

RESEARCH

Open Access



# Assembly and comparative analysis of the complete mitochondrial genome of *Brassica rapa* var. *Purpuraria*

Yihui Gong<sup>1\*</sup>, Xin Xie<sup>1</sup>, Guihua Zhou<sup>1</sup>, Meiyu Chen<sup>1</sup>, Zhiyin Chen<sup>1</sup>, Peng Li<sup>3</sup> and Hua Huang<sup>2\*</sup>

## Abstract

**Background** Purple flowering stalk (*Brassica rapa* var. *purpuraria*) is a widely cultivated plant with high nutritional and medicinal value and exhibiting strong adaptability during growing. Mitochondrial (mt) play important role in plant cells for energy production, developing with an independent genetic system. Therefore, it is meaningful to assemble and annotate the functions for the mt genome of plants independently. Though there have been several reports referring the mt genome of in *Brassica* species, the genome of mt in *B. rapa* var. *purpuraria* and its functional gene variations when compared to its closely related species has not yet been addressed.

**Results** The mt genome of *B. rapa* var. *purpuraria* was assembled through the Illumina and Nanopore sequencing platforms, which revealed a length of 219,775 bp with a typical circular structure. The base composition of the whole *B. rapa* var. *purpuraria* mt genome revealed A (27.45%), T (27.31%), C (22.91%), and G (22.32%). 59 functional genes, composing of 33 protein-coding genes (PCGs), 23 tRNA genes, and 3 rRNA genes, were annotated. The sequence repeats, codon usage, RNA editing, nucleotide diversity and gene transfer between the cp genome and mt genome were examined in the *B. rapa* var. *purpuraria* mt genome. Phylogenetic analysis show that *B. rapa* var. *Purpuraria* was closely related to *B. rapa* subsp. *Oleifera* and *B. juncea*. Ka/Ks analysis reflected that most of the PCGs in the *B. rapa* var. *Purpuraria* were negatively selected, illustrating that those mt genes were conserved during evolution.

**Conclusions** The results of our findings provide valuable information on the *B. rapa* var. *Purpuraria* genome, which might facilitate molecular breeding, genetic variation and evolutionary researches for *Brassica* species in the future.

**Keywords** *Brassica rapa* var. *Purpuraria*, Mitochondrial genome, Repeat sequence, RNA editing, Phylogenetic analysis

\*Correspondence:

Yihui Gong

gyhzgh@163.com

Hua Huang

huangw0109@gmail.com

<sup>1</sup> Development and Utilization and Quality and Safety Control of Characteristic Agricultural Resources in Central Hunan, College of Agriculture and Biotechnology, Hunan University of Humanities, Science and Technology, Loudi 417000, China

<sup>2</sup> Institute of Fruit Tree Research, Key Laboratory of South Subtropical Fruit Biology and Genetic Resource Utilization, Ministry of Agriculture and Rural Affairs, Guangdong Provincial Key Laboratory of Tropical and Subtropical Fruit Tree Research, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China

<sup>3</sup> Xiangtan Agricultural Science Research Institute, Xiangtan 411100, China



© The Author(s) 2024. **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.

## Introduction

Purple flowering stalks (*Brassica rapa* var. *purpuraria*) is widely distributed in the middle regions of the Yangtze River that belongs to the Cruciferae family [1]. *B. rapa* var. *Purpuraria* is an important and popular vegetable for consumers due to its bright color and delicious taste with abundant anthocyanidins, carotenoids, proanthocyanidins, vitamin C and mineral elements [2, 3]. Mitochondria (mt) is important organelles that involve in many metabolic processes related to the synthesis of these nutritional components of amino acids, lipids and vitamins and energy production [4–6]. The mt genome of angiosperm *Arabidopsis thaliana* was first reported in 1997 [7]. Subsequently, the mt genomes of some field crops and fruits, including rice (*Oryza sativa* L.) [8], rape (*Brassica napus* L.) [9], corn (*Zea mays* L.) [10], grape (*Vitis vinifera*) [11], apple (*Malus domestica*) [12], and kiwifruit (*Actinidia chinensis*) [13], have been successively published. The mt genomes have the characteristics of integrity, polymorphism, and semi-autonomy with a unique expression system, they contain a few genes and limited the types of proteins. Therefore, they need to be coordinated with nuclear genes to maintain normal biological functions [14]. The number, arrangement and composition of genes are conserved in the chloroplast (cp) genomes of higher plants [15]. However, Mt genomes have highly conserved characteristics and evolutionary rates that are different from nuclear genes. Therefore, the mt genome is relatively large, which could provide a large amount of genetic information and solve the problem of classification, identification and evolution of related species [14, 16].

The mt genome has significant differences in length, gene sequence, and gene content in different species, and even varies in different cultivars in the same species [17–19]. The length of mt genome in plants varies from 66 Kb to 11.7 Mb [18, 20]. The plant mt genomes mainly contain 50–60 conserved genes, which involved in oxidative phosphorylation and protein translation, and many unknown function open reading frames (ORFs) [21]. In addition, Some of ORFs in the mt genome play a key role in cytoplasmic male sterility of species [22–24]. Except for the unknown function ORFs, the differences in the number of complex II subunit, ribosomal protein genes, tRNAs and multi-copy genes were the main reason for the difference in the number of mt genes in different species [25]. With the development of sequencing technology and the decreasing of sequencing costs, a variety of mt genomes have been reported. The mt genome of *Brassica* species including *A. thaliana* [26], *Raphanus sativus* L. [27], and *B. napus* [28] have been reported. In addition, the mt genomes of several varieties of *B. rapa* species have also been addressed. The mt genomes in the

varieties of *B. rapa* developed with close full length from 219,736 bp to 219,775 bp, with same number of tRNA (18) and rRNA (3); and exhibited minor differences for gene numbers between 54 and 99, and PCGs ranged from 34 to 78 [29–31]. The characterized mt genome could help to observe structural variations in the evolutionary history of *Brassica* species or varieties. Therefore, assembling and analyzing the mt genomes is important to better understanding of their genetic characteristics and for molecular breeding research.

In this work, the whole mt genome of *B. rapa* var. *Purpuraria* was sequenced and assembled using the Illumina and Nanopore sequencing platforms. The genomic features, repeat sequences, codon usage, RNA editing sites, and comparative genomics were analyzed. We also conducted phylogenetic analysis to understand the genetic variations in *B. rapa* var. *Purpuraria* more effectively. This study enhances our understanding of *B. rapa* var. *Purpuraria* genetics and provides useful information for future researches on identification, molecular breeding and system evolution of mt genomes of Brassicaceae species.

## Materials and methods

### Plant materials, DNA extraction and sequencing

The *B. rapa* var. *Purpuraria* seeds were provided by Peng Li (Xiangtan Agricultural Science Research Institute), and cultivated at the Xiangtan Agricultural Science Research Institute (Yuhu District, Xiangtan, Hunan, China; 27°52'N, 112°50'E) under natural conditions. Fresh young leaves were collected and quickly frozen in liquid nitrogen, and then stored at 80 °C. Plant specimens were conserved in the Hunan University of Humanities, Science and Technology (accession number: 20231218BRP02). The total DNA isolation from the young leaves was performed using CTAB method [32] and purified using Plant DNA Mini Kit (D311, Genepioneer Biotechnologies, Nanjing, China) according to the manufacturer's protocol. The qualified DNA samples was sequenced with 500 bp paired-end (PE) library construction using the VAHTS Universal DNA Library Prep Kit for IlluminaV3 (Vazyme Biotech Co., Ltd, Nanjing, China). About 29,356,547,400 raw data from *B. rapa* var. *Purpuraria* were generated with PE150 sequencing strategy. Subsequently, the qualified DNA was cut into 20-kb fragments using a Covaris tube (Covaris) and purified with AMPure beads. The samples were sequenced with Oxford Nanopore library construction.

### Mitochondrial genome assembly and annotation

We used the Fastp 0.23.4 (<https://github.com/OpenGene/fastp>) software to filter the second-generation raw data. The parameters were set as follows: (1) Cut off

the primer and adapters sequences; (2) Filter out the reads with average quality value lower than Q5. (3) Delete the reads with the number of uncertain bases more than 5. Then, the original tri-generational data was filtered using `filtlongv0.2.1` (<https://link.zhihu.com/?target=https%3A//github.com/rswick/Filtlong>) with parameters set as: `-min_length 1000` and `-min_mean_q7`. The highly quality tri-generational data were aligned with the plant mt gene database ([https://github.com/xul962464/plant\\_mt\\_ref\\_gene](https://github.com/xul962464/plant_mt_ref_gene)) using `minimap2` [33]. The size of sequence more than 50 bp, containing multiple core genes and higher alignment quality was selected as the seed sequence. Then, the original tri-generational data were compared with the seed sequences using `minimap2`, and the sequences with overlap more than 1 kb were screened and added to the seed sequences. The original data were compared to the seed sequence iteratively, and all the third-generation sequencing data of the mt genome were obtained. All the third-generation sequencing data were performed self-correction and assembled using the `canu v2.0` program [34], and `Bowtie2 (v2.3.5.1)` was used to align the second-generation data to the corrected sequence. The second-generation data were compared to stitch the corrected third-generation data using `Unicycler (v0.4.8)` software with default parameters, followed by using `Bandage (v0.8.1)` software to visualize and manually adjust the stitching results. The corrected third-generation sequencing data were aligned to the contig obtained by the second step of `Unicycler` using `minimap2`, and the branch direction was manually determined to obtain the final assembly result.

The encoded proteins and rRNAs were aligned to reported plant mt sequences using BLAST, and further manually adjusted according to closely related species. The tRNAs were annotated using `tRNAscanSE` online tool [35] (<http://lowelab.ucsc.edu/tRNAscan-SE/>). ORFs were annotated using `Open Reading Frame Finder` (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) with a minimum length of 102 bp. The map of *B. rapa* var. *Purpuraria* mt genome was drawn using `OGDRAW` (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) software. A single nucleotide polymorphism (SNP) was detected among six *Brassica* mt genomes using `MUMmer` and `BLAT v35` softwares.

#### Relative synonymous codon usage (RSCU) analysis.

The codon composition of the mt genome of *B. rapa* var. *Purpuraria* was analyzed using a Perl script written by ourselves, to select for a unique coding sequences (CDS) and calculate the RSCU of synonymous codons.

#### Analysis of repeat sequences

Dispersed repeat sequences, including forward repeats, backward repeats, reverse repeats, and complementary repeats, were detected using `blastn v2.10.1` software with parameters set as `-word_size 7` and `evaluate 1e-5`. Tandem repeats were identified using `trf409.linux64` software with parameter set as: `2 7 7 80 10 50 2000 -f -d -m`. SSRs were identified with the `MISA v1.0` tool [36]. The motif length of one- to six- nucleotide SSRs was set as 10, 5, 4, 3, 3 and 3, respectively.

#### Identification of RNA editing sites

RNA editing sites in the PCGs of *B. rapa* var. *Purpuraria* were analyzed using the `PmtREP` program (<http://cloud.genepioneer.com:9929/#/tool/alltool/detail/336>) with the cutoff value set as 0.2 [37].

#### Gene transfer between the cp genome and mt genome

The *B. rapa* var. *Purpuraria* mt genome was aligned with its cp genome (PP191173) by `BLAST` and the selected parameters were set as the matching rate  $\geq 70\%$ ,  $E\text{-value} \leq 1e^{-5}$  and length  $\geq 30$  bp [38].

#### Ka/Ks and Pi analysis

We analyzed the nonsynonymous (Ka) and synonymous (Ks) substitution rates of each PCG between *B. rapa* var. *Purpuraria* and *Cucurbita pepo* (NC\_014050.1), *Helianthus grosseserratus* (NC\_051989.1), *Brassica oleracea* (NC\_016118.1), *Brassica juncea* (NC\_016123.1), *Glycyrrhiza uralensis* (NC\_053919) and *Solanum lycopersicum* (MF034192). Homologous gene pairs were aligned in `mafft v7.310` (<https://mafft.cbrc.jp/alignment/software/>). Ka, Ks, and Ka/Ks values of each PCG were calculated using `KaKs-Calculator v2.0` tool (<https://sourceforge.net/projects/kakscalculator2/>) with `MLWL` calculation method. `Dnasp5` software was used to calculate the Pi value of each gene.

#### Phylogenetic analysis

A total of 25 whole mt genomes were used to make sure the phylogenetic position of *B. rapa* var. *Purpuraria*. The 31 mt PCG genes (*atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFc*, *ccmFn*, *cob*, *cox1*, *cox2*, *cox3*, *mttB*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*, *rpl16*, *rpl2*, *rpl5*, *rps12*, *rps14*, *rps3*, *rps4*, and *rps7*) conserved across the 25 analyzed species were aligned using `Mafft v7.427` software with default parameters. Alignments were trimmed in `trimAl` with substitution model selecting in `ModelFinder` [39, 40]. Subsequently, a maximum likelihood tree was constructed by `IQ-TREE v1.6.12` software using the `mtMet+F+R5` model with a bootstrap of 1000 [41].

*Ginkgo biloba* (NC\_027976) was used as the outgroup in this analysis.

### Results

#### Characteristics of the *B. rapa* var. *Purpuraria* mt genome

In this study, the mt genome of *B. rapa* var. *Purpuraria* was sequenced 29,356,547,400 raw data and 97,855,158 bp clean data (Q20=97.11% and Q30=91.81%) were obtained using the Illumina platform (Table S1). Regarding the Nanopore sequencing, a total of 17,960,842,101 bases and 1,417,067 reads were obtained using the Nanopore sequencing platform. The subreads with N50 and

the mean read were 27,413 bp and 12,674 bp, respectively (Table S2). The whole mt genome of *B. rapa* var. *Purpuraria* was 219,775 bp in length with a typical circular structure (Fig. 1). The nucleotide composition of the whole *B. rapa* var. *Purpuraria* mt genome contains 27.45% of A, 27.31% of T, 22.91% of C, and 22.32% of G, with GC and AT contents accounted for 45.23% and 54.77%, respectively (Table S3). PCGs and cis-spliced introns occupied 13.22% and 12.86% of the complete mt genome, whereas tRNA and rRNA genes only made up 0.79% and 2.34%, respectively. A total of 59 genes, comprising 33 PCGs, 3 rRNAs, and 23 tRNAs, were found



**Fig. 1** The circular map of the *B. rapa* var. *Purpuraria* mt genome

in the *B. rapa* var. *Purpuraria* mt genome (Table 2). Six genes, namely, *ccmFc*, *cox2*, *rpl2*, *rps3*, *trnI-AAT*, and *trnT-GGT* included one intron, genes of *nad1*, *nad2*, *nad5*, and *nad7* contained four introns, and one gene of *nad4* had three introns. Three tRNA genes were identified in two or three copies (*trnH-GTG*, *trnM-CAT* and *trnY-GTA*) (Fig. 1 and Table 1).

Plant mt genomes have significantly different in size, gene order and content [42]. We selected six *Brassica* mt genomes to compare genome characteristics and determine variability of the mt genome of *B. rapa* var. *Purpuraria* (Table 2). The size of selected mt genomes varied from 219,736 bp (*B. rapa* ssp. *rapa*) to 232,145 bp (*B. nigra*). The smallest number of genes (53) was found in *B. nigra*, and the largest (106) in *B. napus*. The number of tRNA genes ranged from 17 in *B. napus*, *B. nigra*, and *B. oleracea* to 23 in *B. rapa* var. *Purpuraria*. In addition, all the selected mt genomes included 3 rRNA genes (Table 2). Combing our present results, it revealed that *B. rapa* var. *Purpuraria* has a high degree of similarity to the mt genome sequences to *B. rapa*, *B. nigra*, and *B. oleracea*. Furthermore, to identify sequence variations in the known genes, SNPs were detected between *B. rapa* var. *Purpuraria* and *B. juncea*, *B. napus*, *B. rapa*, *B. nigra* and *B. oleracea*. A total of 202 SNPs were found among six mt genomes. 135 SNPs were identified between *B. rapa* var. *Purpuraria* and *B. rapa* ssp. *rapa*, followed by 62 SNPs in *B. rapa* var. *Purpuraria* vs *B. nigra* group, 4 SNPs (*cox2*, *atp1*, and two *cox1* genes) in *B. rapa* var. *Purpuraria* vs *B. napus* group, and only one SNP (*atp1* gene) in *B. rapa* var. *Purpuraria* vs *B. oleracea* group. But there were no SNPs identified in *B. rapa* var. *Purpuraria* vs *B. juncea* and *B. rapa* var. *Purpuraria* vs *B. rapa* (Table S4).

### Codon preference analysis of PCGs

The total size of PCGs in *B. rapa* var. *Purpuraria* was 35,927 bp. Except for *nad1* gene with ACG as the start codon, ATG was the start codon for other PCGs, which might be the result of C-to-U RNA editing of the second site (Table 1). Four types of stop codons, including TAA, TGA, TAG, and CGA, were detected, and the C to U for RNA editing phenomenon was discovered in *ccmFc* gene. We also calculated the RSCU of 33 PCGs in the *B. rapa* var. *Purpuraria* mt genome (Fig. 2). The 33 PCGs made up 29,055 bp encoding 9685 codons including termination codons. Leucine (Leu) was the most frequent amino acid, with a total of 1053 codons, accounted for 10.87%, followed by serine (Ser), with a total of 856 codons, accounting for 8.84%, and termination codon (Ter) was the rarest with a total of 33 codons, accounting for 0.034%. We discovered that 30 codons of RSCU value were greater than 1, of which 27 codons (90%) ended with A or U, two codons (6.67%) ended with G, and only

one codon (3.33%) ended with C. It illustrated that the A / U preference at the third codon was positioned in the *B. rapa* var. *Purpuraria* mt genome (Table S5).

### The prediction of RNA editing

RNA-editing sites are widely distributed in the mt genome of plants. In this study, 379 RNA editing sites within 33 PCGs (Table 3) were predicted in the mt genome of *B. rapa* var. *Purpuraria* using PmtREP tool (Figure S1). Among these PCGs, *atp6* had one RNA-editing site, whereas the highest was in *nad4* with 379 RNA-editing sites (33), of which 30.34% (115 sites) occurred with the first position of the triplet codes, 68.07% (258 sites) located at the second base of the triplet codes. In addition, the first and second bases of the triplet codes were edited, leading to an amino acid change from proline (CCC) to phenylalanine (TTC). There were 45.64% (173 positions) of amino acids hydrophobicity remained unchanged after the RNA editing. Besides, 45.38% (172 positions) of the amino acids were varied from hydrophilic to hydrophobic, while 8.71% (33 positions) were ranged from hydrophobic to hydrophilic. Furthermore, only one amino acid was varied from arginine to stop codon (Table 3). The findings in our study showed that most amino acids were changed from serine to leucine (24.01%, 160 sites), proline to leucine (22.69%, 86 sites), and serine to phenylalanine (12.40%, 47 sites).

In addition, we compared the RNA editing sites of *B. rapa* (AP017996), *B. nigra*(AP012989), and *B. oleracea* (AP012988) with representatives from *Brassica* species (Fig. 3). The highest edited transcripts were *ccmB* with 32 editing sites in *B. nigra*, and *nad4* with 33 editing sites in *B. rapa*, *B. oleracea* and *B. rapa* var. *Purpuraria*. From the comparison results of RNA editing sites, we found that *B. rapa* var. *Purpuraria* is highly similar to other three closely related *Brassica* species.

### Repeat sequences analysis

Simple sequence repeats (SSRs), also known as microsatellites, are DNA stretches composing of short unit sequence repeats of 1–6 base pairs in length [43]. In this study, a total of 55 SSRs were detected in the mt genome of *B. rapa* var. *Purpuraria*, containing 20 (36.36%) monomers, 11 (20.00%) dimers, 5 (9.09%) trimers, 18 (32.73%) tetramers, and 1 (1.82%) pentamers (Table 4). No hexanucleotide repeats were detected. Among the 55 SSRs, monomer and tetramer were the main type of SSR motifs, accounting for 69.09% of all detected SSRs. In addition, 90.00% of monomers had A/T contents, and 36.36% of dimers were AT/TA (Table S6). The abundant AT content of SSRs supported with the high AT content (54.77%) of the whole mt genome of *B. rapa* var. *Purpuraria*.

**Table 1** Gene profile and organization of the *B. rapa* var. *Purpuraria* mt genome

Group of genes	Gene name	Length	Start codon	Stop codon	Amino acid
ATP synthase	<i>atp1</i>	1524	ATG	TAG	508
	<i>atp4</i>	579	ATG	TAA	193
	<i>atp6</i>	786	ATG	TAA	262
	<i>atp8</i>	477	ATG	TGA	159
	<i>atp9</i>	225	ATG	TGA	75
Cytochrome c biogenesis	<i>ccmB</i>	621	ATG	TAA	207
	<i>ccmC</i>	744	ATG	TGA	248
	<i>ccmFc<sup>a</sup></i>	1329	ATG	CGA(TGA)	443
	<i>ccmFn</i>	1146	ATG	TGA	382
Ubichinol cytochrome c reductase	<i>cob</i>	1182	ATG	TGA	394
Cytochrome c oxidase	<i>cox1</i>	1584	ATG	TAA	528
	<i>cox2<sup>a</sup></i>	783	ATG	TAA	261
	<i>cox3</i>	798	ATG	TGA	266
Maturases	<i>matR</i>	1974	ATG	TAG	658
Transport membrane protein	<i>mttB</i>	360	ATG	TAG	120
NADH dehydrogenase	<i>nad1****</i>	978	ACG(ATG)	TAA	326
	<i>nad2****</i>	1467	ATG	TAA	489
	<i>nad3</i>	357	ATG	TAA	119
	<i>nad4****</i>	1488	ATG	TGA	496
	<i>nad4L</i>	303	ATG	TAA	101
	<i>nad5****</i>	2010	ATG	TAA	670
	<i>nad6</i>	618	ATG	TAA	206
	<i>nad7****</i>	1185	ATG	TAG	395
Ribosomal proteins (LSU)	<i>rpl10</i>	225	ATG	TAG	75
	<i>rpl16</i>	249	ATG	TAA	83
	<i>rpl2<sup>a</sup></i>	1050	ATG	TGA	350
	<i>rpl5</i>	558	ATG	TAA	186
Ribosomal proteins (SSU)	<i>rps12</i>	378	ATG	TGA	126
	<i>rps14</i>	303	ATG	TAG	101
	<i>rps3<sup>a</sup></i>	1665	ATG	TAG	555
	<i>rps4</i>	1089	ATG	TAA	363
	<i>rps7</i>	447	ATG	TAA	149
Succinate dehydrogenase					
Ribosomal RNAs	<i>rrn18</i>	1847			
	<i>rrn26</i>	3176			
	<i>rrn5</i>	121			

**Table 1** (continued)

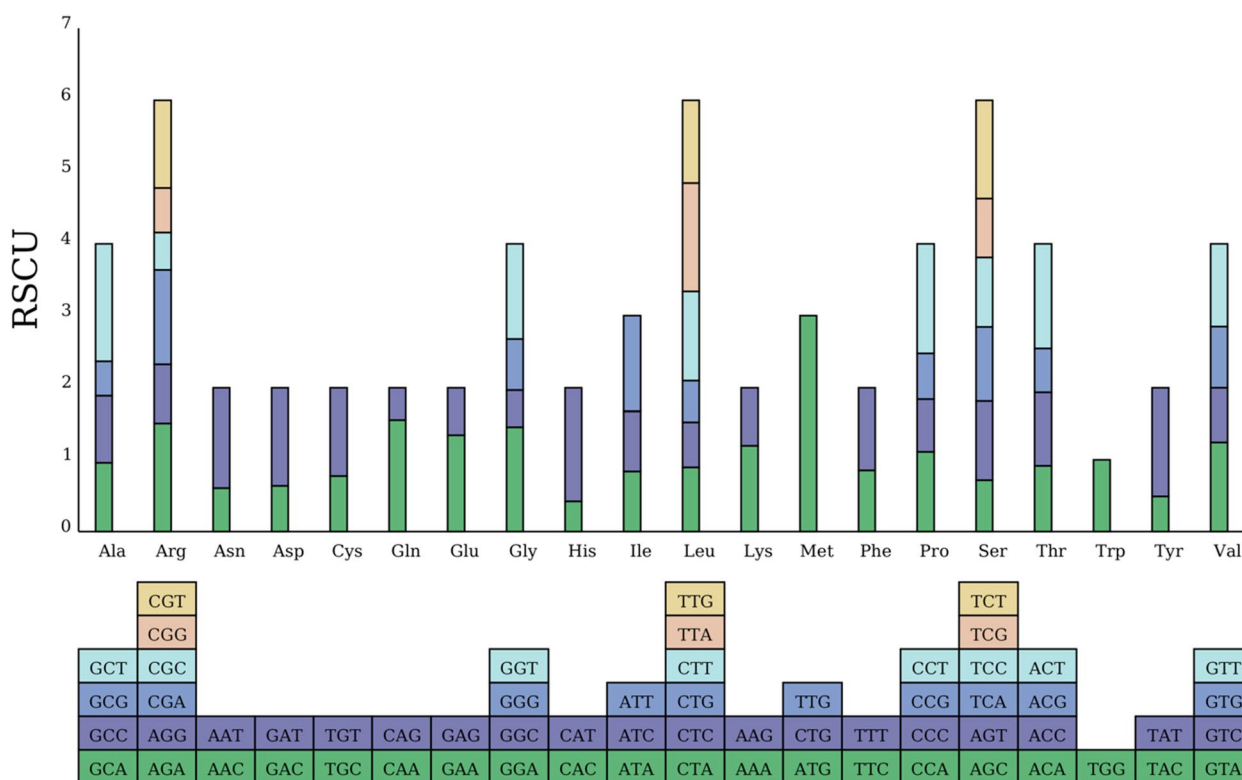
Group of genes	Gene name	Length	Start codon	Stop codon	Amino acid
Transfer RNAs	<i>trnC-GCA</i>	71			
	<i>trnD-GTC</i>	74			
	<i>trnE-TTC</i>	72			
	<i>trnG-GCC</i>	72			
	<i>trnH-GTG</i>	74			
	<i>trnH-GTG</i>	74			
	<i>trnI-AAT<sup>a</sup></i>	69			
	<i>trnK-TTT</i>	73			
	<i>trnL-CAA</i>	85			
	<i>trnM-CAT</i>	73			
	<i>trnM-CAT</i>	74			
	<i>trnM-CAT</i>	74			
	<i>trnN-GTT</i>	72			
	<i>trnP-TGG</i>	75			
	<i>trnQ-TTG</i>	72			
	<i>trnS-GCT</i>	88			
	<i>trnS-GGA</i>	87			
	<i>trnS-TGA</i>	87			
	<i>trnT-GGT<sup>a</sup></i>	69			
	<i>trnW-CCA</i>	74			
<i>trnY-GTA</i>	83				
<i>trnY-GTA</i>	68				
<i>trnY-GTA</i>	68				

<sup>a</sup> Intron number**Table 2** Comparison of gene content among *Brassica* mt genomes

GenBank Accession number	Species	Genome size (bp)	Genes	PCGs	tRNA	rRNA
PP231953	<i>B. rapa</i> var. <i>Purpuraria</i>	219,775	59	33	23	3
JF920288	<i>B. juncea</i>	219,766	99	78	18	3
AP006444.1	<i>B. napus</i>	221,853	106	79	17	3
AP017996	<i>B. rapa</i>	219,775	55	34	18	3
AP012989	<i>B. nigra</i>	232,145	53	33	17	3
AP012988	<i>B. oleracea</i>	219,952	54	34	17	3
MT409179	<i>B. rapa</i> ssp. <i>rapa</i>	219,736	99	78	18	3

Tandem repeats (satellite DNA) are core repeating units about 1–200 bases [44]. As shown in Table 5, 17 tandem repeats with a matching degree over than  $\geq 84\%$  and length varying from 3 to 39 bp were obtained. A total of 252 dispersed repeats ( $\geq 28$  bp), of which 144

palindromic (57.54%) and 108 forward repeats (42.46%) were observed, and no reverse and complementary repeats were found (Fig. 4). The total length of the dispersed repeats was 16251 bp, which occupied 7.39% of the whole mt genome. Most repeats were 25–40 bp (169



**Fig. 2** RSCU analysis of the *B. rapa* var. *Purpuraria* mt genome

repeats, 67.06%), while only one was over than 1 kb being 2427 bp (Table S7).

**Ka/Ks and Pi analysis**

The Ka/Ks ratio was used to evaluate selective pressures during the evolutionary dynamics of PCGs among similar species. In this work, *B. rapa* var. *Purpuraria* was used as a reference to calculate the Ka/Ks value of 33 PCGs in the *B. rapa* var. *Purpuraria* mt genome (Fig. 5). The Ka/Ks values of most of PCGs were less than one, demonstrating that these genes may undergo negative selections during evolution. However, the Ka/Ks value of the *atp4* and *ccmB* genes between *B. rapa* var. *Purpuraria* and *Cucurbita pepo*, the *ccmB* gene between *B. rapa* var. *Purpuraria* and *Glycyrrhiza uralensis*, the *atp4*, *ccmB*, and *mttB* genes between *B. rapa* var. *Purpuraria* and *Helianthus grosseserratus*, the *ccmB* gene between *B. rapa* var. *Purpuraria* and *Solanum lycopersicum* were higher than one, implying positive selection for these genes during evolution. Our findings further indicated that these mt genes might be highly conserved during the evolution process in higher plants.

The nucleotide diversity (Pi) values of 36 PCGs were accounted and varied from 0.01790 to 0.14222, with an average of 0.04417 (Fig. 6 and Table S8). The Pi value

of *gene3.rpl10* region was largest among these regions being 0.1422, and 0.07436 in *gene4.rps4*, 0.07195 in *gene6.atp8*, 0.07182 in *gene29.rpl2* and 0.0709 in *gene21.atp9* were found. The lower Pi values revealed that the mt genome sequences of *B. rapa* var. *Purpuraria* were highly conserved.

**Phylogenetic analysis**

To determine the evolutionary status for the mt genome of *B. rapa* var. *Purpuraria*, the phylogenetic analyses was carried out on *B. rapa* var. *Purpuraria* together with other 24 Cruciferae/Brassica species (Fig. 7). A phylogenetic tree was built based on an aligned data matrix of 31 conserved PCGs, including *atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFc*, *ccmFn*, *cob*, *cox1*, *cox2*, *cox3*, *mttB*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*, *rpl16*, *rpl2*, *rpl5*, *rps12*, *rps14*, *rps3*, *rps4*, and *rps7*, from all tested species. The phylogenetic tree was divided into six groups, being *Brassica*, *Raphanus*, *Arabis*, *Arabidopsis*, *Capsella*, and *Ginkgo*. *B. rapa* var. *Purpuraria* was clustered with the species of genus *Brassica* at first group, and formed sister branches with other related *Brassica* species within the Cruciferae family clade. Furthermore, *B. rapa* var. *Purpuraria* was closely related to *B. rapa* subsp. *Oleifera* (NC\_016125.1) and *B. juncea*



**Table 3** Prediction of RNA editing sites in the *B. rapa* var. *Purpuraria* mt genome

Type	RNA-editing	Number	Percentage
hydrophilic-hydrophilic	CAC (H) = >TAC (Y)	6	12.66%
	CAT (H) = >TAT (Y)	16	
	CGC (R) = >TGC (C)	6	
	CGT (R) = >TGT (C)	20	
hydrophilic-hydrophobic	ACA (T) = >ATA (I)	2	45.38%
	ACC (T) = >ATC (I)	1	
	ACG (T) = >ATG (M)	6	
	ACT (T) = >ATT (I)	4	
	CGG (R) = >TGG (W)	21	
	TCA (S) = >TTA (L)	53	
	TCC (S) = >TTC (F)	16	
	TCG (S) = >TTG (L)	38	
	TCT (S) = >TTT (F)	31	
	hydrophilic-stop	CGA (R) = >TGA (X)	
hydrophobic-hydrophilic	CCA (P) = >TCA (S)	6	8.71%
	CCC (P) = >TCC (S)	5	
	CCG (P) = >TCG (S)	5	
	CCT (P) = >TCT (S)	17	
hydrophobic-hydrophobic	CCA (P) = >CTA (L)	34	32.98%
	CCC (P) = >CTC (L)	7	
	CCC (P) = >TTC (F)	6	
	CCG (P) = >CTG (L)	21	
	CCT (P) = >CTT (L)	24	
	CCT (P) = >TTT (F)	10	
	CTC (L) = >TTC (F)	4	
	CTT (L) = >TTT (F)	8	
	GCA (A) = >GTA (V)	2	
	GCC (A) = >GTC (V)	4	
	GCG (A) = >GTG (V)	4	
	GCT (A) = >GTT (V)	1	
	Total		

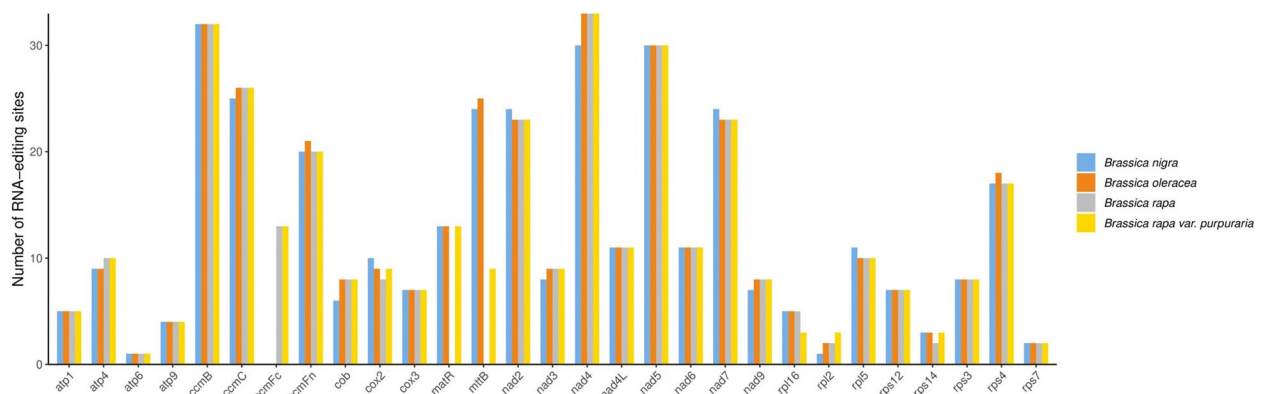
(NC\_016123.1), indicating that *B. rapa* var. *Purpuraria* belongs to the *Brassica* in the Cruciferae family.

### Analysis of homologous fragments of mitochondria and chloroplasts

The total length of homologous sequences on chloroplasts was 13,153 bp, accounting for 8.57% of the whole cp genome. While the total size of homologous sequences on mitochondria was 8961 bp, accounting for 4.08% of the whole mt genome (Table S9). As shown in Table 6, twenty-two homologous fragments with a total length of 13,325 bp were found. The transfer route of the fragments may occur firstly from the chloroplast to nucleus, and then to the mitochondrion in *B. rapa* var. *Purpuraria*, accounting for 6.06% of the whole mt genome. Eight annotated genes, namely, *trnL-CAA*, *trnN-GTT*, *rrn18*, *trnW-CCA*, *trnD-GTC*, *trnM-CAT*, *ccmC*, and *trnI-AAT*, with high similarity to the mitochondria likely originated from the mt genome. While 17 genes, being *rpoB*, *ycf2*, *ycf15*, *trnL-CAA*, *rbcL*, *trnN-GUU*, *ycf1*, *psaA*, *rrn23*, *rrn16*, *psaB*, *trnW-CCA*, *trnD-GUC*, *trnP-UGG*, *trnM-CAU*, *trnI-CAU*, and *trnI-GAU* with high similarity to the cp genes, might be transferred from cp genome, and only partial sequences of those genes were identified in the mt genome (Table 6). Most of transferred genes were tRNA genes, of which those genes were much more conserved in the mt genome than PCGs during the evolution.

### Discussion

Mitochondria is the core of energy source in cells, which exhibited more complex in plant than animals due to its size variations and repetitive sequences [45–47]. In this study, we analyzed the characteristics of mt genome in *B. rapa* var. *Purpuraria*. The total length of the mt genome of *B. rapa* var. *Purpuraria* was similar to that of *B. juncea* [48], being moderate in genome length compared with other reported mt genomes [49]. The GC content in the



**Fig. 3** The distribution and comparison of RNA-editing sites in the PCGs of *B. rapa* var. *Purpuraria* mt genome and three closely related *Brassica* mt genomes

**Table 4** Frequency of identified SSR motifs in the *B. rapa* var. *Purpuraria* mt genome

Motif Type	Number of repeats																Total	Total Proportion (%)
	3	4	5	6	7	8	9	10	11	12	13	14	15	16				
Monomer	-	-	-	-	-	-	-	14	2	-	1	1	-	2	20	36.36		
Dimer	-	-	10	1	-	-	-	-	-	-	-	-	-	-	11	20.00		
Trimer	-	4	-	-	-	1	-	-	-	-	-	-	-	-	5	9.09		
Tetramer	17	1	-	-	-	-	-	-	-	-	-	-	-	-	18	32.73		
Pentamer	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1.82		
Total	18	5	10	1	0	1	0	14	2	0	1	1	0	2	55	100.00		

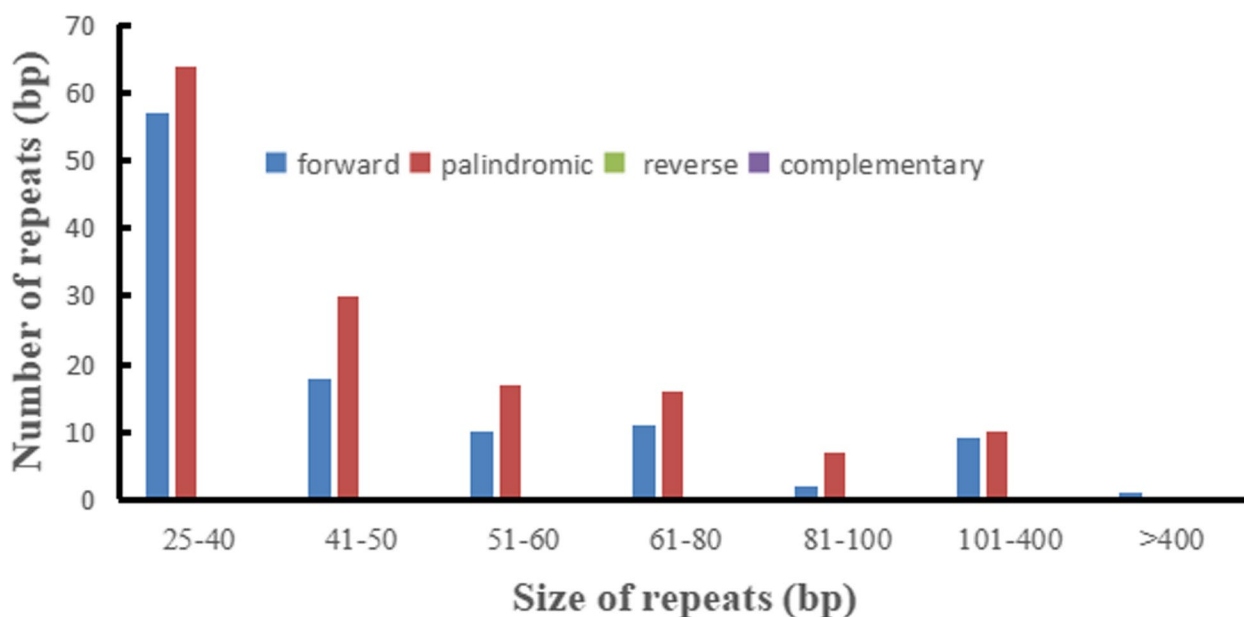
**Table 5** The tandem repeats analysis of *B. rapa* var. *Purpuraria* mt genome

NO	Chr	Size	Copy	Repeat sequence	Percent Matches	Start	End
1	chr1	3	8.3	CTT	100	4183	4207
2	chr1	29	3.1	AAAGGAGAGGTGCTTTAGCAACTCGACTG	98	28,136	28,225
3	chr1	27	2.7	GCTTTCTTGGTTTGATGAGCTTATACAC	95	49,104	49,177
4	chr1	12	2.1	TGAACTGATAGC	100	61,927	61,951
5	chr1	22	2.2	TCGAGATCTTTGAACCTTTCAG	96	69,117	69,165
6	chr1	36	2.4	AGAAGAAGGTTGACTCGGCAATCTCAATTCGTATG	94	71,537	71,623
7	chr1	20	1.9	ATGTTAGTGTTTCAGTATATC	100	71,739	71,776
8	chr1	12	2.2	CTCGAGGAACGC	100	88,939	88,965
9	chr1	14	2.1	CGGAGGCGGGTAAG	100	116,459	116,488
10	chr1	19	2.1	AGATTTTACAAATGGTCTG	100	118,977	119,015
11	chr1	39	2.8	TATCAATTCATAAGAGAAGAAAGATCGTTTTTTTAAAT	100	119,157	119,267
12	chr1	17	2.3	ATATATCCATTCTCATA	95	138,230	138,268
13	chr1	18	2.4	TTCCATCAAGAAGGTACC	84	144,510	144,551
14	chr1	15	2	TAGAAAACCTGGCATC	100	147,080	147,109
15	chr1	34	2	AAGGAGGAACCCAGCTTATCCCCTCTCAGAGGAG	88	186,199	186,266
16	chr1	15	2.1	TCTTTTCTTGCTCTT	94	191,598	191,629
17	chr1	21	2.7	AACAGATAACAACAGCATATT	100	205,173	205,229

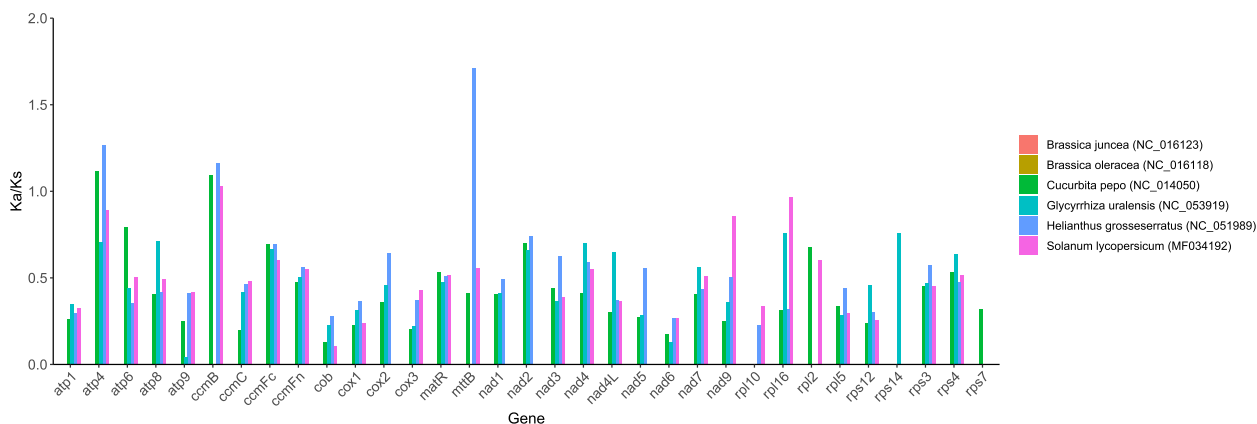
*B. rapa* var. *Purpuraria* mt genome is 45.23%, which is similar to that of other reported plant mt genomes such as *Camellia sinensis* var. *Assamica* cv. *Duntsa*, 45.62% [50]; *Mesona chinensis* Benth, 44.21% [51]; *B. napus*, 45.21% [52], while exhibited greater than the *B. rapa* var. *Purpuraria* cp genome (PP191173, 36.36%) assembled by our research group. Non-coding sequence accounted for 71.7% of the whole *B. rapa* var. *Purpuraria* mt genome, which is similar to most of plant mt genomes [52–54]. In addition, the PCGs occupied 13.22% might be due to the result of increasing sequence repeats during evolution. PCGs usually encoded from initiation codon (ATG) to stop codons (TGA, TAA, and TGA), and the distribution of amino acids compositions was consistent with *Acer yangbiense* [55] and *A. thaliana* [56]. The *nad1* gene using ACG as start codon in consistent with *Salix suchowensis* and *Phaseolus vulgaris* might be induced by RNA editing [46, 47].

Codon usage bias refers that synonymous codons exist in a non-random manner in different species [57]. The analysis of codon usage patterns is helpful to understand the molecular mechanism of biological adaptation and explore the evolutionary relationship among species [58]. Previous studies have shown that codons prefer to use A / U endings in the plant mt genomes [44–47]. In total of 30 high-frequency codons were detected in the *B. rapa* var. *Purpuraria* mt genome, of which 90% codons ended with A or U, which might be the result of natural selection, mutation pressure and genetic drift [59]. In addition, we found that leucine was the most frequently used amino acid, which was consistent with *S. glauca* [44] and *Acer truncatum* Bunge [60].

The number of RNA editing sites varies among different plants, and occurred commonly in gymnosperm and angiosperm mt genomes. We obtained 379 RNA editing sites within all the 33 PCGs in the *B. rapa* var. *Purpuraria*



**Fig. 4** Dispersed repeats analysis in the *B. rapa* var. *Purpuraria* mt genome

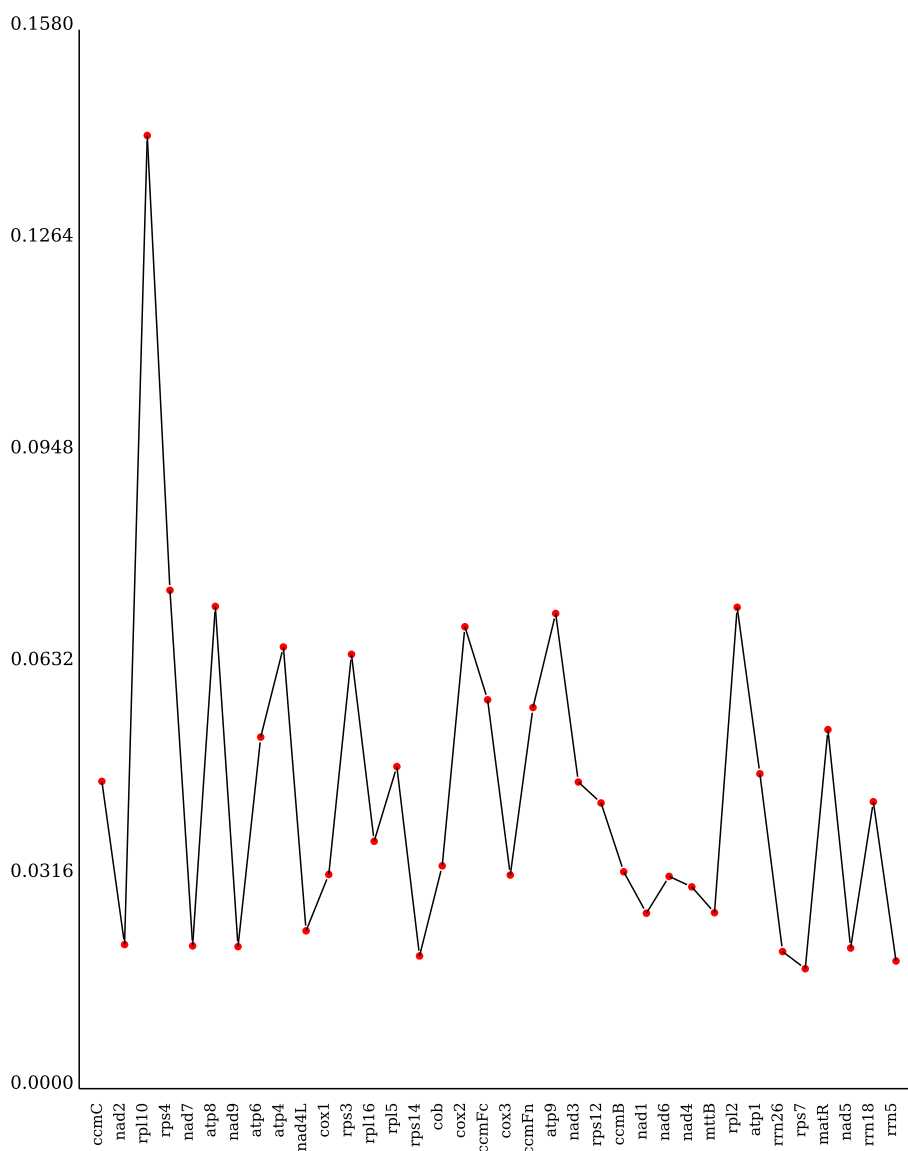


**Fig. 5** Ka/Ks ratios of 33 PCGs between *B. rapa* var. *Purpuraria* and six species

mt genome, which exhibited much less than those in *Diospyros oleifera* (515) [61], *Bupleurum chinense* DC (517) [38], *Macadamia integrifolia* (688), *M. ternifolia* (689) and *M. tetraphylla* (688) [62], and higher than those in *S. glauca* (261)[44] and *Pereskia aculeata* (362) [63]. The selection of RNA editing sites in the *B. rapa* var. *Purpuraria* genome exhibited a high degree of compositional bias. Most of the RNA editing sites were the C-T editing type, being similar as in other plant mt genomes [64–66]. Previous studies showed that about 50% of RNA editing generated at the second bases of the triplet codes [44, 65]. About 68.07% (258) RNA editing sites occurred at the second codon position in the *B. rapa* var. *Purpuraria*

mt genome, greater than that at the first codon position (115, 30.34%). In addition, 1.58% (6) RNA editing sites occurred at both of the first and second codon position. The similar phenomenon was observed in the *D. oleifera* mt genome [61]. There is no RNA editing sites predicted at the third codon position in *B. rapa* var. *Purpuraria* mt genome [44, 66].

The repeat sequences, including SSR, tandem and dispersed repeats, were widely distributed in the plant mt genomes [8, 67]. Previous studies have reported that repeat sequences were important for intermolecular recombination, which play a vital role in forming the mt genome [68]. Because of its high variability and recessive

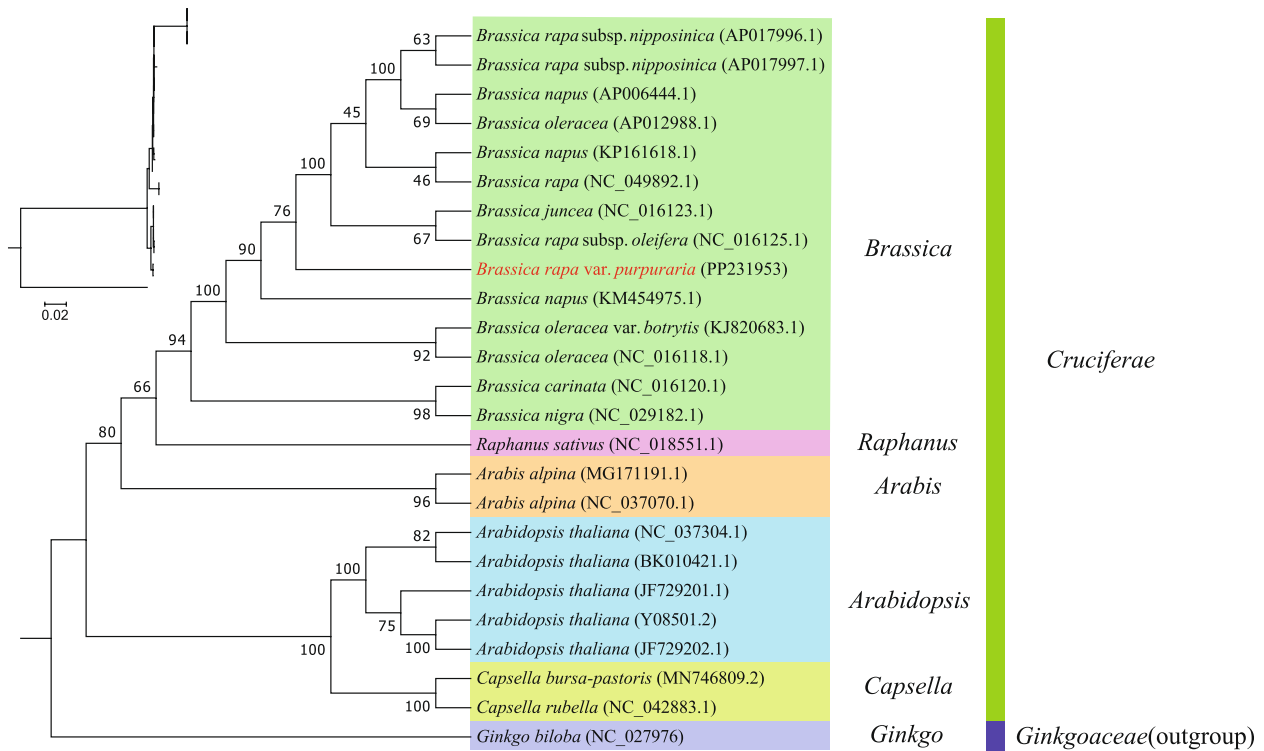


**Fig. 6** Nucleotide diversity of *B. rapa* var. *Purpuraria* mt genome

inheritance, SSR has been widely used to confirm phylogenetic relationships, genetic diversity studies and species identification [69]. The mt genome of *B. rapa* var. *Purpuraria* contained 55 SSRs, of which 90.00% of monomers being A or T, resulting 54.77% of AT in the *B. rapa* var. *Purpuraria* mt genome. The high AT content in mt genome were also detected in *Scutellaria tsinyunensis* [70]. and *Magnolia biondii* [71]. In addition, 252 dispersed repeats were discovered in this study, which was much greater than *B. oleracea* var. *Italica* in the genus *Brassica* [72].

The ratio of Ka/Ks provides useful information for reconstructing phylogenetic relationships, and

contributes to understand the evolutionary dynamics of PCGs among closely related species [73]. Most of mt genes with Ka/Ks ratios < 1 exhibited negative selections, and a few genes with Ka/Ks ratios > 1 showed positive selections during the evolution in plants [44, 61]. The Ka/Ks ratio of *ccmB* gene was greater than 1 in *S. glauca* mt genome exhibited positive selection [44], whereas Five genes, namely, *atp4*, *nad1*, *ccmC*, *mttB*, and *rpl2*, showed positive selections in *Capsicum pubescens* Ruiz & Pav mt genome [74]. Two genes, being *mttB* and *rpl5* exhibited positive selections in European-Asian species [75]. However, three genes with Ka/Ks ratios > 1, including *atp4*, *ccmB*, and *mttB*, exhibited positive selections



**Fig. 7** Maximum-likelihood phylogenetic tree based on 31 conserved PCGs among 25 species. *Ginkgo biloba* (NC\_027976) used as the out group

**Table 6** Fragments transferred from cp to mt in *B.rapa* var. *Purpuraria*

NO	Length	Identity%	Mismatches	Gap	Cp start	Cp end	Mt start	Mt end	Gene (cp)	Gene (mt)
1	2186	98.262	33	1	23,007	25,192	108,053	110,233	<i>rpoB<sup>a</sup></i>	
2	1879	97.339	34	6	91,634	93,500	33,492	31,618	<i>ycf2<sup>a</sup>, ycf15, trnL-CAA</i>	<i>trnL-CAA</i>
3	1879	97.339	34	6	143,266	145,132	31,618	33,492	<i>trnL-CAA, ycf15, ycf2<sup>a</sup></i>	<i>trnL-CAA</i>
4	1365	96.41	39	3	53,575	54,934	142,589	141,230	<i>rbcl<sup>a</sup></i>	
5	1159	99.655	4	0	108,069	109,227	63,319	62,161	<i>trnN-GUU, ycf1<sup>a</sup></i>	<i>trnN-GTT</i>
6	1159	99.655	4	0	127,539	128,697	62,161	63,319	<i>ycf1<sup>a</sup>, trnN-GUU</i>	<i>trnN-GTT</i>
7	659	96.055	20	3	40,170	40,826	38,184	37,530	<i>psaA<sup>a</sup></i>	
8	229	96.943	6	1	103,753	103,980	89,675	89,447	<i>rrn23<sup>a</sup></i>	
9	229	96.943	6	1	132,786	133,013	89,447	89,675	<i>rrn23<sup>a</sup></i>	
10	878	74.374	170	43	135,617	136,469	210,650	209,803	<i>rrn16<sup>a</sup></i>	<i>rrn18<sup>a</sup></i>
11	878	74.374	170	43	100,297	101,149	209,803	210,650	<i>rrn16<sup>a</sup></i>	<i>rrn18<sup>a</sup></i>
12	91	98.901	1	0	38,693	38,783	48,932	49,022	<i>psaB<sup>a</sup>, psaA<sup>a</sup></i>	
13	143	86.713	11	8	65,123	65,264	34,481	34,346	<i>trnW-CCA</i>	<i>trnW-CCA</i>
14	74	97.297	2	0	29,508	29,581	71,853	71,780	<i>trnD-GUC</i>	<i>trnD-GTC</i>
15	118	85.593	17	0	65,457	65,574	34,130	34,013	<i>trnP-UGG<sup>a</sup></i>	
16	77	92.208	6	0	50,951	51,027	52,691	52,615	<i>trnM-CAU</i>	<i>trnM-CAT</i>
17	43	100	0	0	100,387	100,429	117,447	117,405	<i>rrn16<sup>a</sup></i>	
18	43	100	0	0	136,337	136,379	117,405	117,447	<i>rrn16<sup>a</sup></i>	
19	76	85.526	7	3	85,434	85,508	5573	5501	<i>trnI-CAU</i>	<i>ccmC<sup>a</sup>, trnI-AAT</i>
20	76	85.526	7	3	151,258	151,332	5501	5573	<i>trnI-CAU</i>	<i>ccmC<sup>a</sup>, trnI-AAT</i>
21	42	97.619	1	0	134,348	134,389	17,596	17,555	<i>trnI-GAU<sup>a</sup></i>	
22	42	97.619	1	0	102,377	102,418	17,555	17,596	<i>trnI-GAU<sup>a</sup></i>	

<sup>a</sup> Represents the partial sequence identified in mt genome

in our study is consistent with previous studies [44, 74, 75], illustrating that these genes might be selected for future researches on the gene selection and phylogenetic of *Brassica* species. Changes in the size and structure of the plant mt genomes have been obviously observed, whereas the functional genes are still conserved [76]. Previous studies indicated that Pi could reveal the variation of nucleic acid sequences in different plants, and the highly variable regions might be selected as potential molecular markers for population genetics [77, 78]. Pi analysis reflected the variation of nucleotide sequences among different species (Fig. 6). Our results revealed that the Pi value of *rpl10* gene was the largest in these regions, illustrating that *rpl10* gene might be used as molecular markers for the mt genome analysis in *B. rapa* var. *purpuraria*. Except for *Reclinomonas*, plants are the only group of eukaryotes that still remain the *rpl10* gene in their mt genomes [48, 49]. However, the mt *rpl10* gene has been missed in some Brassicaceae species, and replaced by an additional copy of the nuclear gene that normally encodes cp RPL10 protein [79]. Five highly variable regions, being *ccmB*, *ccmFC*, *rps1*, *rps10*, and *rps14*, might be used as molecular markers in *Phaseolus vulgaris* mt genome. In addition, the overall low Pi values showed that the mt genome sequences of *B. rapa* var. *Purpuraria* were highly conserved.

Phylogenetic analysis of plants have developed to use the complete genome data to construct the relationship among different species [44–48]. Here, the phylogenetic tree was constructed according to the mt genomes of 25 species. *B. rapa* var. *Purpuraria* was well clustered with the species of genus *Brassica* and stayed closely to *B. rapa* subsp. *Oleifera* and *B. juncea*, suggesting that *B. rapa* var. *Purpuraria* belongs to the *Brassica* species in the Cruciferae family. Shao et al. (2021) found that *B. oleracea* stayed a closely relationship with *B. rapa* subsp. *Campes- tris* with 100% support rate [80]. Brassicaceae is a superfamily containing over 3800 species, in which *Brassica* is the most important genus as having many important vegetables and oil crops. It has been mentioned that based on cp genome, *B. napus* always clustered with *B. rapa* morphotypes, but did not cluster into a monophyletic group, and were distantly separated by *B. juncea* and *B. oleracea*. The obtained phylogenetic tree revealed a clear phylogenetic relationship among the species. *B. rapa* var. *Purpuraria* as a local vegetable planted around Yangtze River, may also develop some different characteristics in mt genome to regulate the biosynthesis in anthocyanin, and other nutritional compounds. Therefore, assembling and analyzing the mt genomes with those difference genome information will help to better understand their genetic characteristics and selecting their differences for further investigation. DNA sequence transfer between cp

and mt genomes has been frequently discovered in plant mt genomes [81]. In higher plants, the total size of transferred DNA ranges from 50 kb (*A. thaliana*) to 1.1 Mb (*O. sativa* subsp. *japonica*) depending on the plant species [82]. In total, 13,325 bp of cp DNA has been transferred into the mt genome of *B. rapa* var. *Purpuraria*, accounting for 6.06% of the *B. rapa* var. *Purpuraria* mt genome. In comparison, the proportion in *M. integrifolia*, *Liriodendron tulipifera*, and *Nicotiana tabacum* is 5.4%, 3%, and 2.5%, respectively [38, 71, 83]. About 22 fragments transferred from the cp genome to the mt genome, containing eight annotated genes, with six tRNA genes (*trnL-CAA*, *trnN-GTT*, *trnW-CCA*, *trnD-GTC*, *trnM-CAT*, and *trnI-AAT*), *rrn18*, and *ccmC*. The tRNA genes transferred from cp to mt genomes has been commonly discovered in angiosperms [44, 84, 85]. These results were consistent with the previous study, which showed much more conserved for tRNA genes than the PCGs during the evolution, and tRNA genes played indispensable roles in mt genome [46].

## Conclusions

In this work, we sequenced and successfully assembled the genome with a circular molecule structure in the mt genome of *B. rapa* var. *Purpuraria*. The full length of mt genome for *B. rapa* var. *Purpuraria* was 219,775 bp, containing 59 genes with 33 PCGs, 23 tRNA and 3 rRNA genes. The codon preferences, repeat sequences, RNA editing sites in the *B. rapa* var. *Purpuraria* mt genome have also been analyzed subsequently. Phylogenetic analysis confirmed that *B. rapa* var. *Purpuraria* exhibited a close relationship with *B. oleracea*. Gene conservation between mt and cp genome was also discovered in *B. rapa* var. *Purpuraria* via analyzing gene migration. The Ka/Ks analysis showed that most of the PCGs exhibited negative selections, demonstrating the conservation of these mt genes during the process of evolution. This study provides useful genetic information about the *B. rapa* var. *Purpuraria* mt genome and forms an important theoretical might also help to analyze the genetic variation, systematic evolution, and breeding of *B. rapa* var. *Purpuraria*.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10457-1>.

Supplementary Material 1.

Supplementary Material 2.

## Authors' contributions

Yihui Gong and Hua Huang conceived and designed the project. Xin Xie and Peng Li collected the plant materials. Xin Xie, Guihua Zhou, Meiyu Chen, and Zhiyin Chen performed the experiments and analyzed the data. Yihui

Gong and Hua Huang wrote and revised the manuscript. All authors read and approved the final manuscript.

### Funding

This work was supported by Hunan Natural Science Regional Joint Fund Project (grant numbers 2024JJ7235).

### Availability of data and materials

The *B. rapa* var. *Purpuraria* mt genome sequence was uploaded in the GenBank database (accession number PP231953).

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

Received: 2 February 2024 Accepted: 26 May 2024

Published online: 01 June 2024

### References

- Zhang X, Zhang K, Wu J, et al. QTL-Seq and sequence assembly rapidly mapped the gene *BrMYBL2.1* for the purple trait in *Brassica rapa*. *Sci Rep*. 2020;10(1):2328.
- Wang Y, Di H, Cheng W, et al. Variation in the main health-promoting compounds and antioxidant activity of different edible parts of purple flowering stalks (*Brassica campestris* var. *purpuraria*) and green flowering stalks (*Brassica campestris* var. *campestris*). *Plants*. 2022;11(13):1664.
- Guo P, Zhang B, Hu Z, et al. Anthocyanin accumulation and transcriptional regulation in purple flowering stalk (*Brassica campestris* L. var. *purpurea* Bailey). *Plant Molec Biol*. 2023;111(1–2):57–72.
- Millar AH, Heazlewood JL, Kristensen BK, et al. The plant mitochondrial proteome[J]. *Trends Plant Sci*. 2005;10:36–43.
- Gualberto JM, Milesina D, Wallet C, et al. The plant mitochondrial genome: dynamics and maintenance[J]. *Biochimie*. 2014;100:107–20.
- Maréchal A, Brisson N. Recombination and the maintenance of plant organelle genome stability[J]. *New Phytol*. 2010;186(2):299–317.
- Unsel M, Marienfeld JR, Brandt P, et al. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides[J]. *Nat Genet*. 1997;15(1):57–61.
- Notsu Y, Masood S, Nishikawa T, et al. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants[J]. *Mol Genet Genomics*. 2002;268(4):434–45.
- Handa H. The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. *Nucleic Acids Res*. 2003;31(20):5907–16.
- Clifton SW, Minx P, Fauron CMR, et al. Sequence and comparative analysis of the maize NB mitochondrial genome[J]. *Plant Physiol*. 2004;136(3):3486–503.
- Goremykin VV, Salamini F, Velasco R, et al. Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer[J]. *Mol Biol Evol*. 2009;26(1):99–110.
- Goremykin VV, Lockhart PJ, Viola R, et al. The mitochondrial genome of *Malus domestica* and the import-driven hypothesis of mitochondrial genome expansion in seed plants[J]. *Plant J*. 2012;71(4):615–26.
- Wang S, Li D, Yao X, et al. Evolution and diversification of kiwifruit mitochondrial genomes through extensive whole-genome rearrangement and mosaic loss of intergenic sequences in a highly variable region[J]. *Genome Biol Evol*. 2019;11(4):1192–206.
- Kan SL, Shen TT, Gong P, et al. The complete mitochondrial genome of *Taxus cuspidata* (Taxaceae): eight protein-coding genes have transferred to the nuclear genome[J]. *BMC Evol Biol*. 2020;20:1–17.
- Wu Y, Zhou H. Research progress of sugarcane chloroplast genome[J]. *Agricultural Science & Technology*. 2013;14(12):1693.
- Janouškovec J, Liu SL, Martone PT, et al. Evolution of red algal plastid genomes: ancient architectures, introns, horizontal gene transfer, and taxonomic utility of plastid markers[J]. *PLoS ONE*. 2013;8(3):e59001.
- Davila JI, Arrieta-Montiel MP, Wamboldt Y, et al. Double-strand break repair processes drive evolution of the mitochondrial genome in *Arabidopsis*[J]. *BMC Biol*. 2011;9(1):1–14.
- Skippington E, Barkman TJ, Rice DW, et al. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes[J]. *Proc Natl Acad Sci*. 2015;112(27):E3515–24.
- Sloan DB, Alverson AJ, Chackalovcak JP, et al. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates[J]. *PLoS Biol*. 2012;10(1):e1001241.
- Putintseva YA, Bondar EI, Simonov EP, et al. Siberian larch (*Larix sibirica* Ledeb) mitochondrial genome assembled using both short and long nucleotide sequence reads is currently the largest known mitogenome [J]. *BMC Genomics*. 2020;21:654.
- Gualberto JM, Newton KJ. Plant mitochondrial genomes: dynamics and mechanisms of mutation[J]. *Annu Rev Plant Biol*. 2017;68:225–52.
- Koizuka N, Imai R, Fujimoto H, et al. Genetic characterization of a pentatricopeptide repeat protein gene, orf687, that restores fertility in the cytoplasmic male-sterile Kosenia radish[J]. *Plant J*. 2003;34(4):407–15.
- Bhatnagar-Mathur P, Gupta R, Reddy PS, et al. A novel mitochondrial orf147 causes cytoplasmic male sterility in pigeonpea by modulating aberrant anther dehiscence[J]. *Plant Mol Biol*. 2018;97:131–47.
- Xie H, Peng X, Qian M, et al. The chimeric mitochondrial gene orf182 causes non-pollen-type abortion in Dongxiang cytoplasmic male-sterile rice[J]. *Plant J*. 2018;95(4):715–26.
- Alverson AJ, Wei XX, Rice DW, et al. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae)[J]. *Mol Biol Evol*. 2010;27(6):1436–48.
- Jongsun P, Hong XI. The complete mitochondrial genome of *Arabidopsis thaliana* (Brassicaceae) isolated in Korea[J]. *Korean J Plant Taxon*. 2021;51(2):176–80.
- Tanaka Y, Tsuda M, Yasumoto K, et al. A complete mitochondrial genome sequence of Ogura-type male-sterile cytoplasm and its comparative analysis with that of normal cytoplasm in radish (*Raphanus sativus* L.)[J]. *BMC Genom*. 2012;13:352. <https://doi.org/10.1186/1471-2164-13-352>.
- Wang J, Jiang J, Li X, et al. Complete sequence of heterogenous-composition mitochondrial genome (*Brassica napus*) and its exogenous source[J]. *BMC Genomics*. 2012;13(1):1–9.
- Hatono S, Nishimura K, Murakami Y, et al. Complete mitochondrial genome sequences of *Brassica rapa* (Chinese cabbage and mizuna), and intraspecific differentiation of cytoplasm in *B. rapa* and *Brassica juncea*[J]. *Breed Sci*. 2017;67(4):357–62.
- Xue JY, Wang Y, Dong S, et al. Maternal inheritance of U's triangle and evolutionary process of *Brassica* mitochondrial genomes[J]. *Front Plant Sci*. 2020;11:540342.
- Ren Y. The complete mitochondrial genome of turnip (*Brassica rapa* ssp. *rapa*)[J]. *Mitochondrial DNA Part B*. 2021;6(4):1566–1567.
- Arseneau JR, Steeves R, Laflamme M. Modified low-salt CTAB extraction of high-quality DNA from contaminant-rich tissues[J]. *Mol Ecol Resour*. 2017;17(4):686–93.
- Li H. Minimap2: pairwise alignment for nucleotide sequences[J]. *Bioinformatics*. 2018;34(18):3094–100.
- Koren S, Walenz BP, Berlin K, et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation[J]. *Genome Res*. 2017;27(5):722–36.
- Chan P P, Lowe T M. tRNAscan-SE: searching for tRNA genes in genomic sequences[M]. Springer N Y. 2019;1962:1–14.
- Thiel T, Michalek W, Varshney R, et al. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.)[J]. *Theoret Appl Genet*. 2003;106(3):411–22.
- Mower JP. PREP-Mt: predictive RNA editor for plant mitochondrial genes[J]. *BMC Bioinformatics*. 2005;6:1–15.

38. Qiao Y, Zhang X, Li Z, et al. Assembly and comparative analysis of the complete mitochondrial genome of *Bupleurum chinense* DC[J]. *BMC Genomics*. 2022;23(1):1–17.
39. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses[J]. *Bioinformatics*. 2009;25(15):1972–3.
40. Kalyaanamoorthy S, Minh BQ, Wong TKF, et al. ModelFinder: fast model selection for accurate phylogenetic estimates[J]. *Nat Methods*. 2017;14(6):587–9.
41. Nguyen LT, Schmidt HA, Von Haeseler A, et al. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies[J]. *Mol Biol Evol*. 2015;32(1):268–74.
42. Alverson AJ, Zhuo S, Rice DW, et al. The mitochondrial genome of the legume *Vigna radiata* and the analysis of recombination across short mitochondrial repeats[J]. *PLoS ONE*. 2011;6(1): e16404.
43. Xu Y, Cheng W, Xiong C, et al. Genetic diversity and association analysis among germplasms of *Diospyros kaki* in Zhejiang Province based on SSR markers[J]. *Forest*. 2021;12(4):422.
44. Cheng Y, He X, Priyadarshani S, et al. Assembly and comparative analysis of the complete mitochondrial genome of *Suaeda glauca*[J]. *BMC Genomics*. 2021;22(1):1–15.
45. Kozik A, Rowan BA, Lavelle D, et al. The alternative reality of plant mitochondrial DNA: One ring does not rule them all[J]. *PLoS Genet*. 2019;15(8): e1008373.
46. Ye N, Wang X, Li J, et al. Assembly and comparative analysis of complete mitochondrial genome sequence of an economic plant *Salix suchowensis*[J]. *PeerJ*. 2017;5: e3148.
47. Bi C, Lu N, Xu Y, et al. Characterization and analysis of the mitochondrial genome of common bean (*Phaseolus vulgaris*) by comparative genomic approaches[J]. *Int J Mol Sci*. 2020;21(11):3778.
48. Wu Z, Hu K, Yan M, et al. Mitochondrial genome and transcriptome analysis of five alloplasmic male-sterile lines in *Brassica juncea*[J]. *BMC Genomics*. 2019;20:1–15.
49. Chang S, Yang T, Du T, et al. Mitochondrial genome sequencing helps show the evolutionary mechanism of mitochondrial genome formation in Brassica[J]. *BMC Genomics*. 2011;12:1–12.
50. Li J, Tang H, Luo H, et al. Complete mitochondrial genome assembly and comparison of *Camellia sinensis* var. *Assamica* cv. *Duntsa*[J]. *Front Plant Sci*. 2023;14:1117002. <https://doi.org/10.3389/fpls.2023.1117002>.
51. Tang D, Huang S, Quan C, et al. Mitochondrial genome characteristics and phylogenetic analysis of the medicinal and edible plant *Mesona chinensis* Benth[J]. *Front Genet*. 2023;13:1056389.
52. Jiang M, Ni Y, Li J, et al. Characterisation of the complete mitochondrial genome of *Taraxacum mongolicum* revealed five repeat-mediated recombinations[J]. *Plant Cell Rep*. 2023;42(4):775–89.
53. Yang JX, Dierckxsens N, Bai MZ, et al. Multichromosomal mitochondrial genome of *Paphiopedilum micranthum*: Compact and fragmented genome, and rampant intracellular gene transfer[J]. *Int J Mol Sci*. 2023;24(4):3976.
54. Liu D, Qu K, Yuan Y, et al. Complete sequence and comparative analysis of the mitochondrial genome of the rare and endangered *Clematis acerifolia*, the first clematis mitogenome to provide new insights into the phylogenetic evolutionary status of the genus[J]. *Front Genet*. 2023;13:1050040.
55. Yang J, Wariss HM, Tao L, et al. De novo genome assembly of the endangered *Acer yangbiense*, a plant species with extremely small populations endemic to Yunnan Province, China[J]. *Gigasci*. 2019;8(7):giz085.
56. Sloan DB, Wu Z, Sharbrough J. Correction of persistent errors in *Arabidopsis* reference mitochondrial genomes[J]. *Plant Cell*. 2018;30(3):525–7.
57. Li C, Zhou L, Nie J, et al. Codon usage bias and genetic diversity in chloroplast genomes of *Elaeagnus* species (Myrtiflorae: Elaeagnaceae)[J]. *Physiol Mol Biol Plants*. 2023;29(2):239–51.
58. Ding H, Gao J, Yang J, et al. Genome evolution of *Buchnera aphidicola* (Gammaproteobacteria): Insights into strand compositional asymmetry, codon usage bias, and phylogenetic implications[J]. *Int J Biol Macromol*. 2023;253: 126738.
59. Bulmer M. The selection-mutation-drift theory of synonymous codon usage[J]. *Genetics*. 1991;129(3):897–907.
60. Ma Q, Wang Y, Li S, et al. Assembly and comparative analysis of the first complete mitochondrial genome of *Acer truncatum* Bunge: A woody oil-tree species producing nervonic acid[J]. *BMC Plant Biol*. 2022;22(1):1–17.
61. Xu Y, Dong Y, Cheng W, et al. Characterization and phylogenetic analysis of the complete mitochondrial genome sequence of *Diospyros oleifera*, the first representative from the family Ebenaceae[J]. *Heliyon*. 2022;8(7):e09870. <https://doi.org/10.1016/j.heliyon.2022.e09870>.
62. Niu Y, Lu Y, Song W, et al. Assembly and comparative analysis of the complete mitochondrial genome of three *Macadamia* species (*M. integrifolia*, *M. ternifolia* and *M. tetraphylla*)[J]. *Plos One*. 2022;17(5):e0263545.
63. Zhang X, Shan Y, Li J, et al. Assembly of the complete mitochondrial genome of *Pereskia aculeata* revealed that two pairs of repetitive elements mediated the recombination of the genome[J]. *Int J Mol Sci*. 2023;24(9):8366.
64. Edera AA, Sanchez-Puerta MV. Computational detection of plant RNA editing events[M]. *Methods Mol Biol*. 2021;2181:13–34. <https://doi.org/10.1007/978-1-0716-0787-9>.
65. Robles P, Quesada V. Organelle genetics in plants[J]. *Int J Mol Sci*. 2021;22(4):2104.
66. Verhage L. Targeted editing of the *Arabidopsis* mitochondrial genome[J]. *Plant J*. 2020;104(6):1457–8.
67. Guo W, Zhu A, Fan W, et al. Complete mitochondrial genomes from the ferns *Ophioglossum californicum* and *Psilotum nudum* are highly repetitive with the largest organellar introns[J]. *New Phytol*. 2017;213(1):391–403.
68. Dong S, Zhao C, Chen F, et al. The complete mitochondrial genome of the early flowering plant *Nymphaea colorata* is highly repetitive with low recombination[J]. *BMC Genomics*. 2018;19(1):1–12.
69. Ping J, Feng P, Li J, et al. Molecular evolution and SSRs analysis based on the chloroplast genome of *Callitropsis funebris*[J]. *Ecol Evol*. 2021;11(9):4786–802.
70. Li J, Xu Y, Shan Y, et al. Assembly of the complete mitochondrial genome of an endemic plant, *Scutellaria tsinyunensis*, revealed the existence of two conformations generated by a repeat-mediated recombination[J]. *Planta*. 2021;254:1–16.
71. Dong S, Chen L, Liu Y, et al. The draft mitochondrial genome of *Magnolia biondii* and mitochondrial phylogenomics of angiosperms[J]. *PLoS ONE*. 2020;15(4): e0231020.
72. Zhang Z, Tao M, Shan X, et al. Characterization of the complete chloroplast genome of *Brassica oleracea* var. *italica* and phylogenetic relationships in Brassicaceae[J]. *Plos One*. 2022;17(2):e0263310.
73. Fay JC, Wu CI. Sequence divergence, functional constraint, and selection in protein evolution[J]. *Annu Rev Genomics Hum Genet*. 2003;4(1):213–35.
74. Xu L, Wang J, Zhang T, et al. Characterizing complete mitochondrial genome of *Aquilegia amurensis* and its evolutionary implications[J]. *BMC Plant Biol*. 2024;24(1):142.
75. Wu D, Fu W, Fan G, et al. Characteristics and comparative analysis of the special-structure (non-single-circle) mitochondrial genome of *Capsicum pubescens* Ruiz & Pav[J]. *Genes*. 2024;15(2):152.
76. Hong Z, Liao X, Ye Y, et al. A complete mitochondrial genome for fragrant Chinese rosewood (*Dalbergia odorifera*, Fabaceae) with comparative analyses of genome structure and intergenomic sequence transfers[J]. *BMC Genomics*. 2021;22(1):1–13.
77. Feng L, Wang Z, Wang C, et al. Multichromosomal mitochondrial genome of *Punica granatum*: comparative evolutionary analysis and gene transformation from chloroplast genomes[J]. *BMC Plant Biol*. 2023;23(1):512.
78. Bi S, Song Y, Liu L, et al. Complete mitochondrial genome of *Piophilidae* (Diptera: Piophilidae): Genome description and phylogenetic implications[J]. *Genes*. 2023;14(4):883.
79. Kubo N, Arimura SI. Discovery of the *rpl10* gene in diverse plant mitochondrial genomes and its probable replacement by the nuclear gene for chloroplast RPL10 in two lineages of angiosperms[J]. *DNA Res*. 2010;17(1):1–9.
80. Shao D, Ma Y, Li X, et al. The sequence structure and phylogenetic analysis by complete mitochondrial genome of kohlrabi (*Brassica oleracea* var. *gongylodes* L.)[J]. *Mitochondrial DNA Part B*. 2021;6(9):2714–6.
81. Straub SCK, Cronn RC, Edwards C, et al. Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (Apocynaceae)[J]. *Genome Biology and evolution*. 2013;5(10):1872–85.
82. Smith DR, Crosby K, Lee RW. Correlation between nuclear plastid DNA abundance and plastid number supports the limited transfer window hypothesis[J]. *Genome Biol Evol*. 2011;3:365–71.



83. Sugiyama Y, Watase Y, Nagase M, et al. The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants[J]. *Mol Genet Genomics*. 2005;272:603–15.
84. Bergthorsson U, Adams KL, Thomason B, et al. Widespread horizontal transfer of mitochondrial genes in flowering plants[J]. *Nature*. 2003;424(6945):197–201.
85. Bi C, Paterson A H, Wang X, et al. Analysis of the complete mitochondrial genome sequence of the diploid cotton *Gossypium raimondii* by comparative genomics approaches[J]. *BioMed Res Int*. 2016;2016:5040598.

### **Publisher's Note**

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