

Research

Assembly and characterization of the complete mitochondrial genome of *Descurainia sophia*: a medical plant producing glucosinolate

Le Zhao^{1,2} · Hang Yan¹ · Yun-hao Zhu^{1,2} · Hao-yu Jia¹ · Yong-guang Han¹ · Xiao-ke Zheng^{1,2} · Min Wang^{3,4} · Wei-sheng Feng^{1,2}

Received: 8 January 2024 / Accepted: 4 March 2024

Published online: 13 March 2024

© The Author(s) 2024 [OPEN](#)

Abstract

As a commonly used traditional Chinese medicine, *Descurainia sophia* has various reported pharmacological activities such as prevention of chronic diseases, cardiovascular protection, neuroprotection, anticancer, antioxidant, and has high nutritional values, but its mitochondrial genome (mitogenome) has not been reported. In this study, the *D. sophia* mitogenome was sequenced, assembled and annotated using next generation sequencing technologies on the Illumina HiSeq and PacBio Sequel platform. The results showed that the *D. sophia* mitogenome was a circular DNA molecule with the length of 265,457 bp and the GC content of 44.78%, encoding unique 56 genes, including 35 protein coding genes, 18 tRNA genes and 3 rRNA genes. We analyzed codon preference, repeated sequences, DNA sequence transformation, phylogenetic relationship, RNA editing and syntenicity in the *D. sophia* mitogenome. Based on the DNA sequences of mitogenomes, the phylogenetic and syntenicity analysis indicated that *D. sophia* was more closely related to *Boechera stricta*, and the mitogenome of *D. sophia* underwent frequent genome recombination. In this study, we obtained the complete mitogenome of *D. sophia*, which can provide valuable data support for future studies on genetic characteristics, phylogenetic relationships, species identification and development of new molecular markers in *D. sophia*.

Article Highlights

- The results of this study provide valuable information about the mitogenome of *D. sophia*, and will contribute to the research of evolution and mitogenome variation in the genus *Descurainia*.
- The *D. sophia* mitogenome was an unbranched circular DNA molecule, and the syntenicity analysis revealed that genomic recombination occurred frequently in the mitogenome of *D. sophia*, which was evolutionarily closest related to *Boechera stricta*.
- A total of 57 SSRs and 406 potential C → U RNA editing sites were identified in mitogenome, which can be used to develop new molecular markers.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s42452-024-05775-8>.

✉ Min Wang, wangmin_216@163.com; ✉ Wei-sheng Feng, fwsh@hactcm.edu.cn | ¹School of Pharmacy, Henan University of Chinese Medicine, No. 156 Jinshui East Road, Zhengzhou 450046, Henan, China. ²The Engineering and Technology Research Center for Chinese Medicine Development of Henan Province, Zhengzhou 450046, China. ³College of Chemistry and Materials Engineering, Beijing Technology & Business University, No.11, Fucheng Road, Haidian District, Beijing 100048, China. ⁴Beijing Key Laboratory of Plant Research and Development, Beijing 100048, China.



Keywords *Descurainia sophia* · Mitochondrial genome · Repeated sequences · Phylogenetic analysis · RNA editing

1 Introduction

Descurainia sophia, also known as flaxweed, belongs to the genus *Descurainia* of the Cruciferae family and is mainly distributed in the northern part of China. The dried mature seeds of the *D. sophia* are called “Tinglizi”, which is the mainstream species of “Tinglizi” commonly used in clinical medicine nowadays and is recorded in the Pharmacopoeia of the People’s Republic of China (2020 edition) [1]. Traditional Chinese medical theory suggests its seeds have the functions of cough relief, asthma prevention, urination promotion and edema alleviation [2]. In addition, the seeds of *D. sophia* was found to have significant pharmacological activities in improving cardiovascular function and regulating blood lipid levels [3]. Through analyzing the chemical composition, various secondary metabolites were isolated from *D. sophia*, such as glucosinolates, isothiocyanates, flavonoids, cardiac glycosides, phenylpropanoids, organic acids, and fatty oils, etc. [3]. A total of 10 glucosinolates [4–6] and 10 isothiocyanates [7, 8] have been identified from *D. sophia*, and in 2016 we isolated 4 benzenic glucosinolates compounds from the seeds of *D. sophia* [9]. Glucosinolates and isothiocyanates were not only involved in the defense system of the plant, but also had preventive effects against diabetes and hypertension, neuroprotection, and anticancer, and were the basis of important medicinal substances for *D. sophia* [10]. Moreover, the oil content of *D. sophia* seeds was as high as 39–44%, and the oil was mostly unsaturated fatty acids, including 38.13–40.92% linolenic acid [11]. As an essential fatty acid, linolenic acid had a variety of biological activities, such as lowering blood lipids, cardiovascular protection, anticancer, antioxidant, etc. [12], and had a high medicinal and nutritional value. At present, most studies on *D. sophia* have focused on the isolation of chemical components and pharmacological activities, but there are no reports about the nuclear and mitochondrial genomes of *D. sophia*.

Mitochondria are one of the important organelles in eukaryotes and are the site of aerobic oxidation in the organism, providing energy for various physiological activities of the cell. Mitochondria possess their own genetic material and genetic system, and are semi-autonomous organelles [13]. The first plant mitochondrial genome (mitogenome) sequenced was that of the terrestrial plant- *Marchantia polymorpha*, reported in 1992 [14]. With the development of sequencing technology, the mitogenome data of the dicotyledons *Arabidopsis thaliana*, *Beta vulgaris*, *Nicotiana tabacum*, *Brassica napus*, the monocotyledons *Zea mays*, *Oryza sativa* [15], and the medicinal plants *Salvia miltiorrhiza* [16], *Cannabis sativa* [17] have been released. The size of the mitogenome varies widely among different species of plants and animals, with most animals having a mitogenome size of about 15–17 kb, while angiosperms generally have a mitogenome size between 200 and 750 kb, with that of cucumber reaching 1156 kb [15]. Research on the mitogenome has progressed much less than that on the chloroplast genome because of its complex structure, few coding genes, and difficulty in assembly. As of 2023-Nov-15, 10,387 chloroplast genomes, 596 mitochondrial genomes and 1296 plastid genomes data have been recorded by the NCBI GenBank database. The main characteristics of plant mitogenomes include huge variation in genome size and structure, highly conserved genes, sparse gene distribution, massive amount of non-coding sequences, and a large number of RNA editing sites.

In recent decades, with the rapid development of second- and third-generation high-throughput sequencing technologies and the decreasing cost of sequencing, more and more plant mitogenomes have been reported, and plant mitogenomes are used for molecular identification and phylogenetic analysis [18]. Therefore, the mitogenome of *D. sophia* was of great value for studying the origin and evolution of *D. sophia*, determining the phylogenetic status and the conserving germplasm resources. In this study, the complete mitogenome of *D. sophia* was assembled using high-throughput sequencing technology, and the structure and composition of its mitogenome were investigated, which provided the molecular basis for studying the genetic diversity, evolutionary relationship and the development of new molecular markers in *D. sophia*.

2 Materials and methods

2.1 Plant materials and mitogenome sequencing

The leaves of *D. sophia* was collected from the Henan Provincial Medicinal Botanical Garden, Henan University of Chinese Medicine, Zhengzhou City, Henan Province, China (34°46′41″N, 113°48′28″E) in April 2022. High-quality genomic DNA was isolated from leaves of *D. sophia* using the modified CTAB method [19], followed by high-throughput sequencing on the PacBio Sequel II and Illumina HiSeq X Ten platforms.

2.2 Assembly and annotation of mitogenome

The Illumina short-reads were assembled using GetOrganelle v1.7.7.0 [20] with the default parameters for assembling plant mitochondrial genomes, and selected the plant mitochondrial genome database “embplant_mt”. The PacBio long-reads were assembled directly using the default parameters of Flye assembler v2.9.2 [21] to obtain graphical assembly results in GFA format. Then the PacBio long-reads were aligned to the mitogenome by BWA v0.7.17 [22], and the aligned PacBio long-reads were exported through the SAMtools [23], which was used to resolve the repeated sequence regions of the graphical mitogenome assembled from the Illumina short-reads. The results of the short-reads and long-reads were then aligned to finally obtain the structure and composition of *D. sophia* mitogenome.

Arabidopsis thaliana (NC_037304), *Capsella rubella* (NC_042883.1) and *Liriodendron tulipifera* (NC_021152.1) were selected as reference genomes and Geseq v2.03 [24] was adopted to annotate the protein coding genes (PCGs) of the *D. sophia* mitogenome. The tRNAscan-SE v2.0.11 [25] and BLASTN v2.13.0 [26] (parameter: -evalue 1e-5 -outfmt 6 -max_hsps 10 -word_size 7 -task blastn-short) were applied to annotate tRNA and rRNA genes, respectively. The Apollo v1.11.8 [27] was used to manually correct mitogenome annotation errors.

2.3 Analysis of codon preference and repeated sequences

The Phylosuite v1.1.16 [28] was adopted to extract the PCGs of the mitogenome, and then the PCGs were analyzed for codon preference and RSCU (Relative Synonymous Codon Usage) values were calculated using MEGA v7.0 [29]. Repeated sequences including simple sequence repeats (SSRs), tandem repeats and dispersed repeats were identified by MISA v2.1 [30], TRF v4.09 [31] and REPuter program [32] (default parameters), respectively. The results were visualized by Excel 2021.

2.4 Identification of mitochondrial plastid DNAs (MTPTs)

The GetOrganelle v1.7.7.0 [20] and CPGAVAS2 [33] (default parameters) were adopted to assemble and annotate the chloroplast genome of *D. sophia* from Illumina short-reads, respectively. The annotation results of the chloroplast genome were corrected using CPGView software [34]. The BLASTN v2.13.0 [26] was applied to identify the mitochondrial plastid DNAs (MTPTs) between mitogenome and chloroplast genome. The results were visualized by using Circos v0.69.9.

2.5 Phylogenetic and synteny analysis

The 29 plant species, including *D. sophia*, from three orders of angiosperms were selected according to their genetic relationships to construct phylogenetic tree, and their complete mitogenomes were downloaded from the NCBI GenBank database. These species were *D. sophia* (OQ916154), *Brassica napus* (NC_008285.1), *Brassica rapa* (NC_016125.1), *Brassica juncea* (NC_016123.1), *Brassica carinata* (NC_016120.1), *Brassica oleracea* (NC_016118.1), *Brassica nigra* (NC_029182.1), *Boechera stricta* (NC_042143.1), *Arabidopsis thaliana* (NC_037304.1), *Arabis alpina* (NC_037070.1), *Capsella bursa-pastoris* (MN746809.2), *C. rubella* (NC_042883.1), *Sinapis arvensis* (NC_031896.1), *Schrenkiella parvula* (KT988071.2), *Eruca vesicaria* (KF442616.1), *Raphanus sativus* (NC_018551.1), *Carica papaya* (NC_012116.1), and *Batis maritima* (NC_024429.1) of the order Brassicales; *Theobroma grandiflorum* (NC_066895.1), *Theobroma cacao* (NC_066894.1), *Aquilaria sinensis* (NC_054354.1), *Bombax ceiba* (NC_038052.1), *Gossypium thurberi* (NC_035074.1), *Hibiscus cannabinus* (NC_035549.1), *Gossypium arboreum* (NC_035073.1), *Gossypium davidsonii* (NC_035075.1), and *Gossypium trilobum* (NC_035076.1) of the order Malvales; *Cotinus coggygria* (NC_064986.1) and *Mangifera longipes* (NC_060990.1) of the order Sapindales. The mitogenomes of *C. coggygria* and *M. longipes* were set as outgroup. The 21 common genes (*atp6*, *atp8*, *ccmB*, *ccmC*, *ccmFC*, *cob*, *cox1*, *cox3*, *matR*, *nad1*, *nad2*, *nad4*, *nad5*, *nad6*, *nad7*, *nad9*, *rpl5*, *rpl16*, *rps3*, *rps14*, *sdh4*) that were 100% conserved in all plant mitogenomes were extracted with PhyloSuite v1.1.16 [28]. The multiple sequence alignment analysis was performed using MAFFT v7.505 [35], and then IQ-TREE v1.6.12 [36] was used to construct phylogenetic tree. The iTOL v4.0 [37] software was adopted to visualize the results of phylogenetic tree.

According to the results of phylogenetic analysis, the mitogenomes of seven plant species (belonging to Brassicales) were chosen for synteny analysis, with the following species: *C. bursa-pastoris*, *C. rubella*, *A. thaliana*, *B. stricta*, *D. sophia*, *B. maritima*, and *C. papaya*. Among them, *C. bursa-pastoris*, *C. rubella*, *A. thaliana*, *B. stricta* were more closely related to *D. sophia*, while *B. maritima*, and *C. papaya* were more distantly related to *D. sophia*. To analyze the collinearity relationship of these seven mitogenomes, BLASTN results for two-by-two alignment of mitogenomes with each species were

acquired based on the BLAST program. Homologous sequences longer than 500 bp were reserved as conserved collinearity regions, and multiple synteny plot were generated by MCscanX software [38].

2.6 Prediction of RNA editing sites

The DNA sequences of 35 PCGs encoded by the *D. sophia* mitogenome were used as input files, and then the C → U RNA editing sites of mitochondrial PCGs were predicted by Deepred-Mt software [39] with the parameter of “cutoff value=0.9”. Deepred-Mt was based on the Convolutional Neural Networks (CNN) model for prediction, which had a high accuracy compared to other prediction software.

3 Results

3.1 Characteristics of *D. sophia* Mitogenome

We assembled mitogenome of *D. sophia* based on PacBio Sequel II and Illumina HiSeq X Ten platforms. A total of 14.8 Gb raw data from PacBio long-reads and 12.9 Gb raw data from Illumina short-reads were generated, achieving the 60.76× average depth of *D. sophia* mitogenome coverage with no gap (Fig. S1). The draft mitogenome assembled from Illumina short-reads was visualized using Bandage software [40], resulting in a unitig graph, which represented the mitogenome structure of *D. sophia*. The yellow blocks in unitig graph indicated the repeated sequence regions of branch nodes (Fig. 1A). Afterwards, the branch nodes caused by repeated sequences were resolved based on PacBio long-reads, and then a circular contig molecule was generated from the unitig graph (Fig. 1C). This was consistent with the results obtained by assembling PacBio long-reads directly using Flye software (Fig. 1B), which differ in length by only two bases. The above results showed that the assembly results based on Illumina short-reads and PacBio long-reads were consistent. Considering the high accuracy of Illumina sequencing platform, we preferred the assembly results based on Illumina short-reads. The main structure of the *D. sophia* mitogenome was a single circular molecule. After excluding repeated regions with PacBio long-reads, one closed-loop DNA molecule without branches was assembled with a total length of 265,457 bp and GC content of 44.78% (Fig. 2). The assembled mitogenome sequences and raw data (Illumina and PacBio) of *D. sophia* had been submitted to the NCBI GenBank database (accession number OQ916154) and SRA database (accession number PRJNA1078884).

The *D. sophia* mitogenome was annotated with 56 unique genes, including 35 PCGs, 3 rRNA genes (*rrn5*, *rrn18*, and *rrn26*) and 18 tRNA genes (of which *trnM-CAU*, *trnS-GCU* and *trnY-GUA* were multi-copied genes) (Fig. 2, Table 1). PCGs contained 25 core genes and 10 variable genes (Table 1). The 25 core genes were divided into 7 classes, including 9 NADH dehydrogenase genes, 1 ubiquinol cytochrome c reductase gene, 3 cytochrome c oxidase genes, 5 ATP synthase genes, 5 cytochrome c biogenesis genes, 1 maturase gene, and 1 transport membrane protein gene. The 10 variable genes were composed of ribosomal protein large subunit (4 genes), ribosomal protein small subunit (5 genes), and succinate dehydrogenase (1 gene).

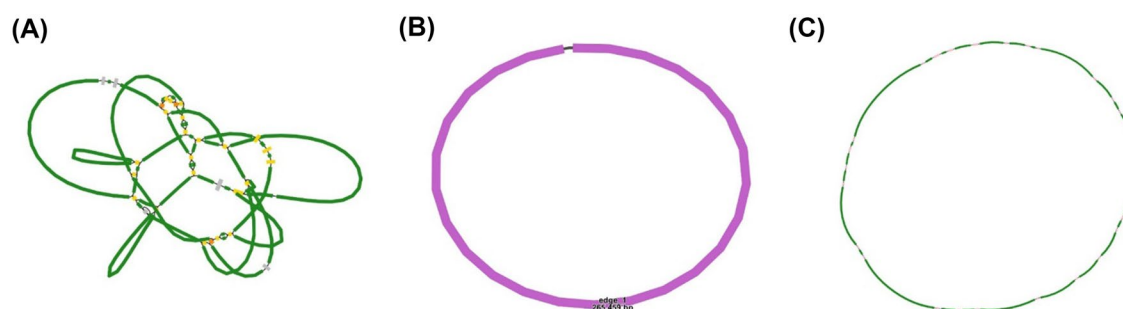


Fig. 1 Structure analysis of *D. sophia* mitogenome. **A** The unitig graph of *D. sophia* mitogenome assembled from Illumina short-reads, the yellow blocks represent the repeated sequence regions of branch nodes, and the green lines indicate a single loop contig; **B** Mitogenome assembly map based on PacBio long-reads; **C** A circular contig molecule are obtained after resolving the branch nodes caused by repeated sequences based on PacBio long-reads

Fig. 2 The complete mitochondrial genome map of *D. sophia*. The genes inside the circle are transcribed clockwise, whereas the genes outside the circles are transcribed counterclockwise. Genes are represented by different colors according to their functional classification. The arrows indicate the positive and negative strands. The gray inner circle stands for the GC content

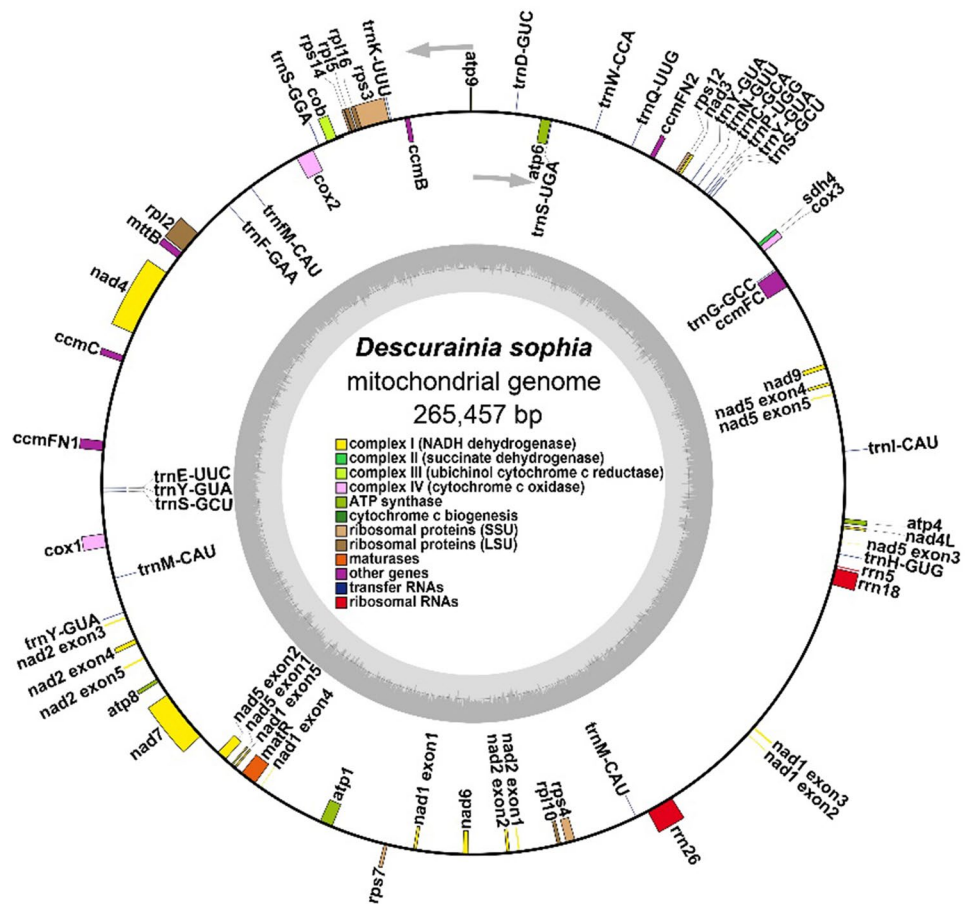


Table 1 Gene composition in the mitogenome of *D. sophia*

Group of genes		Name of genes
Core genes	Complex I (NADH dehydrogenase)	<i>nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7, nad9</i>
Variable genes	Complex III (ubiquinol cytochrome c reductase)	<i>cob</i>
	Complex IV (cytochrome c oxidase)	<i>cox1, cox2, cox3</i>
	ATP synthase	<i>atp1, atp4, atp6, atp8, atp9</i>
	Cytochrome c biogenesis	<i>ccmB, ccmC, ccmFC, ccmFN1, ccmFN2</i>
	Maturases	<i>matR</i>
	Protein transport subunit	<i>mttB</i>
	Complex II (succinate dehydrogenase)	<i>sdh4</i>
	Ribosomal protein large subunit	<i>rpl2, rpl5, rpl10, rpl16</i>
rRNA genes	Ribosomal protein small subunit	<i>rps3, rps4, rps7, rps12, rps14</i>
	Ribosome RNA	<i>rrn5, rrn18, rrn26</i>
tRNA genes	Transfer RNA	<i>trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU, trnG-GCC, trnH-GUG, trnI-CAU, trnK-UUU, trnM-CAU (× 2) *, trnN-GUU, trnP-UGG, trnQ-UUG, trnS-GCU (× 2) *, trnS-GGA, trnS-UGA, trnW-CCA, trnY-GUA (× 4) *</i>

* Multi-copied genes, e.g. (x2) means there are two copies of the gene

3.2 Analysis of codon preference in *D. sophia* Mitogenome

The 35 PCGs of *D. sophia* mitogenome were analyzed for codon preference, and the codon usage for each amino acid was shown in Fig. 3. Codons with RSCU (Relative Synonymous Codon Usage) value greater than 1 were considered to be used preferentially by amino acids (Table S1). Apart from the RSCU values of 1 for both the start codon AUG and tryptophan (UGG), there was also a general codon usage preference for mitochondrial PCGs. For instance, the stop codon had a high usage preference for UAA with the highest RSCU value of 1.68 among mitochondrial PCGs, followed by alanine (Ala) with a usage preference for GCU with an RSCU value of 1.6. Notably, the maximum RSCU value of phenylalanine (Phe) was less than 1.2, which did not have a strong codon usage preference.

3.3 Analysis of repeated sequences

A total of 57 SSRs were identified in *D. sophia* mitogenome. The monomeric, dimeric, trimeric, tetrameric and pentameric forms of SSRs were 14, 8, 10, 22, and 3, accounting for 24.56, 14.03, 17.54, 38.60 and 5.26% of the total SSRs, respectively, and no hexameric SSRs were detected (Fig. 4A). Among the 14 monomeric SSRs, 6 adenine (A) monomeric repeats were identified, taking up 42.86% of the monomeric SSRs (Table S2). In the mitogenome of *D. sophia*, there were 21 tandem repeats with a match greater than 73% and length of 8–71 bp (Table S3). Moreover, 391 pairs of dispersed repeats with length greater than or equal to 30 bp were detected, including 210 pairs of palindromic repeats and 181 pairs of forward repeats (Fig. 4B). However, no reverse repeats and complementary repeats were identified, with the longest palindromic repeat being 1062 bp and the longest forward repeat being 451 bp (Table S4).

3.4 Mitochondrial plastid DNA analysis

During mitochondrial evolution, some chloroplast genome fragments were transferred into the mitogenome to produce mitochondrial plastid DNAs (MTPTs). The 12 homologous fragments were identified between chloroplast genome and mitogenome of *D. sophia* (Fig. 5, Table 2), with a total length of 6,004 bp, accounting for 2.26% of the mitogenome. Among them, there were two fragments more than 1,000 bp, and the longest aligned fragment was MTPT7 with a length of 2,485 bp. Through annotating these 12 homologous fragments, 7 complete genes were detected (Table 2), including 1 chloroplast PCG (*psaA*) and 6 tRNA genes (*trnD-GUC*, *trnN-GUU*, *trnI-CAU*, *trnM-CAU*, *trnS-GGA*, and *trnW-CCA*). Many tRNA genes remained intact in both chloroplast and mitochondrial organelle genomes and had high sequence similarity, indicating that they can still play a role in the mitogenome of *D. sophia*.

Fig. 3 Analysis of codon preference in *D. sophia* mitogenome. Horizontally, the 21 amino acids and their corresponding codons are displayed on the x-axis. Vertically, rectangles of different color indicate RSCU values for different codons of the same amino acid

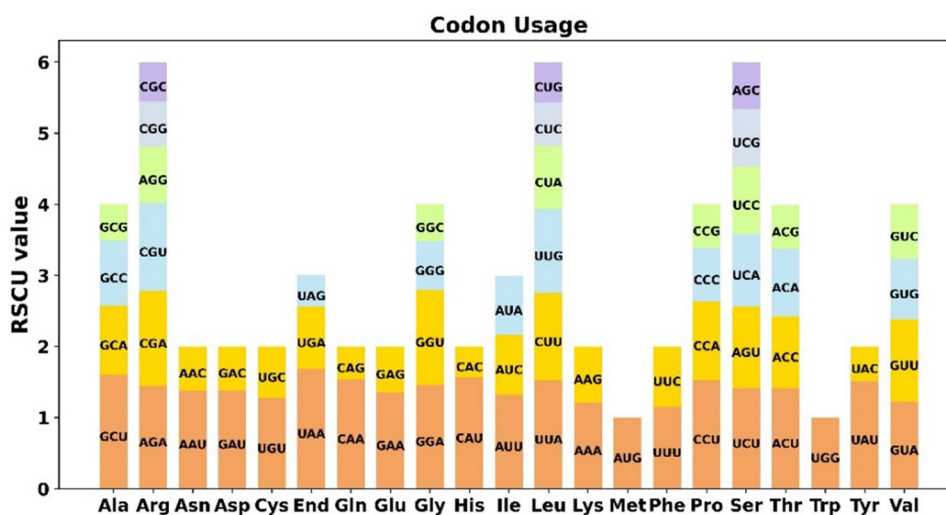


Fig. 4 Analysis of repeated sequences in *D. sophia* mitogenome. **A** Types of SSRs; **B** Types of tandem repeats and dispersed repeats

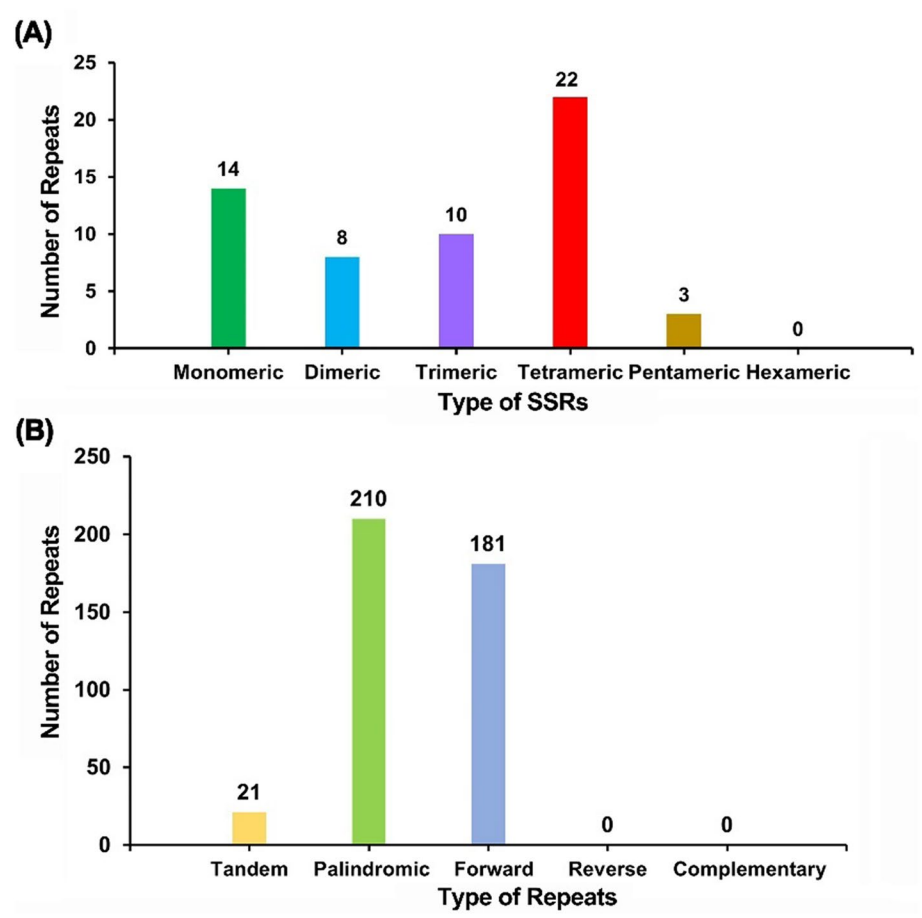


Fig. 5 Homologous fragments between chloroplast genome and mitogenome in *D. sophia*. The blue arc represents the mitogenome (mtDNA), the green arc indicates the chloroplast genome (cpDNA), and the yellow lines stand for the homologous fragments between two organelle genomes

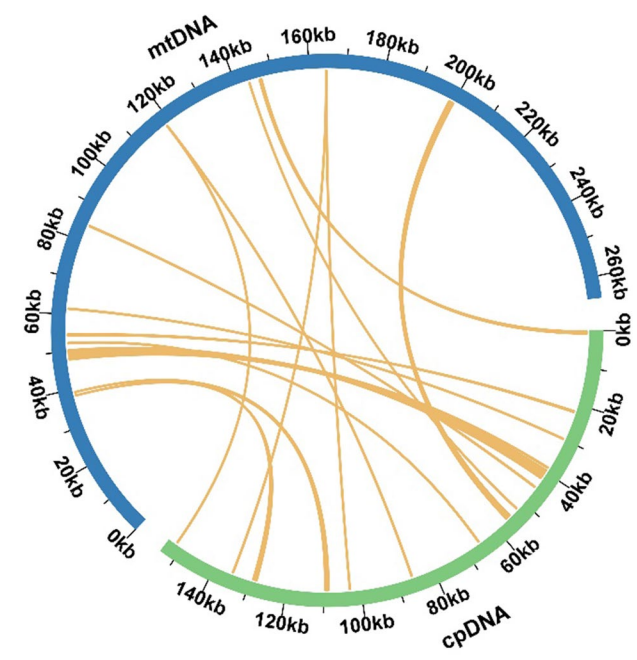


Table 2 Transferred fragments from the chloroplast genome to the mitogenome in *D. sophia*

Number	Alignment length (bp)	Identity%	Chloroplast genome		Mitochondrial genome		MTPT annotation
			Start	End	Start	End	
MTPT 1	934	98.287	128,205	129,134	38,192	39,125	Complete (<i>trnN-GUU</i>), Partial (<i>ycf1</i>)
MTPT2	84	97.619	65,845	65,928	52,356	52,274	Complete (<i>trnW-CCA</i>)
MTPT3	1,223	97.465	55,007	56,229	19,9976	198,771	Partial (<i>rbcl</i>)
MTPT4	75	97.333	29,838	29,912	61,577	61,503	Complete (<i>trnD-GUC</i>)
MTPT5	498	96.586	346	843	146,493	146,981	Partial (<i>psbA</i>)
MTPT6	58	96.552	102,961	103,018	164,531	164,474	Partial (<i>trnI-GAU</i>)
MTPT7	2,485	96.016	39,484	41,955	47,973	50,429	Partial (<i>psaB</i>), Complete (<i>psaA</i>)
MTPT8	77	92.208	52,333	52,409	143,662	143,587	Complete (<i>trnM-CAU</i>)
MTPT9	113	92.035	44,763	44,875	84,383	84,495	Complete (<i>trnS-GGA</i>)
MTPT10	127	85.827	38,586	38,708	49,279	49,405	Partial (<i>psaB</i>)
MTPT11	253	85.771	21,921	22,166	54,389	54,637	Partial (<i>rpoC1</i>)
MTPT12	77	85.714	86,011	86,085	118,693	118,621	Complete (<i>trnI-CAU</i>)
Total	6004						

3.5 Phylogenetic and synteny analysis

To investigate the phylogenetic relationships of *D. sophia* mitogenome, based on the DNA sequences of 21 common mitochondrial PCGs, we performed the phylogenetic analysis of 29 plant species from three angiosperm orders, including *D. sophia* (Fig. 6). The mitogenomes of *M. longipes* and *C. coggygia*, two species of Sapindales, were set as outgroups. The results showed that plant species from three orders, Brassicales, Malvales and Sapindales, were clustered into three separate branches. According to the topological structure of mitogenome phylogeny and the latest Angiosperm Phylogeny Group (APG IV) classification [41], *D. sophia* belonged to the Cruciferae family and was more closely related to *B. stricta* (Fig. 6).

According to the sequence similarity, MCscanX software was used to generate the multiple synteny plot of *D. sophia* with closely related species. A great deal of homologous collinearity regions were identified in the mitogenomes of *D. sophia* and six other Brassicales species, but these collinearity regions had short lengths (Fig. 7). Furthermore, some gaps were detected, indicating these sequences were unique in *D. sophia* and had no homology with six other plant species. The results showed that the collinearity regions between the mitogenomes of these seven species were highly non-conservative in the order of arrangement, and the mitogenome of *D. sophia* underwent frequent genome recombination with closely related species.

3.6 Prediction of RNA editing sites in mitogenome of *D. sophia*

The RNA editing sites of 35 PCGs in the *D. sophia* mitogenome were analyzed using Deepred-Mt software with the cutoff value = 0.9. A total of 406 potential RNA editing sites were identified in 35 mitochondrial PCGs, all of which were base C → U editing sites (Fig. S2, Table S5). Among all the PCGs of *D. sophia* mitogenome, 34 RNA editing sites were identified for the *ccmB* gene with the highest number of edits, followed by the *nad4* gene with 32 RNA editing sites. The *atp6* and *rps14* genes each have only one RNA editing site.

4 Discussion

No mitogenome data have been released for a single species of the genus *Descurainia* in the Cruciferae family. In this study, the second- and third-generation sequencing technologies (Illumina Hiseq X Ten and PacBio Sequel II platform) were adopted to resolve the mitogenome of *D. sophia*. The mitogenome of *D. sophia* had a circular structure with the size of 265,457 bp and GC content of 44.78%, which was consistent with *A. thaliana* (44.8%), *C. bursa-pastoris* (44.74%),

Fig. 6 Phylogenetic analysis of *D. sophia*. Two plant species of Sapindales, *M. longipes* and *C. coggygia*, were set as outgroups

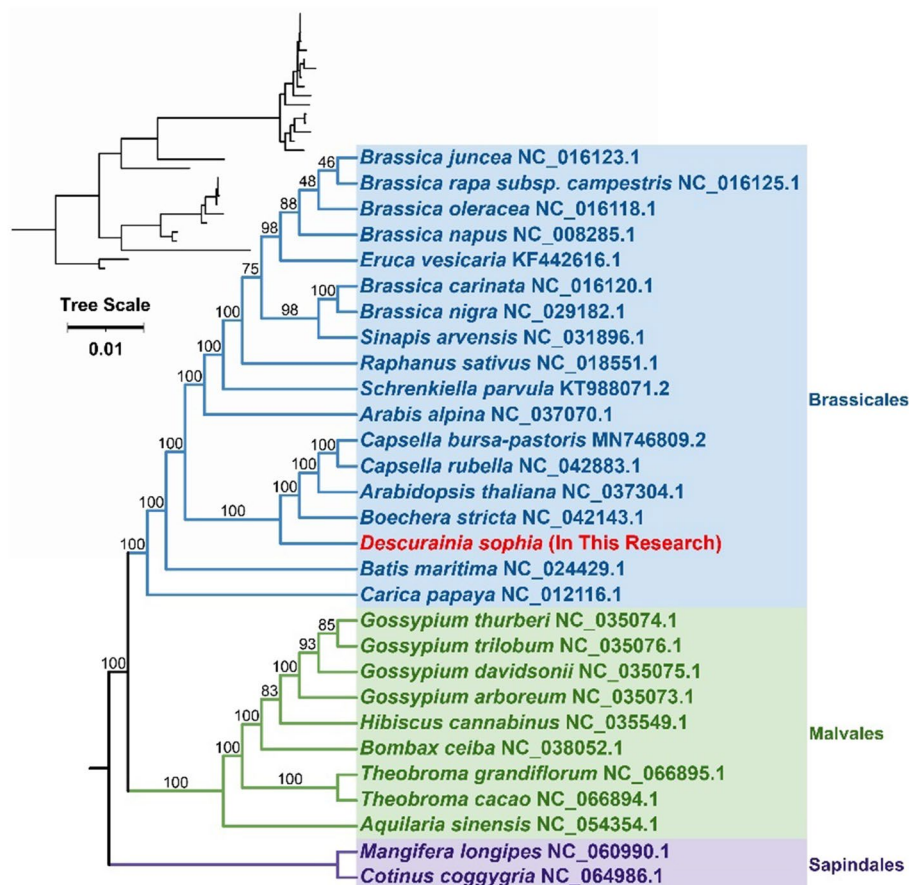
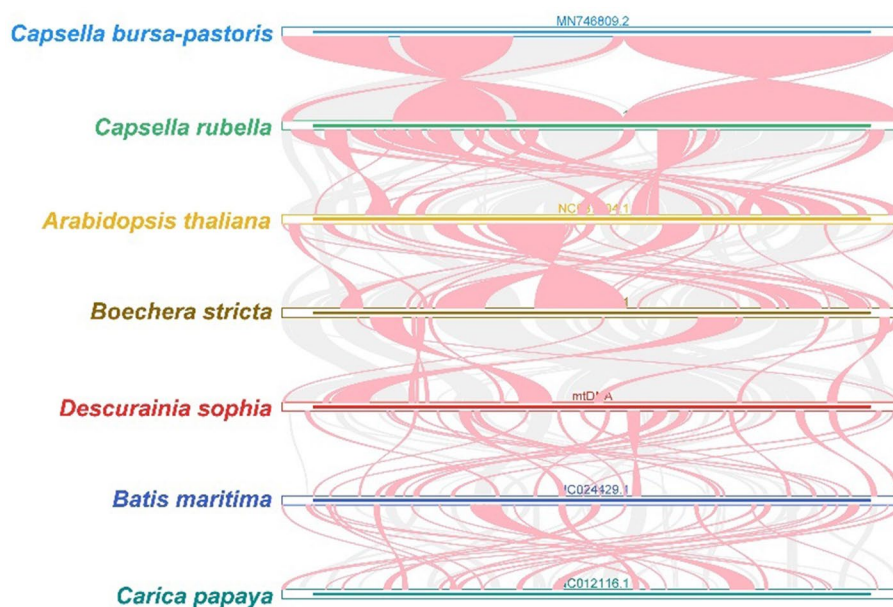


Fig. 7 Synteny analysis of *D. sophia* mitogenomes and closely related species. The horizontal bars represent mitogenomes of different plant species with different colors. The red arcs show regions where inversions occur, and the gray arcs show regions of good homology



O. sativa (43.8%), *Z. mays* (43.9%), and *Morus alba* (45.50%), indicating that the GC content was highly conserved in higher plants [42]. There were great differences in the usage of genomic codons among different species. This preference was the result of the relative balance that had gradually developed within the cells of organisms during

long-term evolutionary selection. In the mitogenome of *D. sophia*, most of the PCGs had typical ATG as start codons, and the distribution of amino acid composition was similar to that of other angiosperms [43].

The PCGs only accounted for about 11.94% of the *D. sophia* mitogenome, tRNA and rRNA genes occupied 0.68% and 1.97%, respectively, and the rest were non-coding sequences accounting for 85.41%. Mitogenome coding sequences were more conserved than non-coding sequences, but non-coding sequences were the main source of mitogenome variation [44]. The non-coding sequences of the mitogenome were mainly composed of repeated sequences, chloroplast genome homologous sequences, and nuclear genome homologous sequences. The repeated sequences can be divided into two categories: tandem repeats and dispersed repeats. The repeated sequences were essential for recombination in the mitogenome. The longest repeated sequence in a species (generally more than 1 kb in angiosperms) will constitute homologous recombination, resulting in the isomerization of the mitogenome [45]. In the mitogenome of *D. sophia*, there was only one pair of dispersed repeats longer than 1 kb with a length of 1062 bp, and the rest of the dispersed repeats were 30–713 bp in length. Homologous recombination mediated by repeated sequences was commonly found in plant mitogenomes. The size of repeated sequences has been reported to be closely associated to the frequency of recombination, e.g., recombination mediated by short repeated sequences tends to be less frequent than that mediated by long repeated sequences [15]. For example, it has been found that in the mitogenomes of *Scutellaria tsinyunensis* [46], *Abelmoschus esculentus* [47] and *Prunus salicina* [48], the recombination frequency mediated by long repeated sequences was high, while that of short repeat sequences was low. In the mitogenome of *D. sophia*, only a high frequency of recombination with a pair of long-repeated sequences was detected, and some short-repeated sequences that may be involved in recombination have not been identified. The degree of variation and structural heterogeneity of mitogenome varies significantly among Brassicaceae. Through analysis of 16 mitogenomes (Brassicaceae plants), Liu found that the mitogenome of *Meniocus linifolius* has four conformations, with more loss of PCGs and ambiguous phylogenetic status. The mutation rate and structural heterogeneity of *M. linifolius* were significantly accelerated compared with other Brassicaceae plants [49]. With the release of more mitogenomes of genus *Descurainia* plants, it is possible to analyze the mitogenome size, the number of genes (gene loss), repetitive sequences, variation of genes, and the evolutionary impact of these variations at the genus level.

In the process of plant mitochondrial evolution, some chloroplast genome fragments were transferred into the mitogenome, and the length and sequence similarity of the transferred fragments varied among species. The 12 fragments were identified to migrate from the chloroplast genome to the *D. sophia* mitogenome, including 7 complete genes (1 PCG and 6 tRNA genes) with a total length of 6 kb, accounting for 2.26% of the mitogenome. The 23 chloroplast migrated fragments were detected in the mitogenome of *Quercus acutissima*, including 13 complete genes (2 PCGs, 10 tRNAs and 1 rRNA genes) with a total length of 15.69 kb, representing 3.49% of the mitogenome [50]. These were the typical ratios for angiosperms, with the total length of chloroplast migrated fragments ranging from 4.4 kb in *Arabidopsis* [51] to 138 Kb in *Amborella* [52]. In angiosperms, the migration of tRNA genes from chloroplasts to mitochondria was common, indicating that tRNA genes were more conserved in the mitogenome than PCGs genes, and they may play an indispensable role in mitochondria [53]. Those transferred chloroplast PCGs contained only partial sequences, indicating that they experienced a certain degree of gene loss. Considering the low substitution rate of the mitogenome, mitochondrial genes were important reference gene sequences for phylogenetic analysis at the taxonomic level [54]. Phylogenetic and synteny analysis revealed that *D. sophia* was closely related to *B. stricta*. The chromosome number of *B. stricta* was $n=7$ [55], while the genome of *D. sophia* was $2n=4x=28$ (unpublished data). This may have occurred with chromosomal changes such as chromosome fusion and whole genome duplication.

RNA editing was prevalent in the mitogenomes and chloroplast genomes of plants, and was one of the required steps for mitochondrial and chloroplast gene expression. RNA editing belonged to the post-transcriptional modification, which contributed to improve protein folding and function [18]. It was reported that two key RNA editing sites in cotton mitochondrial gene *atp1* mRNA were necessary for cotton fiber cell elongation and were closely associated with cotton fiber length [56]. RNA editing factor SIORRM4 was required during tomato ripening. When the SIORRM4 gene was knocked out, the RNA editing sites of many mitochondrial genes (e.g., *nad3*, *cytc1*, *cox2*) transcripts were reduced, resulting in delayed maturation of tomato [57]. These results indicated that mitochondrial RNA editing was closely associated to important cultivation traits in plants. Previous studies reported that 441 RNA editing sites were identified in 36 genes of the *A. thaliana* mitogenome [51], and 491 sites in 34 genes of the *O. sativa* mitogenome [58]. We identified 406 C → U RNA editing sites in 35 PCGs of the *D. sophia* mitogenome. The number of RNA editing sites varied widely among different mitochondrial genes, with the cytochrome *c* biogenesis gene (*ccmB*) having the highest number, followed by the NADH dehydrogenase gene (*nad4*). Currently, an efficient mitochondrial genome editor (mitoTALECD) had been developed in *A. thaliana*, which enabled editing of the *A. thaliana* mitogenome C → U with editing efficiency up to 100%,

and the resulting homogeneous mutations can be stably inherited to the next generation [59]. Therefore, the analysis of RNA editing sites provided a basis for predicting the gene function of new codons in the future, and was more helpful to understand the expression of mitochondrial and chloroplast genes in plants.

5 Conclusions

In this study, we sequenced, assembled and annotated the *D. sophia* mitogenome using second- and three-generation high-throughput sequencing technology on the Illumina HiSeq X Ten and PacBio Sequel II platform. According to the assembly results, the main structure of the *D. sophia* mitogenome was a circular DNA molecule without branches. The size of the *D. sophia* mitogenome was 265,457 bp, with the GC content of 44.78%, which encoded 56 unique genes, including 35 PCGs, 18 tRNA and 3 rRNA genes. In addition, we analyzed codon preference, repeated sequences, DNA sequence transformation, phylogenetic relationship, synteny and RNA editing sites in the *D. sophia* mitogenome. The phylogenetic and synteny analysis indicated that *D. sophia* was closely related to *B. stricta*, and the *D. sophia* mitogenome underwent frequent genome recombination. This study provided the essential information for future studies on genetic breeding, species identification and development of new molecular markers in *D. sophia*.

Acknowledgements We thank Mrs. Kai Tan for assisting in the preparation of this manuscript. We also thank the experimentalists and bioinformatics analysts of Wuhan Bena Technology Co., Ltd. and MitoRun Research Group for their participation in this project.

Author contributions LZ, YhZ, HyJ and HY performed the experiments; LZ, YgH, HY, and MW analyzed the data; LZ and YhZ drafted the manuscript; MW, XkZ and WsF reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding This work was supported by National Key Research and Development Project (2019YFC1708802), Major Science and Technology Project in Henan Province (171100310500), Henan Province High-Level Personnel Special Support “ZhongYuan One Thousand People Plan” (ZYQR201810080), Training Program for Young Teachers in Colleges and Universities of Henan Province (2021GGJS086).

Data availability The assembled mitogenome of *D. sophia* had been submitted to the NCBI GenBank database (accession number of OQ916154). The raw data (Illumina and PacBio) had been submitted to SRA database (accession number PRJNA1078884).

Code availability Not applicable.

Declarations

Competing interests The authors declare that they have no competing interest.

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/>.

References

1. Chinese Pharmacopoeia Commission. The Pharmacopoeia of the People's Republic of China, 2020 edition, vol. 1. Beijing: China Medical Science Press; 2020:348.
2. Li H, Zheng X, Gong J, Feng W. Research progress in chemical constituents of *Lepidium apetalum* and *Descurainia sophia* and their pharmacological activities. Drug Eval Res. 2013;36(3):235–40. <https://doi.org/10.7501/j.issn.1674-6376.2013.03.019>.
3. Zhou X, Tang L, Zhou G, Kou Z, Wang T, Wang Z. Advances on *Lepidii Semen* and *Descurainiae Semen*. China J Chin Mater Med. 2014;39(24):4699–708.
4. Lee YJ, Kim NS, Kim H, Yi J, Oh S, Bang O, Lee J. Cytotoxic and anti-inflammatory constituents from the seeds of *Descurainia sophia*. Arch Pharm Res. 2013;36(5):536–41. <https://doi.org/10.1007/s12272-013-0066-x>.
5. Sun K, Li X, Liu JM, Wang JH, Li W, Sha Y. A novel sulphur glycoside from the seeds of *Descurainia sophia* (L.). J Asian Nat Prod Res. 2005;7(6):853–6. <https://doi.org/10.1080/1028602042000204072>.
6. Chen J, Guan R. Extraction and identification of several glucosinolates in *Descurainia sophia* seeds. Acta Bot Boreali-Occident Sin. 2006;26(6):1231–5.
7. Afsharypuor S, Lockwood GB. Glucosinolate degradation products, alkanes and fatty acids from plants and cell cultures of *Descurainia sophia*. Plant Cell Rep. 1985;4(6):341–4. <https://doi.org/10.1007/bf00269894>.

8. Lockwood GB, Afsharypour S. Comparative study of the volatile aglucones of glucosinolates from in vivo and in vitro grown *Descurainia sophia* and *Alyssum minimum* using gas chromatography—mass spectrometry. *J Chromatogr A*. 1986;356:438–40. [https://doi.org/10.1016/s0021-9673\(00\)91513-9](https://doi.org/10.1016/s0021-9673(00)91513-9).
9. Feng W, Li C, Zheng X, Li L, Chen W, Zhang Y, Cao Y, Gong J, Kuang H. Three new sulphur glycosides from the seeds of *Descurainia sophia*. *Nat Prod Res*. 2016;30(15):1675–81. <https://doi.org/10.1080/14786419.2015.1135141>.
10. Sønderby IE, Geu-Flores F, Halkier BA. Biosynthesis of glucosinolates—gene discovery and beyond. *Trends Plant Sci*. 2010;15(5):283–90. <https://doi.org/10.1016/j.tplants.2010.02.005>.
11. Qian L, Jiang J, Gao X, Zhang X. Analysis of fatty oil from seeds of *Descurainia sophia* by GC-MS. *J Plant Resour Environ*. 2006;15(1):76–7.
12. Kim K, Nam YA, Kim HS, Hayes AW, Lee B. α -Linolenic acid: nutraceutical, pharmacological and toxicological evaluation. *Food Chem Toxicol*. 2014;70:163–78. <https://doi.org/10.1016/j.fct.2014.05.009>.
13. Møller IM, Rasmussen AG, Van Aken O. Plant mitochondria: past, present and future. *Plant J*. 2021;108(4):912–59. <https://doi.org/10.1111/tpj.15495>.
14. Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T, Ohyama K. Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA: a primitive form of plant mitochondrial genome. *J Mol Biol*. 1992;223(1):1–7. [https://doi.org/10.1016/0022-2836\(92\)90708-R](https://doi.org/10.1016/0022-2836(92)90708-R).
15. Gualberto JM, Milesina D, Wallet C, Niaz AK, Weber-Lotfi F, Dietrich A. The plant mitochondrial genome: dynamics and maintenance. *Biochimie*. 2014;100:107–20. <https://doi.org/10.1016/j.biochi.2013.09.016>.
16. Yang H, Chen H, Ni Y, Li J, Cai Y, Ma B, Yu J, Wang J, Liu C. *De Novo* hybrid assembly of the *Salvia miltiorrhiza* mitochondrial genome provides the first evidence of the multi-chromosomal mitochondrial DNA structure of *Salvia* species. *Int J Mol Sci*. 2022;23(22):14267. <https://doi.org/10.3390/ijms232214267>.
17. Liu J, Ni Y, Liu C. Polymeric structure of the *Cannabis sativa* L. mitochondrial genome identified with an assembly graph model. *Gene*. 2023;853:147081. <https://doi.org/10.1016/j.gene.2022.147081>.
18. Cheng Y, He X, Priyadarshani SVGN, Wang Y, Ye L, Shi C, Ye K, Zhou Q, Luo Z, Deng F, Cao L, Zheng P, Aslam M, Qin Y. Assembly and comparative analysis of the complete mitochondrial genome of *Suaeda glauca*. *BMC Genomics*. 2021;22(1):167. <https://doi.org/10.1186/s12864-021-07490-9>.
19. Allen GC, Flores-Vergara MA, Krasynanski S, Kumar S, Thompson WF. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat Protoc*. 2006;1(5):2320–5. <https://doi.org/10.1038/nprot.2006.384>.
20. Jin J, Yu W, Yang J, Song Y, dePamphilis CW, Yi T, Li D. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. 2020;21(1):241. <https://doi.org/10.1186/s13059-020-02154-5>.
21. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 2019;37(5):540–6. <https://doi.org/10.1038/s41587-019-0072-8>.
22. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.
23. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPPD. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
24. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq—versatile and accurate annotation of organelle genomes. *Nucl Acid Res*. 2017;45(W1):W6–11. <https://doi.org/10.1093/nar/gkx391>.
25. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucl Acid Res*. 1997;25(5):955–64. <https://doi.org/10.1093/nar/25.5.955>.
26. Chen Y, Ye W, Zhang Y, Xu Y. High speed BLASTN: an accelerated MegaBLAST search tool. *Nucl Acid Res*. 2015;43(16):7762–8. <https://doi.org/10.1093/nar/gkv784>.
27. Lewis SE, Searle SMJ, Harris N, Gibson M, Iyer V, Richter J, Wiel C, Bayraktaroglu L, Birney E, Crosby MA, Kaminker JS, Matthews BB, Prochnik SE, Smith CD, Tupy JL, Rubin GM, Misra S, Mungall CJ, Clamp ME. Apollo: a sequence annotation editor. *Genome Biol*. 2002;3(12):research0082.1–14. <https://doi.org/10.1186/gb-2002-3-12-research0082>.
28. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT. 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. <https://doi.org/10.1111/1755-0998.13096>.
29. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33(7):1870–4. <https://doi.org/10.1093/molbev/msw054>.
30. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics*. 2017;33(16):2583–5. <https://doi.org/10.1093/bioinformatics/btx198>.
31. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucl Acid Res*. 1999;27(2):573–80. <https://doi.org/10.1093/nar/27.2.573>.
32. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucl Acid Res*. 2001;29(22):4633–42. <https://doi.org/10.1093/nar/29.22.4633>.
33. Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucl Acid Res*. 2019;47(W1):W65–73. <https://doi.org/10.1093/nar/gkz345>.
34. Liu S, Ni Y, Li J, Zhang X, Yang H, Chen H, Liu C. CPGView: a package for visualizing detailed chloroplast genome structures. *Mol Ecol Resour*. 2023;23(3):694–704. <https://doi.org/10.1111/1755-0998.13729>.
35. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80. <https://doi.org/10.1093/molbev/mst010>.
36. 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. <https://doi.org/10.1093/molbev/msu300>.
37. Letunic I, Bork P. Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucl Acid Res*. 2019;47(W1):W256–9. <https://doi.org/10.1093/nar/gkz239>.

38. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T-h, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH. MCSanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucl Acid Res.* 2012;40(7):e49. <https://doi.org/10.1093/nar/gkr1293>.
39. Edera AA, Small I, Milone DH, Sanchez-Puerta MV. Deepred-Mt: deep representation learning for predicting C-to-U RNA editing in plant mitochondria. *Comput Biol Med.* 2021;136: 104682. <https://doi.org/10.1016/j.compbio.2021.104682>
40. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics.* 2015;31(20):3350–2. <https://doi.org/10.1093/bioinformatics/btv383>.
41. The Angiosperm Phylogeny G. 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. <https://doi.org/10.1111/boj.12385>.
42. Liangliang G, Yisu S, Mengmeng W, Ackah M, Peng G, Danyan Z, Lin Q, Zhao W. The complete mitochondrial genome sequence variation and phylogenetic analysis of mulberry. *J Data Min Genom Proteom.* 2022;13:249. <https://doi.org/10.4172/2153-0602.22.13.249>.
43. Sloan DB, Wu Z, Sharbrough J. Correction of persistent errors in *Arabidopsis* reference mitochondrial genomes. *Plant Cell.* 2018;30(3):525–7. <https://doi.org/10.1105/tpc.18.00024>.
44. Christensen AC. Plant mitochondrial genome evolution can be explained by DNA repair mechanisms. *Genome Biol Evol.* 2013;5(6):1079–86. <https://doi.org/10.1093/gbe/evt069>.
45. Wynn EL, Christensen AC. Repeats of unusual size in plant mitochondrial genomes: identification, incidence and evolution. *G3 Genes Genomes Genet.* 2019;9(2):549–59. <https://doi.org/10.1534/g3.118.200948>.
46. Li J, Xu Y, Shan Y, Pei X, Yong S, Liu C, Yu J. Assembly of the complete mitochondrial genome of an endemic plant, *Scutellaria tsinyunensis*, revealed the existence of two conformations generated by a repeat-mediated recombination. *Planta.* 2021;254(2):36. <https://doi.org/10.1007/s00425-021-03684-3>.
47. 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(1):481. <https://doi.org/10.1186/s12864-022-08706-2>.
48. Fang B, Li J, Zhao Q, Liang Y, Yu J. Assembly of the complete mitochondrial genome of Chinese plum (*Prunus salicina*): characterization of genome recombination and RNA editing sites. *Genes.* 2021;12(12):1970. <https://doi.org/10.3390/genes12121970>.
49. Liu J, Hu JY, Li DZ. Remarkable mitochondrial genome heterogeneity in *Meniocus linifolius* (Brassicaceae). *Plant Cell Rep.* 2024;43(2):36. <https://doi.org/10.1007/s00299-023-03102-w>.
50. Liu D, Guo H, Zhu J, Qu K, Chen Y, Guo Y, Ding P, Yang H, Xu T, Jing Q, Han S, Li W, Tong B. Complex physical structure of complete mitochondrial genome of *Quercus acutissima* (Fagaceae): a significant energy plant. *Genes.* 2022;13(8):1321. <https://doi.org/10.3390/genes13081321>.
51. Giegé P, Brennicke A. RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. *Proc Nat Acad Sci USA.* 1999;96(26):15324–9. <https://doi.org/10.1073/pnas.96.26.1532>.
52. Bergthorsson U, Richardson AO, Young GJ, Goertzen LR, Palmer JD. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. *Proc Nat Acad Sci USA.* 2004;101(51):17747–52. <https://doi.org/10.1073/pnas.040833610>.
53. Bi C, Paterson AH, Wang X, Xu Y, Wu D, Qu Y, Jiang A, Ye Q, Ye N. Analysis of the complete mitochondrial genome sequence of the diploid cotton *Gossypium raimondii* by comparative genomics approaches. *BioMed Res Int.* 2016;2016:5040598. <https://doi.org/10.1155/2016/5040598>.
54. Hiesel R, von Haeseler A, Brennicke A. Plant mitochondrial nucleic acid sequences as a tool for phylogenetic analysis. *Proc Nat Acad Sci USA.* 1994;91(2):634–8. <https://doi.org/10.1073/pnas.91.2.634>.
55. Schranz ME, Windsor AJ, Song BH, Lawton-Rauh A, Mitchell-Olds T. Comparative genetic mapping in *Boechea stricta*, a close relative of *Arabidopsis*. *Plant Physiol.* 2007;144(1):286–98. <https://doi.org/10.1104/pp.107.096685>.
56. He P, Xiao G, Liu H, Zhang L, Zhao L, Tang M, Huang S, An Y, Yu J. Two pivotal RNA editing sites in the mitochondrial *atp1* mRNA are required for ATP synthase to produce sufficient ATP for cotton fiber cell elongation. *New Phytol.* 2018;218(1):167–82. <https://doi.org/10.1111/nph.14999>.
57. Yang Y, Zhu G, Li R, Yan S, Fu D, Zhu B, Tian H, Luo Y, Zhu H. The RNA editing factor SIORRM4 is required for normal fruit ripening in tomato. *Plant Physiol.* 2017;175(4):1690–702. <https://doi.org/10.1104/pp.17.01265>.
58. Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol Genet Genomics.* 2002;268(4):434–45. <https://doi.org/10.1007/s00438-002-0767-1>.
59. Nakazato I, Okuno M, Zhou C, Itoh T, Tsutsumi N, Takenaka M, Arimura S-i. Targeted base editing in the mitochondrial genome of *Arabidopsis thaliana*. *Proc Nat Acad Sci USA.* 2022;119(20): e2121177119. <https://doi.org/10.1073/pnas.2121177119>.