated using Phylosuite [56]. They were then aligned using MAFFT [57]. Gblocks was utilised to select the optimal multiple sequence alignment regions with default parameters [58]. Both the cpgenome and the mitogenome of S. divaricata and 10 related species were subjected to phylogenetic analysis by using IQTREE [59]. Two Solanum species were selected as the outgroups. Phylogenetic analysis was conducted with the best evolutionary model ‘TVM+F+I+G4’ and ‘GR+F+G4’ based on Bayesian Information Criterion scores for the cpgenomes and the mitogenomes, respectively. Bootstrap analysis was performed with 1000 replicates by using UFBoot2 (v 1.6.12) [39]. The newick format tree was visualised using iTOL6 (https://itol.embl.de/) [60].
Selective pressure analysis
The dN/dS ratios of 14 protein-coding sequences among mitogenomes from *S. divaricata* and 10 campanulids were calculated using PAML (version 4.9) [61]. The yn00 module was selected to estimate nonsynonymous substitution rate (dN) and synonymous substitution rate (dS) with the following parameters: ‘verbose = 0, icode = 0, weighting = 0, commonf3x4 = 0, ndata = 1’. A boxplot of pairwise dN/dS values was created using the R package ggplot2 [62].

Prediction of RNA-editing sites and polycistronic transcript units
The transcriptome data (SRR11365146) of *S. divaricata* were downloaded from the SRA database (http://www.ncbi.nlm.nih.gov/sra). The raw data were mapped to the *S. divaricata* organelle genomes by using TopHat2 [63]. RNA-editing sites were calculated using REDITools with the parameters ‘coverage ≥ 5, frequency ≥ 0.1, p-value ≤ 0.5’ [64]. The raw data were de novo assembled by using the Trinity program [65]. The 50 longest transcripts were selected for comparison with the genes from the organelle genome to predict polycistronic transcript units.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12864-022-08821-0.

Additional file 1: Table S1. Gene compositions of the *S. divaricata* plastome. Table S2. The lengths of introns and exons for the splitting genes in the *S. divaricata* plastome. Table S3. The lengths of introns and exons for the splitting genes in the *S. divaricata* mitogenome. Table S4. Microsatellite repeats in the *S. divaricata* plastome. The structure of microsatellite repeats is presented as repeat units surrounded with parenthesis and the numbers of repeat units. Table S5. Microsatellite repeats in the *S. divaricata* mitogenome. The structure of microsatellite repeats is presented as repeat units surrounded with parenthesis and the numbers of repeat units. Table S6. Tandem repeats in the *S. divaricata* plastome. Table S7. Tandem repeats in the *S. divaricata* mitogenome. Table S8. Dispersed repeats in the *S. divaricata* plastome. Table S9. Dispersed repeats in the *S. divaricata* mitogenome. Table S10. DNA transfer of *S. divaricata* organellar genomes. Table S11. The dN, dS of the common genes of *S. divaricata* mitogenome. Table S12. RNA editing sites of *S. divaricata* organellar genomes. Table S13. PTUs identified in organellar genomes of *S. divaricata*. Table S14. The MtPT fragments in the Apiales species. Table S15. The common MtPT DNA fragments in Apiales species. Supplementary Figure 1. Gis-splicing gene map generated for the chloroplast genome of *S. divaricata*. Supplementary Figure 2. Trans-splicing gene map generated for the chloroplast genome of *A. thaliana*. Supplementary Figure 3. The dotplot of two *S. divaricata* chloroplast genomes. Supplementary Figure 4. The difference between the two chloroplast genomes identified by BLASTN. Supplementary Figure 5. Comparison of the cpgenome and mitogenome sequences suggest the transferring of DNA fragments from the cpgenome to the mitogenome.

Acknowledgments
We would like to thank JianJun Jin, the author of GetOrganelle, for guiding the assembly process. Thanks to Ziqi Zhou for help with preparing the figures.

Authors’ contributions
CL conceived the study; YN and JLL assembled and annotated the mitogenome; JNY collated the data; YN carried out the comparative analysis; YN and JLL wrote the manuscript; CL, HMC and PHC reviewed the manuscript critically. All authors read and approved the manuscript.

Funding
The study was supported by CAMS Innovation Fund for Medical Sciences (CIFMS) (2021-I2M-1-022), the National Science & Technology Fundamental Resources Investigation Program of China (2018FY1100705), National Natural Science Foundation of China (81872996), Open Fund of the National Sugarcane Engineering and Technology Research Center (KJG16005R), Science and Technology Innovation Special Fund of Fujian Agriculture and Forestry University (KFA17263A), [KF2015080], [KF2015118]. The funders were not involved in the study design, data collection, analysis, publication decision, or manuscript preparation.

Availability of data and materials
The cpgenome and mitogenome sequences supporting the conclusions of this article are available in GenBank (https://www.ncbi.nlm.nih.gov/) with accession numbers: MZ089852 and MZ128146, respectively. The sample has been deposited in the Institute of Medicinal Plant Development (Beijing, China) with the accession numbers 20170491. The raw data has been submitted to the SRA database (BioSample: SAMN20926830; BioProject: PRJNA756825; SRA: SRR11563639).

Declarations
Ethics approval and consent to participate
We collected fresh leaf materials of *Saposnikovia divaricata* for this study. The plant sample was identified by Professor Zhao Zhang in the Institute of Medicinal Plant. We processed the voucher specimens and deposited them in the Institute of Medicinal Plant Development (Beijing, China) with the accession numbers 20170491. The study, including plant samples, complies with relevant institutional, national, and international guidelines and legislation. No specific permits are required for plant collection.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing and conflicting interests.

Received: 9 November 2021 Accepted: 4 August 2022
Published online: 10 August 2022

References
1. Kreiner J, Pang E, Lenon GB, Yang AWH. Saposnikoviae divaricata: a phytochemical, pharmacological, and pharmacokinetic review. Chin J Nat Med. 2017;15(4):255–64.
2. Group TAP, Chase MW, Christenhusz MJM, Fay MF, Byng JW, Judd WS, et al. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc. 2016;181(1):1–20.
3. Uragatavo BM, Shults EE, Taraskin VV, Radnaeva LD, Petrova TN, Rybalova TV, et al. Chromones and coumarins from Saposnikovia divaricata (Turcz.) Schischk. Growing in Buryatia and Mongolia and their cytotoxicity. J Ethnopharmacol. 2002;61:12517.
4. You W, Wang H. High yield cultivation of *S. divaricata*. Spec Econ Plant Anim (Chinese Journal). 2004;7(4):23.
5. Yang M, Wang C-C, Xu J-P, Wang J, Zhang C-H, Li M-H. Saposnikovia divaricata—an Ethnopharmacological, phytochemical and pharmacological review. Chin J Integr Med. 2020;26(1):1–8.
6. Markelch A, Brisson N. Recombination and the maintenance of plant organelle genome stability. New PhytoL 2010;186(2):299–317.
7. Wolfsberg TG, Schafer S, Tatusov RL, Tatusova TA. Organelle genome resources at NCBI. Trends Biochem Sci. 2001;26(3):199–203.
8. Johnston IG. Tension and resolution: dynamic, evolving populations of organelle genomes within plant cells. Mol Plant. 2019;12(6):764–83.

9. Morley SA, Ahmad N, Nielsen BL. Plant organelle genome replication. Plants. 2019;8(10):358.

10. Chen Z, Zhao N, Li S, Grover CE, Nie H, Wendel JF, et al. Plant mitochondrial genome evolution and cytoplasmonic male sterility. Crt Rev Plant Sci. 2017;36(1):55–69.

11. Bao Z, Zhu Z, Zhang H, Zhong Y, Wang W, Zhang J, et al. The complete chloroplast genome of Saponoskivia divaricata. Mitochondrial DNA Part B. 2020;5(1):360–1.

12. Kim C-K, Jin M-W, Kim Y-K. The complete mitochondrial genome sequences of Bupleurum falcatum (Apiaceae: Apiales). Mitochondrial DNA Part B. 2020;5:2576–7.

13. Iorizzo M, Pottorff M, Bostan H, Ellison S, Cavagnaro P, Senalik D, et al. Recent advances in carotid genomics. In: II International Symposium on Carrot and Other Apiaceae 1264, vol. 2018, 2018. p. 75–90.

14. Powell W, Machray GC, Provan J. Polymorphism revealed by simple sequence repeats. Trends Plant Sci. 1996;1(7):215–22.

15. Fan H, Chu J-Y. A brief review of short tandem repeat mutation. Genomics Proteomics Bioinform. 2007;5(1):7–14.

16. Verstrepen KJ, Jansen A, Lewitter F, Fink GR. Intragenic tandem repeats generate functional variability. Nat Genet. 2005;37(9):966–90.

17. Smyth DR. Dispersed repeats in plant genomes. Chromosoma. 1991;100(6):355–69.

18. Waksuigi T, Hirose T, Horihata M, Tsudzuki T, Kössel H, Sugiura M. Creation of a novel protein-coding region at the RNA level in black pine chloroplasts: the pattern of RNA editing in the gymnosperm chloroplast is different from that in angiosperms. PNAS. 1996;93(16):8766–70.

19. Ellis J. Promiscuous DNA—chloroplast genes inside plant mitochondria. Proc Natl Acad Sci. 1991;88(20):9199–90.

20. Knoop V. The mitochondrial DNA of land plants: peculiarities in phylogenetic analysis. Bioinformatics. 1996;12(1):67–9.

21. Fan H, Chu J-Y. A brief review of short tandem repeat mutation. Genomics Proteomics Bioinform. 2007;5(1):7–14.

22. Verstrepen KJ, Jansen A, Lewitter F, Fink GR. Intragenic tandem repeats generate functional variability. Nat Genet. 2005;37(9):966–90.

23. Smyth DR. Dispersed repeats in plant genomes. Chromosoma. 1991;100(6):355–69.

24. Waksuigi T, Hirose T, Horihata M, Tsudzuki T, Kössel H, Sugiura M. Creation of a novel protein-coding region at the RNA level in black pine chloroplasts: the pattern of RNA editing in the gymnosperm chloroplast is different from that in angiosperms. PNAS. 1996;93(16):8766–70.

25. Ellis J. Promiscuous DNA—chloroplast genes inside plant mitochondria. Proc Natl Acad Sci. 1991;88(20):9199–90.

26. Verstrepen KJ, Jansen A, Lewitter F, Fink GR. Intragenic tandem repeats generate functional variability. Nat Genet. 2005;37(9):966–90.

27. Smyth DR. Dispersed repeats in plant genomes. Chromosoma. 1991;100(6):355–69.

28. Waksuigi T, Hirose T, Horihata M, Tsudzuki T, Kössel H, Sugiura M. Creation of a novel protein-coding region at the RNA level in black pine chloroplasts: the pattern of RNA editing in the gymnosperm chloroplast is different from that in angiosperms. PNAS. 1996;93(16):8766–70.

29. Ellis J. Promiscuous DNA—chloroplast genes inside plant mitochondria. Proc Natl Acad Sci. 1991;88(20):9199–90.

30. Verstrepen KJ, Jansen A, Lewitter F, Fink GR. Intragenic tandem repeats generate functional variability. Nat Genet. 2005;37(9):966–90.

31. Smyth DR. Dispersed repeats in plant genomes. Chromosoma. 1991;100(6):355–69.

32. Verstrepen KJ, Jansen A, Lewitter F, Fink GR. Intragenic tandem repeats generate functional variability. Nat Genet. 2005;37(9):966–90.

33. Smyth DR. Dispersed repeats in plant genomes. Chromosoma. 1991;100(6):355–69.

34. Waksuigi T, Hirose T, Horihata M, Tsudzuki T, Kössel H, Sugiura M. Creation of a novel protein-coding region at the RNA level in black pine chloroplasts: the pattern of RNA editing in the gymnosperm chloroplast is different from that in angiosperms. PNAS. 1996;93(16):8766–70.

35. Ellis J. Promiscuous DNA—chloroplast genes inside plant mitochondria. Proc Natl Acad Sci. 1991;88(20):9199–90.

36. Verstrepen KJ, Jansen A, Lewitter F, Fink GR. Intragenic tandem repeats generate functional variability. Nat Genet. 2005;37(9):966–90.

37. Smyth DR. Dispersed repeats in plant genomes. Chromosoma. 1991;100(6):355–69.

38. Waksuigi T, Hirose T, Horihata M, Tsudzuki T, Kössel H, Sugiura M. Creation of a novel protein-coding region at the RNA level in black pine chloroplasts: the pattern of RNA editing in the gymnosperm chloroplast is different from that in angiosperms. PNAS. 1996;93(16):8766–70.

39. Ellis J. Promiscuous DNA—chloroplast genes inside plant mitochondria. Proc Natl Acad Sci. 1991;88(20):9199–90.

40. Verstrepen KJ, Jansen A, Lewitter F, Fink GR. Intragenic tandem repeats generate functional variability. Nat Genet. 2005;37(9):966–90.

41. Smyth DR. Dispersed repeats in plant genomes. Chromosoma. 1991;100(6):355–69.

42. Waksuigi T, Hirose T, Horihata M, Tsudzuki T, Kössel H, Sugiura M. Creation of a novel protein-coding region at the RNA level in black pine chloroplasts: the pattern of RNA editing in the gymnosperm chloroplast is different from that in angiosperms. PNAS. 1996;93(16):8766–70.

43. Ellis J. Promiscuous DNA—chloroplast genes inside plant mitochondria. Proc Natl Acad Sci. 1991;88(20):9199–90.

44. Verstrepen KJ, Jansen A, Lewitter F, Fink GR. Intragenic tandem repeats generate functional variability. Nat Genet. 2005;37(9):966–90.

45. Smyth DR. Dispersed repeats in plant genomes. Chromosoma. 1991;100(6):355–69.

46. Verstrepen KJ, Jansen A, Lewitter F, Fink GR. Intragenic tandem repeats generate functional variability. Nat Genet. 2005;37(9):966–90.

47. Smyth DR. Dispersed repeats in plant genomes. Chromosoma. 1991;100(6):355–69.

48. Waksuigi T, Hirose T, Horihata M, Tsudzuki T, Kössel H, Sugiura M. Creation of a novel protein-coding region at the RNA level in black pine chloroplasts: the pattern of RNA editing in the gymnosperm chloroplast is different from that in angiosperms. PNAS. 1996;93(16):8766–70.

49. Ellis J. Promiscuous DNA—chloroplast genes inside plant mitochondria. Proc Natl Acad Sci. 1991;88(20):9199–90.

50. Verstrepen KJ, Jansen A, Lewitter F, Fink GR. Intragenic tandem repeats generate functional variability. Nat Genet. 2005;37(9):966–90.

51. Smyth DR. Dispersed repeats in plant genomes. Chromosoma. 1991;100(6):355–69.

52. Waksuigi T, Hirose T, Horihata M, Tsudzuki T, Kössel H, Sugiura M. Creation of a novel protein-coding region at the RNA level in black pine chloroplasts: the pattern of RNA editing in the gymnosperm chloroplast is different from that in angiosperms. PNAS. 1996;93(16):8766–70.

53. Ellis J. Promiscuous DNA—chloroplast genes inside plant mitochondria. Proc Natl Acad Sci. 1991;88(20):9199–90.
55. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80.
56. Zhang D, Gao F, Jakovlje 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(1):348–55.
57. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
58. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000;17(4):540–52.
59. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–74.
60. Letunic I, Bork P. Interactive tree of life (ITOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;47(W1):W256–9.
61. Yang Z. PAAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.
62. Wickham H, et al. Comput Stat. 2011;3(2):180–5.
63. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14(1):1–13.
64. Picardi E, Pesole G. REDItools: high-throughput RNA editing detection made easy. Bioinformatics. 2013;29(14):i813–4.
65. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644–52.

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