scientific reports

Check for updates

OPEN

Assembly and comparative analysis of the complete mitochondrial genome of *Lactuca sativa* var. *ramosa* Hort

Yihui Gong[1]✉, Yalin Qin[1], Rong Liu[1], Yuanyuan Wei[1], Haotian Liu[1], Peng Li[2] & Guihua Zhou[1]✉

Lettuce (*Lactuca sativa* var. *ramosa* Hort) is an important leaf vegetable that widely cultivates due to its high-quality, short growth cycle, and less diseases. *L. sativa* var. *ramosa* Hort belongs to Asteraceae family and its evolutionary relationships with related species of Asteraceae are not completely assessed based on genome sequences. Here, we assembled the whole mitochondrial (mt) genome of *L. sativa* var. *ramosa* Hort, and performed a comparative with other related species. The *L. sativa* var. *ramosa* Hort mt genome has a typical circular structure with a length of 363,324 bp, within GC content accounted for 45.35%. In total of 71 genes, comprising 35 protein-coding genes (PCGs), 6 rRNAs, 28 tRNAs, and 2 pseudogenes were annotated. Codon preference, RNA-editing sites, repetitive sequences, and genes migrating from chloroplast (cp) to mt genomes were investigated in the *L. sativa* var. *ramosa* Hort mt genome. Nucleotide diversity (Pi) showed that the *L. sativa* var. *ramosa* Hort mt genome was relatively conserved. A Bayesian phylogenetic tree showed that *L. sativa* var. *ramosa* Hort was closely to *L. sativa* var. *capitata* and *L. virosa*, which belonged to the *Lactuca* genus in the Asteraceae family. Our findings will provide useful information to explore genetic variation, genetic diversity, and molecular breeding on the *Lactuca* genus.

Keywords *Lactuca sativa* var. *ramosa* Hort, Mitochondrial genome, Repeats, Phylogenetic analysis

Mitochondria are double-membrane semi-autonomous organelles that widely present in eukaryotes, and are the place of cell oxidative metabolism[1]. Mitochondria are involved in the regulation of cell growth, division, apoptosis, and the synthesis and metabolism of some compounds, which also play a vital role in the development of plants[2]. The mitochondrial (mt) genomes of most plants have a circular double stranded DNA, and their lengths rang from several thousand to several million base pairs[3,4]. In the current study, the mt genome of *Brassica napus* has the lowest length of 221 kb, while the *Silene conoidea* mt genome has the largest size of 11.3 Mb[5]. Although plant mitochondria display great diversity in terms of genome size, most of the protein-coding genes (PCGs) are highly conserved, mainly composed of 24 core conserved genes and 17 variant genes, and could be divided into complex I (*nad*), complex II (*sdh*), complex III (*cob*), complex IV (*cox*), complex V (*atp*), Cytohrome c biogenesis (*ccm*), and transfer RNAs, etc.[6]. Except for complex II, ribosomal protein, and tRNA genes, other genes are relatively conserved in the mt genome of higher plants[7,8]. Unlike chloroplast (cp) genomes that use their own unique genetic codes, the genetic codes among plant mt genomes are universal across species. In addition to directly inheriting from ancestral mitochondria, tRNA genes also originate from the migration of their own cp genome sequences[9,10]. There are also abundant genetic variations in the mt genomes of higher plants, which are widely used as potential molecular markers for studying the origin and evolution of species and population genetic diversity[11]. The mt genome not only has the characteristics of fast evolution speed and low recombination rate, but also has the advantages of small genome size and easy sequencing research. It has become an ideal tool for comparative genetics and systematics researches among different plants[12,13].

Lettuce is an important raw vegetable in the world that belongs to the Asteraceae family in the *Lactuca* genus, which originates from Mediterranean coast[14]. Lettuce is one of the main cultivated vegetables in plant factories, and is favored by consumers due to its abundant vitamin C content[15]. *Lactuca* is a big genus including about 100 species, which is divided into four groups: the cultivated one, being *L. sativa*, and three wild species,

[1]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. [2]Xiangtan Agricultural Science Research Institute, Xiangtan 411100, China. ✉email: gyhzgh@163.com; zghgyh@163.com

being *L. serriola*, *L.virosa*, and *L. saligna*. *L. serriola* has the phenotypic characteristics of harder prickles on its stem and lobed leaves[16]. While *L. saligna* has a narrow and long leaf phenotype. *L.virosa* represents many phenotypes, including some with lobed leaves or not, some with prickles on its leaves others not, but all species have wide leaves[16]. *L. sativa* var. *ramosa* Hort with long obovate leaves and dense into cabbage-like leaf balls is eaten raw, crisp and refreshing, and slightly sweet. Morphological differences could distinguish between *L. sativa*, the cultivated lettuce and wild-type varieties. Four types of lettuces, being Butterhead lettuces (var. *capitata* L. nidus tenerrima), Crisphead lettuces (var. *capitata* L. salinas), leaf lettuces (var. *acephala* Alef.), and Cos lettuces (var. *longifolia*), have strong competitive advantages in the market[17]. Besides, *L. sativa* var. *ramosa* Hort is a representative lettuce variety, which widely planted in China due to its short growth cycle and high nutritional value.

The process leading to the domestication of *L. sativa* is still unclear. *L. serriola* was confirmed as the direct ancestor and one of the closest related species of *L. sativa*[18–20]. With the progress of sequencing technologies and the report of sequencing genomes, it is helpful to explain the relationship between *Lactuca* spp. The *rnL-F* and *ndhF* genes were developed as cp marker for *Lactuca* species[21]. The *Lactuca* spp is classified into several clades, namely, *L. sativa*, *L. serriola*, *L. virosa*, and *L. saligna*[21]. In addition, the mt genomes of several *Lactuca* species have also been reported. The ones of *L. sativa* (363,324 bp, MK642355), *L. serriola* (363,328 bp, MK820672), and *L. saligna* (368,269 bp, MK759657), which makes it possible to develop a new set of markers[22]. The characterized mt genome of *L. sativa* var. *ramosa* comparison with between wild and cultivated *Lactuca* species can contribute to finding genetic or structural variations in the evolutionary history of *Lactuca* cultivars. Therefore, the assembly and analysis of the mt genome is of great significance for better understanding its genetic features and for molecular marker research.

In this study, we sequenced and assembled the whole mtDNA sequence of *L. sativa* var. *ramosa* by using the Illumina and Nanopore sequencing platforms and described its genome features. Its genome characteristics and evolutionary relationships were conducted a comparison with the other related *Lactuca* species. The findings obtained in this study provide available genetic information to explore species identification, genetic variation, and genetic relationship for the *Lactuca* species in the future.

Materials and methods

Plant materials, genome sequencing and assembly

The *L. sativa* var. *ramosa* Hort plants were cultivated in the greenhouse at the the Loudi Ziyuan Agricultural Science and Technology Development Co., Ltd. (Tongzi Village, Shanshan Town, Louxing District, Loudi, Hunan, China, 27°47′N, 112°1′E). We collected approximately 5 g of 30-day-old leaves, transported them using dry ice, and sent them to the Genepioneer Biotechnologies (Nanjing, China). Total genomic DNA of *L. sativa* var. *ramosa* Hort was isolated from the young leaves using the HiPure Universal DNA kit D301(Genepioneer Biotechnologies). The DNA purity was detected with 1.0% agarose gel, and then was sequenced using the Illumina Novaseq6000 and Oxford Nanopore PromethION sequencing platforms. To obtain high-quality reads of *L. sativa* var. *ramosa* Hort mt genome, Fastp v0.23.4 (https://github.com/OpenGene/fastp) software was used to filter the Illumina sequencing raw data, and delete the sequencing adaptors and primer sequences in the reads, filter out reads with an average quality value lower than Q5, and discard reads with the number over than 5. Then, the Nanopore sequencing raw data was filtered via using filtlongv0.2.1 software and the parameters were set as follows:–min_length 1000 and –min_mean_q 7. The Nanopore sequencing raw data was assembled via using Minimap2[23], of which mt sequences were aligned with the plant mt gene database (https://github.com/xul962464/plant_mt_ref_gene) . Sequences with sizes > 50 bp, comprising multiple core genes, were screened as the seed sequences according to their alignment. Subsequently, Minimap2 was used to compare the original Nanopore sequencing raw data with the seed sequences, and sequences with overlap > 1 kb were selected and added to the seed sequences, and iteratively aligned the original Nanopore sequencing data with the seed sequences to obtain all the mt genome sequence of *L. sativa* var. *ramosa* Hort. All the Nanopore seuqencing data were conducted self-correction via using Canu[24], and Bowtie2 (v2.3.5.1) was used to compare the Illumina sequencing data to the corrected sequence. The corrected Illumina sequencing data were stitched with the corrected Nanopore sequencing data using Unicycler (v0.4.8) with default parameters. The stitching results were visualized and manually adjusted using Bandage software (v0.8.1), and finally mt genome sequence of *L. sativa* var. *ramosa* Hort was obtained.

Genome annotation

The PCGs and rRNA genes of the *L. sativa* var. *ramosa* Hort mt genome was annotated using MIFOFY[5]. Then, the tRNA genes was analyzed using tRNAscan-SE 2.0[25]. Finally, the annotation results were manually adjusted and corrected based on the related species. Open Reading Frame Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) was used to identify the ORFs with the length ≥ 102 bp, and delete the redundant sequences and known genes with overlap sequences. OGDRAW program was used to draw the circle map of *L. sativa* var. *ramosa* Hort mt genome[26].

Repeat sequence analysis

Interspersed repeats, comprising forward repeats, palindromic repeats, reverse repeats, and complementary repeats, were identified using blastn v2.10.1 with removing redundancy and tandem repeats, and the parameters was set as follows : -word_size 7 and evalue 1e-5. Subsequently, the interspersed repeats were visualized using circos v0.69-5. Tandem repeats were analyzed using online tool Tandem Repeats Finder (http://tandem.bu.edu/trf/trf.basic.submit.html) with parameters set as default. Misa v1.0 software was used to detect simple sequence repeats (SSRs). The repeats of one to six bases with 10, 5, 4, 3, 3, and 3 repeats numbers, respectively, were analyzed in this analysis.

RNA-editing analysis in PCGs and Pi analysis

The RNA- editing sites of 31 PCGs of *L. sativa* var. *ramosa* Hort and other five mt genomes (*L. saligna*, *L. sativa*, *L. sativa* var. *capitata*, *L. serriola*, and *L. virosa*) were identified using the PREP-Mt online tool (http://prep.unl.edu/) with cutoff value set as 0.2[27]. We calculated the nucleotide diversity (Pi) value of each PCG between *L. sativa* var. *ramosa* Hort and *L. saligna*, *L. sativa*, *L. sativa* var. *capitata*, *L. serriola*, and *L. virosa*. The homologous gene sequences from six *Lactuca* species were globally aligned using mafft software v7.427 with auto mode. The Pi value of each PCG was determined using Dnasp5.

Phylogenetic analyses

A total of 28 entire mt genomes, including 27 representative Asteraceae species and one Ginkgoaceae species, were used to confirm the phylogenetic position of *L. sativa* var. *ramosa* Hort. The 31 mt PCGs, being *atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFc*, *ccmFn*, *cob*, *cox1*, *cox2*, *cox3*, *matR*, *mttB*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*, *rpl10*, *rpl16*, *rpl5*, *rps12*, *rps13*, *rps3*, and *rps4,* conserved across the 28 tested species were aligned in MAFFT v7.427 with –auto mode. The aligned sequences were connected end-to-end, and were trimmed using trimAl (v1.4.rev15) in ModelFinder[28,29]. A Bayesian phylogenetic tree was created using MrBayes v3.2.7 software with the Markov Chain Monte Carlo (MCMC) iterative operation for 1 million generations, sampling every 100 generations. The initial 25% of the phylogenetic tree was deleted (burn-in), and then the majority-rule consensus tree was obtained.

Identification of homologous fragments from cp genome to mt genome

To obtain the homologous fragments from cp genome to mt genome, BLASTN software was used to compare the *L. sativa* var. *ramosa* Hort mt genome with its cp genome (PP999684). The parameters were set as follows: the matching rate ≥ 70%, E-value ≤ 1e − 5 and the minimum length = 30 bp[30].

Synteny analysis

Using *L. sativa* var. *ramosa* Hort as the reference, genome alignment between other *Lactuca* sequences and *L. sativa* var. *ramosa* Hort sequences was conducted using nucmer (4.0.0beta2) software with the maxmatch parameter to produced dot-plot plots,

Results

Features of the *L. sativa* var. *ramosa* Hort mt genome.

The *L. sativa* var. *ramosa* Hort mt genome was generated 16,089,057,852 raw data and 53,275,026 bp clean data (Q20 = 98.71% and Q30 = 96.34%) were obtained via the Illumina sequencing (Table S1). Then, in total of 17,652,114,316 bases and 1,711,468 reads were obtained via Nanopore sequencing with a mean read size of 10,314 bp. Te subreads with N50 value was 24,620 bp in length (Table S2). The *L. sativa* var. *ramosa* Hort mt genome exhibited a typical circular structure with full length of 363,324 bp (Fig. 1). The nucleotide composition of the entire *L. sativa* var. *ramosa* Hort mt genome included 27.33% for A, 27.31% for T, 22.65% for C, and 22.70% for G, with GC content of 45.35% (Table S3). PCGs and cis-spliced introns accontted for 9.46% and 6.42% of the entire mt genome, while tRNA and rRNA genes only occupied 0.57% and 3.12%, respectively. 71 annotated genes, consisting of 35 PCGs, 6 rRNAs, 28 tRNAs, and 2 pseudogenes, were detected in the *L. sativa* var. *ramosa* Hort mt genome (Table 1). Six genes, being *ccmFc*, *cox2*, *nad4*, *rps3*, *trnS-GCT*(2), and *trnT-TGT*(3), had one intron; whereas four genes, namely, *nad1*, *nad2*, *nad5*(2), and *nad7*, included four introns. 11 genes, including *atp1*, *ccmB*, *nad5*, *rpl10*, *rrn18*, *rrn26*, *rrn5*, *trnD-GTC*, *trnK-TTT*, *trnQ-TTG*, and *trnS-GCT*, were found in two copies, while *trnT-TGT* and *trnM-CAT* genes were detected in three or five copies.

Genome size and gene content vary from species to species[31,32]. Five representative *Lactuca* species were used to compare genome features and find variability of the genome of *L. sativa* var. *ramosa* Hort (Table 2). The lengths of all the tested species were between 363,324 bp (*L. sativa* var. *ramosa* Hort, *L. sativa*, and *L. sativa* var. *capitata*) and 373,019 bp (*L. virosa*). The lowest number of genes (69) were identified in *L. sativa* and *L. serriola*, and the highest (79) in *L. virosa*. The PCGs ranged from 35 in *L. sativa* var. *ramosa* Hort to 43 in *L. virosa*, and tRNAs were between 25 and 29. Excep for *L. sativa* var. *ramosa* Hort, all the *Lactuca* species had the same number in rRNA (6) and intron (24). The AT and GC contents exhibited a minor difference in all the detected species. Overall, *L. sativa* var. *ramosa* showed a minor difference in characteristics with other *Lactuca* species.

Codon usage analysis of PCGs

Except for *cox1* gene with ACG and *mttB* gene with ATT as the start codon, other PCGs were used ATG as the start codon, which resulted in C-to-U RNA editing of the second site and G-to-U RNA editing of the third site, respectively (Table 1). The RSCU values of 35 PCGs were calculated with our Perl script in the *L. sativa* var. *ramosa* Hort mt genome (Fig. 2). Except for stop codons, the 35 PCGs encoded 9,868 codons with the total length of 34,353 bp. The highest frequent amino acid was leucine (Leu), encoded by CUA, CUC, CUG, CUU, UUA, and UUG, with 1,051 codons, followed by serine (Ser), encoded by AGC, AGU, UCA, UCC, UCG and UCU, with 936 codons, and cysteine (Cys) encoded by UGC and UGU was the lowest with 134 codons. 29 codons with RSCU > 1 were observed in the *L. sativa* var. *ramosa* Hort mt genome, of which 27 codons (93.10%) ended with A or U, and two condons (6.90%) ended with C or G. In addition, the methionine (Met) and tryptophan (Trp) with RSCU = 1 showed no preference (Table S4).

Prediction of RNA-editing sites

RNA-editing is a means of maintaining the normal biological function of cp and mt, and widely exists in all eukaryotes[33]. In this work, 500 RNA-editing sites in 35 PCGs (Table 3) were discovered in the *L. sativa* var. *ramosa* Hort mt genome using the PREP-Mt online tool. The *atp8* gene had the least RNA-editing sites (3), while

Fig. 1. Circular map of the *L. sativa* var. *ramosa* Hort mt genome.

the largest was in *ccmFn* gene with 37 RNA-editing sites (Figure S1). Among 500 RNA-editing sites, 64.80% (324 sites) changed at the second position of the triplet codes, followed 33.60% (168 sites) changed with the first base of the triplet codes. while 1.6% (8 sites) changed with the first and second bases of the triplet codes, which resulted in an amino acid change from proline (CCC) to phenylalanine (TTC). Additionally, 48% (240) sites were changed from hydrophilic to hydrophobic, followed 31.4% (157) from hydrophobic to hydrophobic, and 0.40% (2 sites) was the least from hydrophilic to stop. Furthermore, 113 sites (about 22.6%) were varied from serine (S) to leucine (L) , and 110 sites (about 22%) were change from proline (P) to leucine (L).

Furthermore, we compared the RNA editing sites of *L. saligna*, *L. sativa*, *L.* sativa var. *capitata*, *L. serriola* and *L.virosa* with representatives from *Lactuca* species (Fig. 3). The largest edited transcripts were *ccmB* and *ccmFn* both with 36 RNA editing sites in *L. saligna*, and *L . sativa* var. *capitata*, and the *ccmFn* gene with 38–39 RNA editing sites (38 for *L. sativa* var. *ramosa* Hort, *L. serriola*, and *L.virosa*; 39 for *L.* sativa). From the comparison of RNA editing sites among six *Lactuca* species, we found that they have no interspecies differences in the number of RNA editing sites for *ccmB*.

Group of genes	Gene name	Length	Start codon	Stop codon	Amino acid
ATP synthase	*atp1*	1,947	ATG	TAA	649
	atp1	1,533	ATG	TAA	511
	atp4	576	ATG	TAA	192
	atp6	720	ATG	TAA	240
	atp8	480	ATG	TAA	160
	atp9	261	ATG	TAA	87
Cytohrome c biogenesis	*ccmB*	621	ATG	TGA	207
	ccmB	621	ATG	TGA	207
	ccmC	753	ATG	TGA	251
	*ccmFc**	1,347	ATG	TAA	449
	ccmFn	1,719	ATG	TAG	573
Ubichinol cytochrome c reductase	*cob*	1,176	ATG	TGA	392
Cytochrome c oxidase	*cox1*	1,584	ACG(ATG)	TAA	528
	*cox2**	834	ATG	TAG	278
	cox3	798	ATG	TGA	266
Maturases	*matR*	1,968	ATG	TAG	656
Transport membrance protein	*mttB*	882	ATT	TAG	294
NADH dehydrogenase	*nad1****	978	ATG	TAA	326
	*nad2****	1,467	ATG	TAA	489
	nad3	357	ATG	TAA	119
	*nad4**	1,488	ATG	TGA	496
	nad4L	273	ATG	TAA	91
	*nad5****	2,013	ATG	TAA	671
	*nad5****	2,013	ATG	TAA	671
	nad6	732	ATG	TAG	244
	*nad7****	1,185	ATG	TAG	395
	nad9	573	ATG	TAA	191
Ribosomal proteins (LSU)	*rpl10*	489	ATG	TAA	163
	rpl10	489	ATG	TAA	163
	rpl16	516	ATG	TAA	172
	rpl5	564	ATG	TAA	188
Ribosomal proteins (SSU)	*rps12*	372	ATG	TAG	124
	rps13	351	ATG	TGA	117
	*rps3**	1,677	ATG	TAG	559
	rps4	996	ATG	TAG	332
	#rps14				
Succinate dehydrogenase	*#sdh4*				
Ribosomal RNAs	*rrn18*	1,947			
	rrn18	1,947			
	rrn26	3,612			
	rrn26	3,612			
	rrn5	117			
	rrn5	117			
Transfer RNAs	*trnC-GCA*	71			
	trnD-GTC	74			
	trnD-GTC	74			
	trnE-TTC	72			
	trnF-GAA	74			
	trnG-GCC	73			
	trnH-GTG	74			
	trnK-TTT	73			
	trnK-TTT	73			
	trnM-CAT	73			
	trnM-CAT	73			
	trnM-CAT	73			
	trnM-CAT	74			

Continued

Group of genes	Gene name	Length	Start codon	Stop codon	Amino acid
	trnM-CAT	77			
	trnN-GTT	72			
	trnP-TGG	75			
	trnQ-TTG	75			
	trnQ-TTG	72			
	trnS-GCT	88			
	*trnS-GCT**	71			
	*trnS-GCT**	71			
	trnS-TGA	87			
	trnT-GGT	74			
	*trnT-TGT**	73			
	*trnT-TGT**	73			
	*trnT-TGT**	73			
	trnW-CCA	74			
	trnY-GTA	83			

Table 1. List of encoding genes in the *L. sativa* var. *ramosa* Hort mt genome. *Intron number; #:pseudogene.

GenBank Accession number	Species	Genome length	Gene	PCG	rRNA	tRNA	Pseudo gene	Intron	AT%	GC%
PP999685	*L. sativa* var. *ramosa* Hort	363,324	71	35	6	28	2	29	54.65	45.35
MK642355.1	*L. sativa*	363,324	69	37	6	25	1	24	54.65	45.35
MK759657.1	*L. saligna*	368,269	71	39	6	25	1	24	54.8	45.2
MK820672.1	*L. serriola*	363,328	69	37	6	25	1	24	54.64	45.36
MZ159953	*L. sativa* var. *capitata*	363,324	78	42	6	29	1	24	54.65	45.35
MZ159959	*L. virosa*	373,019	79	43	6	29	1	24	54.73	45.27

Table 2. Comparison of gene content among *Lactuca* mt genomes.

Repeat sequence analysis

Repeat sequences, including SSR, tandem repeats, and interspersed repeats, were widely distributed in the mt genomes of plants, which play a critical role in genome rearrangement[34,35]. SSRs are an efficient molecular marker, which are DNA fragments comprising short sequence repeat units with a size of 1–6 base pairs[36]. In total of 110 SSRs were discovered in the *L. sativa* var. *ramosa* Hort mt genome, consisting of 21.82% (24) for monomers, 20.91% (23) for dimers, 9.09% (10) for trimers, 44.55% (49) for tetramers, 3.63% (4) for pentamers (Table 4). SSRs in monomer, dimer and tetramer motifs occupied 87.28% of all identified SSRs. The monomers included 11 of Adenine (A) and 13 of thymine (T), respectively. The TA SSR motifs were the highest abundant dimers with 30.43% of the total dimers (Table S5). Whereas the hexamers were not yet found in this genome.

Tandem repeats, also named satellite DNA, are widely present in eukaryotic genomes and some prokaryotes[37]. In *L. sativa* var. *ramosa* Hort,15 tandem repeats were identified with a matching degree more than 76%, and the sizes were between 12 and 39 bp (Table 5). Interspersed repeats is another kind of repetitive sequence, which is distributed dispersedly in the genome. A total of 120 interspersed repeats with the size ≧30 bp were obtained, of which 76 palindromic (about 63.33%) and 44 forward repeats (36.67%), and the reverse and complementary were not yet detected in this mt genome (Fig. 4). The whole size of these identified interspersed repeats was 58,124 bp, which accounted for 16% of the total mt genome. Most interspersed repeats were between 30 and 50 bp, and the maximum length of repeat was 34,696 bp (Table S6).

Pi analysis

The Pi values of 35 PCGs were calculated and ranged from 0 to 0.01032 in the *L. sativa* var. *ramosa* Hort mt genome (Fig. 5 and Table S7). The Pi values of *gene16.atp1* were the highest among all the tested regions, being 0.01032, and 0.00082 in *gene20.nad2*, 0.00046 in *gene3.nad6*, 0.0004 in *gene2.cox2*, and 0.00028 in *gene8.cob*. These genetic variations, being *atp1*, *nad2*, *nad6*, *cox2*, and *cob*, might be selected as the available molecular markers for the *Lactuca* species in the future. Most PCGs with low Pi values reflected that the mt genome of *L. sativa* var. *ramosa* Hort were relatively conserved.

Phylogenetic analysis

To affirmed the phylogenetic position of *L. sativa* var. *ramosa* Hort, a Bayesian phylogenetic tree was conducted based on a set of 31 conserved PCGs from all 28 detected mt genomes (Fig. 6). The phylogenetic tree was divided into eight groups, namely, *Lactuca*, *Chrysanthemum*, *Diplostephium*, *Aster*, *Helianthus*, *Ageratum*, *Arctium*, and *Ginkgo*. *L. sativa* var. *ramosa* Hort was well clustered with the species of *Lactuca* genus at first group, and formed sister branches with other related *Lactuca* species in the Asteraceae family clade. the mt genome of *L. sativa* var.

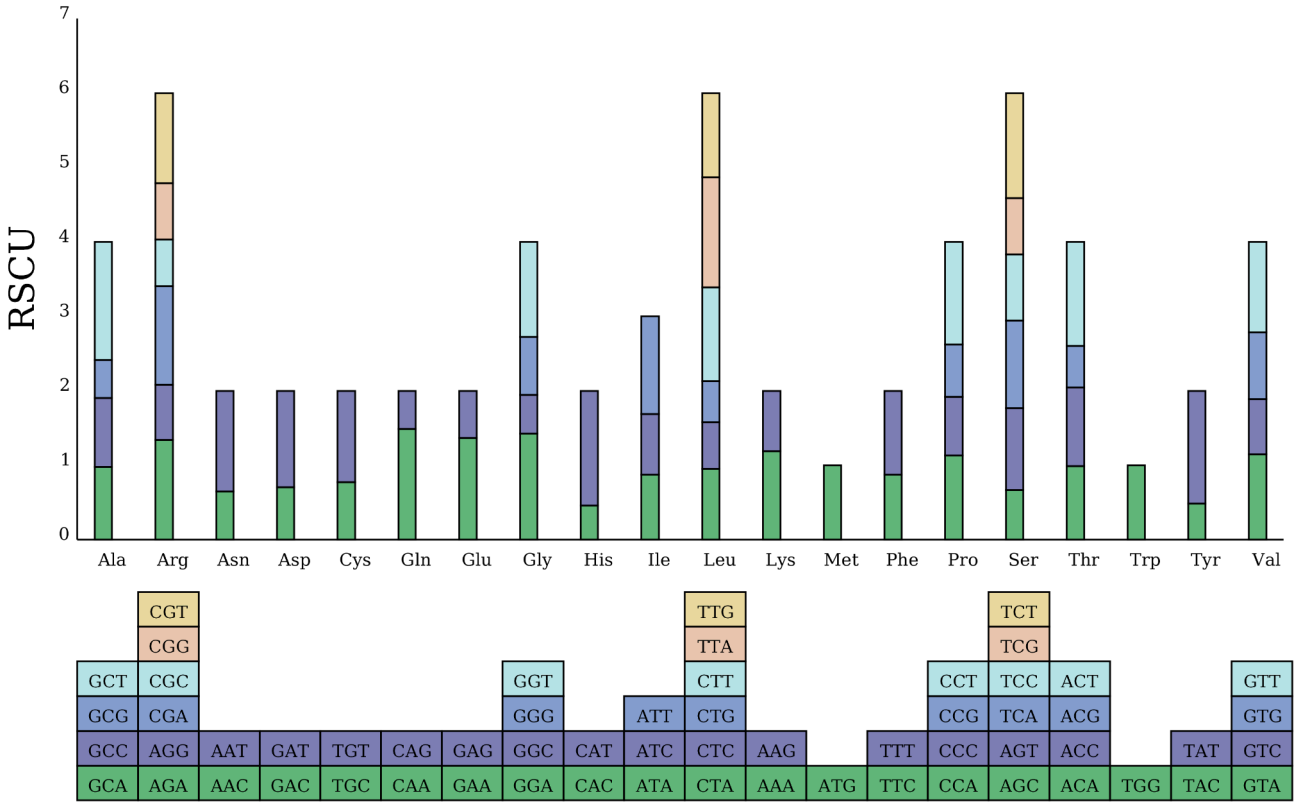

Fig. 2. RSCU analysis of the *L. sativa* var. *ramosa* Hort mt genome.

ramosa Hort was closely related not only to *L. sativa* var. *capitata* (MZ159953), but also to other mt genomes of *L. sativa* var. *capitata* and *L. virosa* to the same extent. Overall, the findings of our mt genomes analysis provide an utilizable information for future researches of the evolutionary relationships of *Lactuca* plants.

Homologous fragments transferred from cp to mt

The cp-like sequences in the mt genome were identified via comparing with the whole cp genome sequence of *L. sativa* var. *ramosa* Hort obtained from the GenBank of NCBI (PP999684). The homologous sequence had a length of 5,511 bp in the cp genome, occupied 3.61% of the entrie cp genome. Whereas the homologous sequences on the mt genome was 5,553 bp in length, accounting for 1.53% of the entrie mt genome (Table S8). A total of 15 fragments were observed in the *L. sativa* var. *ramosa* Hort mt genome, varying in length from 79 bp to 1,219 bp (Table 6). The cp-like sequences was 7,547 bp in length, accounting for 2.08% of the mt genome. Six complete tRNA genes, being *trnW-CCA*, *trnQ-TTG*, *trnD-GTC*, *trnH-GTG*, *trnN-GTT*, and *trnM-CAT*, were identified, with some homologous fragments of *rrn18* genes. We also found that 15 insertion regions in the cp genome of *L. sativa* var. *ramosa* Hort, comprising eight complete genes, including two PCGs (*petL* and *petG*,) and five tRNA genes (*trnW*, *trnP*, *trnD-GUC*, *trnN*, and *trnM*), were detected in the *L.sativa* var. *ramosa* Hort cp genome, with some homologous fragments of *rpoC1*, *rrn16*, *rbcL*, *infA*, *rps8*, and *ycf3* genes. Combined with the above findings, the tRNA genes were more conserved than PCGs and rRNAs in the mt genome of *L. sativa* var. *ramosa* Hort.

Synteny analysis of mt genome sequences

As shown in Fig. 7, the dot-plot analysis indicated that longer synteny sequences with higher similarity were identified among *L.sativa* var. *ramosa* Hort with *L. sativa* var. *capitata* than between *L. sativa* var. *ramosa* Hort and other *Lactuca* species, illustrating that *L. sativa* var. *ramosa* Hort has a similarity structure with *L. sativa* var. *capitata*. The off-diagonal signals in *L. serriola* were due to common repeat sequences. Furthermore, the sequence rearrangement events were found in *L. sativa*, *L. saligna* and *L.virosa*.

Discussion

Mitochondria are indispensable organelles in plants, which are an important place for respiration and energy conversion. Mt genomes have the characteristics of slow evolution and high conservation, which have become an ideal tool for evolutionary analysis of species[38,39]. In this study, we characterized the *L. sativa* var. *ramosa* Hort mt genome, and carried out a comparison with other related *Lactuca* species. The *L. sativa* var. *ramosa* Hort mt genome is a circular structure with a full length of 363,324 bp and 45.35% GC content, which exhibited a high similarity to *L. sativa* and *L. sativa* var. *capitata* (Table 2) . GC content is an important indicator for evaluating species. The GC content in the *L. sativa* var. *ramosa* Hort mt genome was 45.35%, which was comparable to

Type	RNA-editing	Number	Percentage
Hydrophilic–hydrophilic	CAC (H) = > TAC (Y)	6	12.20%
	CAT (H) = > TAT (Y)	16	
	CGC (R) = > TGC (C)	11	
	CGT (R) = > TGT (C)	28	
Hydrophilic–hydrophobic	ACA (T) = > ATA (I)	2	48.00%
	ACC (T) = > ATC (I)	3	
	ACG (T) = > ATG (M)	6	
	ACT (T) = > ATT (I)	4	
	CGG (R) = > TGG (W)	33	
	TCA (S) = > TTA (L)	71	
	TCC (S) = > TTC (F)	28	
	TCG (S) = > TTG (L)	42	
	TCT (S) = > TTT (F)	51	
Hydrophilic–stop	CAG (Q) = > TAG (X)	1	0.40%
	CGA (R) = > TGA (X)	1	
Hydrophobic–hydrophilic	CCA (P) = > TCA (S)	7	8.00%
	CCC (P) = > TCC (S)	11	
	CCG (P) = > TCG (S)	6	
	CCT (P) = > TCT (S)	16	
Hydrophobic–hydrophobic	CCA (P) = > CTA (L)	46	31.40%
	CCC (P) = > CTC (L)	7	
	CCC (P) = > TTC (F)	8	
	CCG (P) = > CTG (L)	37	
	CCT (P) = > CTT (L)	20	
	CCT (P) = > TTT (F)	12	
	CTC (L) = > TTC (F)	6	
	CTT (L) = > TTT (F)	14	
	GCC (A) = > GTC (V)	3	
	GCG (A) = > GTG (V)	2	
	GCT (A) = > GTT (V)	2	
Total		500	100%

Table 3. RNA-editing prediction in the *L. sativa* var. *ramosa* Hort mt genome.

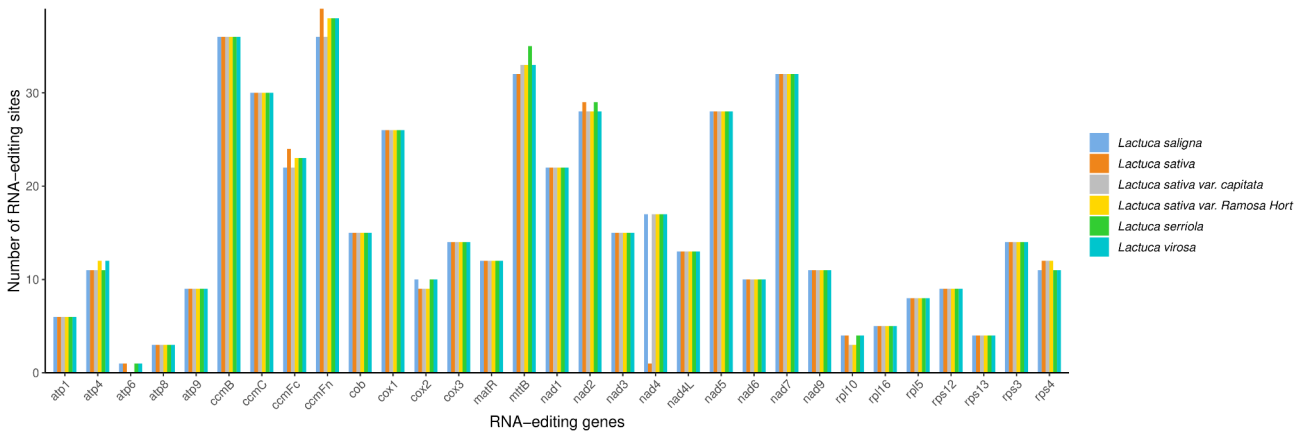

Fig. 3. Comparison of RNA-editing sites among six *Lactuca* species.

other reported mt genomes of *Lactuca* species such as *L. serriola*, 45.36%; *L. virosa* 45.27%; *L.* sativa var Salinas, 43.43%; *L. saligna*, 42.54%;[17,40], whereas showed higher than the *L. sativa* var. *ramosa* Hort cp genome (PP999684, 37.55%) sequenced by our research team. Non-coding sequence occupied 81.62% for the complete *L. sativa* var. *ramosa* Hort mt genome, which is consistent with *Brassica rapa* var. *Purpuraria*[41], *Taraxacum mongolicum*[42] and *Clematis acerifolia*[43]. Besides, the PCGs generally encoded from start codon (ATG) to terminator codon

Motif Type	Number of repeats																Total	Total Proportion (%)
	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		
Monomer	–	–	–	–	–	–	–	13	4	2	2	1	–	1	–	1	24	21.82
Dimer	–	–	18	2	3	–	–	–	–	–	–	–	–	–	–	–	23	20.91
Trimer	–	10	–	–	–	–	–	–	–	–	–	–	–	–	–	–	10	9.09
Tetramer	48	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	49	44.55
Pentamer	4	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	4	3.63
Total	52	10	19	2	3	0	0	13	4	2	2	1	0	1	0	1	110	100.00

Table 4. SSR motifs anaysis of the *L. sativa* var. *ramosa* Hort mt genome.

NO	Size	Copy	Repeat sequence	Percent Matches	Start	End
1	21	2.1	CATAAAGAGGGCCTTAGAATT	91	13,021	13,064
2	18	2.4	GACTATGAAACAGATCGC	81	14,675	14,717
3	27	1.9	AAACAATTCTATGTTCAACTTGACTAC	92	25,658	25,709
4	28	2	AAATGGTTCAACATTTGAAGTTCTTTCC	100	30,655	30,710
5	39	2	AATATCATGATCGGGTCGACCAGGCCAGATCATGAGTGA	95	36,346	36,424
6	12	3	TCATCGTCGCTA	83	42,231	42,266
7	16	2.6	CTTTCTGCGGGATCCT	88	81,439	81,478
8	21	2.1	CATAAAGAGGGCCTTAGAATT	91	125,989	126,032
9	18	2.4	GACTATGAAACAGATCGC	81	127,643	127,685
10	27	1.9	AAACAATTCTATGTTCAACTTGACTAC	92	138,626	138,677
11	28	2	AAATGGTTCAACATTTGAAGTTCTTTCC	100	143,623	143,678
12	15	3.3	AAGAGCTACAGAAGG	76	156,308	156,354
13	19	2.2	TCTTAAGTGAAGAGTAACC	100	164,263	164,303
14	22	2.3	TATAATAGTATAGAAGTCTA	82	169,130	169,175
15	21	2	TTCCTTTCAAGCTACTACCAA	82	267,834	267,877

Table 5. The identified tandem repeats in the *L. sativa* var. *ramosa* Hort mt genome.

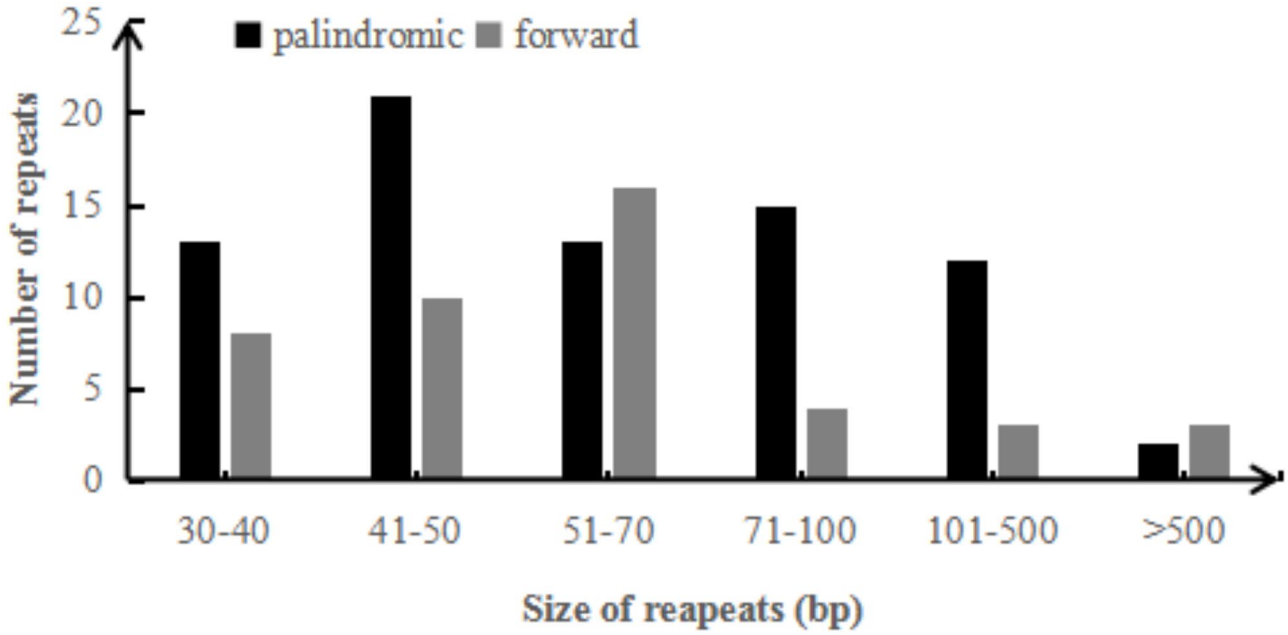

Fig. 4. Interspersed repeats identified in the *L. sativa* var. *ramosa* Hort mt genome.

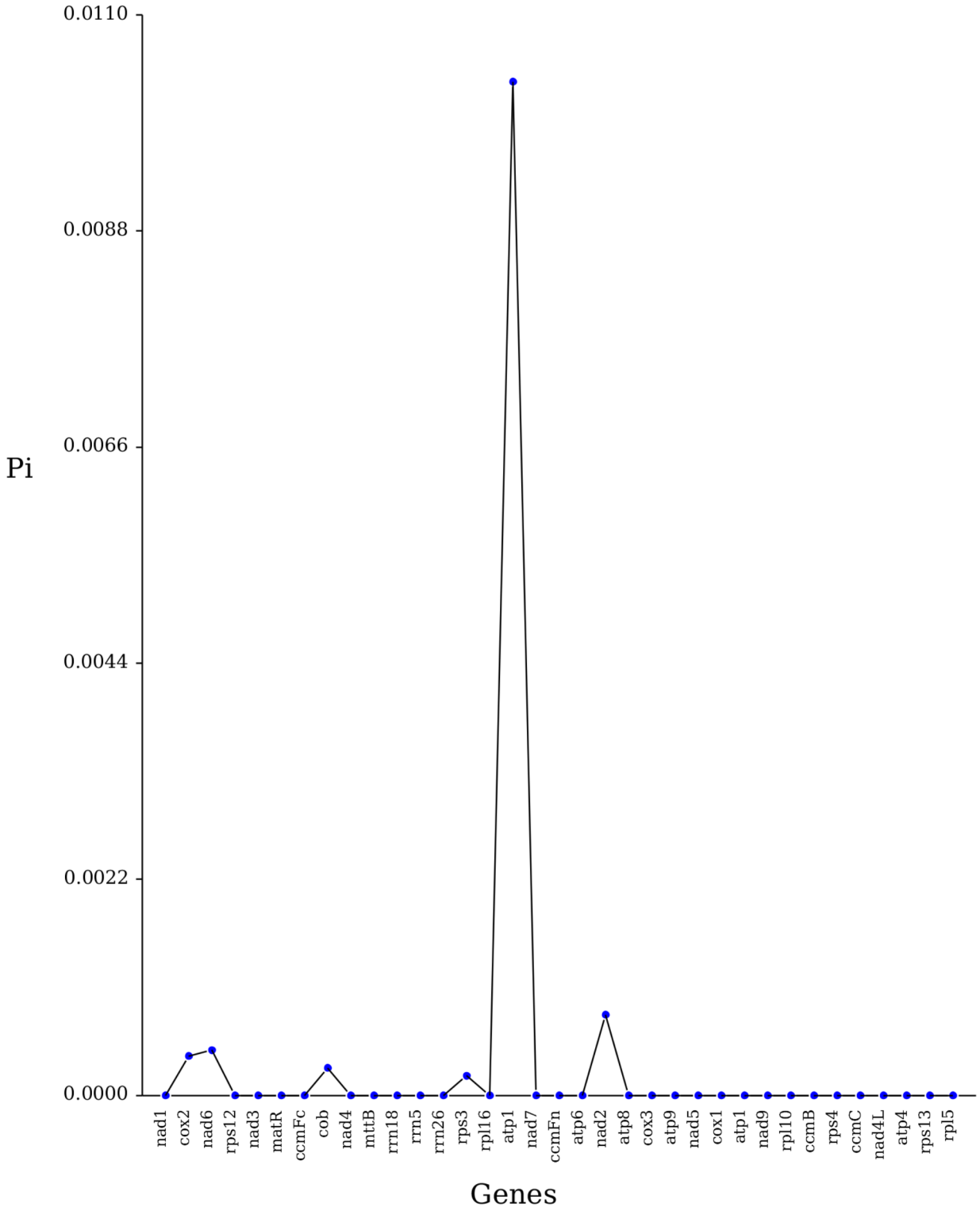

Fig. 5. Nucleotide diversity analysis of the *L. sativa* var. *ramosa* Hort mt genome.

(TGA, TAG and TAA), which accounted for 9.46% of the whole mt genome. This phenomenon was agreed with *Mesona chinensis* Benth[44] and *Luffa cylindrica*[45], which might be resulted in increasing repetitive sequences during evolution. The *cox1* gene using ACG as initiator codon in coherence with *Diospyros oleifera* might be caused by RNA editing[46].

The usage frequency of different codons encoding the same amino acid is different, which is interpreted as codon preference[47]. RSCU is an important index to evaluate the codon usage pattern of mt genome in plants[48]. Codon preference has been widely applied in genetic, domestication and systematic evolution of plant taxa[49–51].

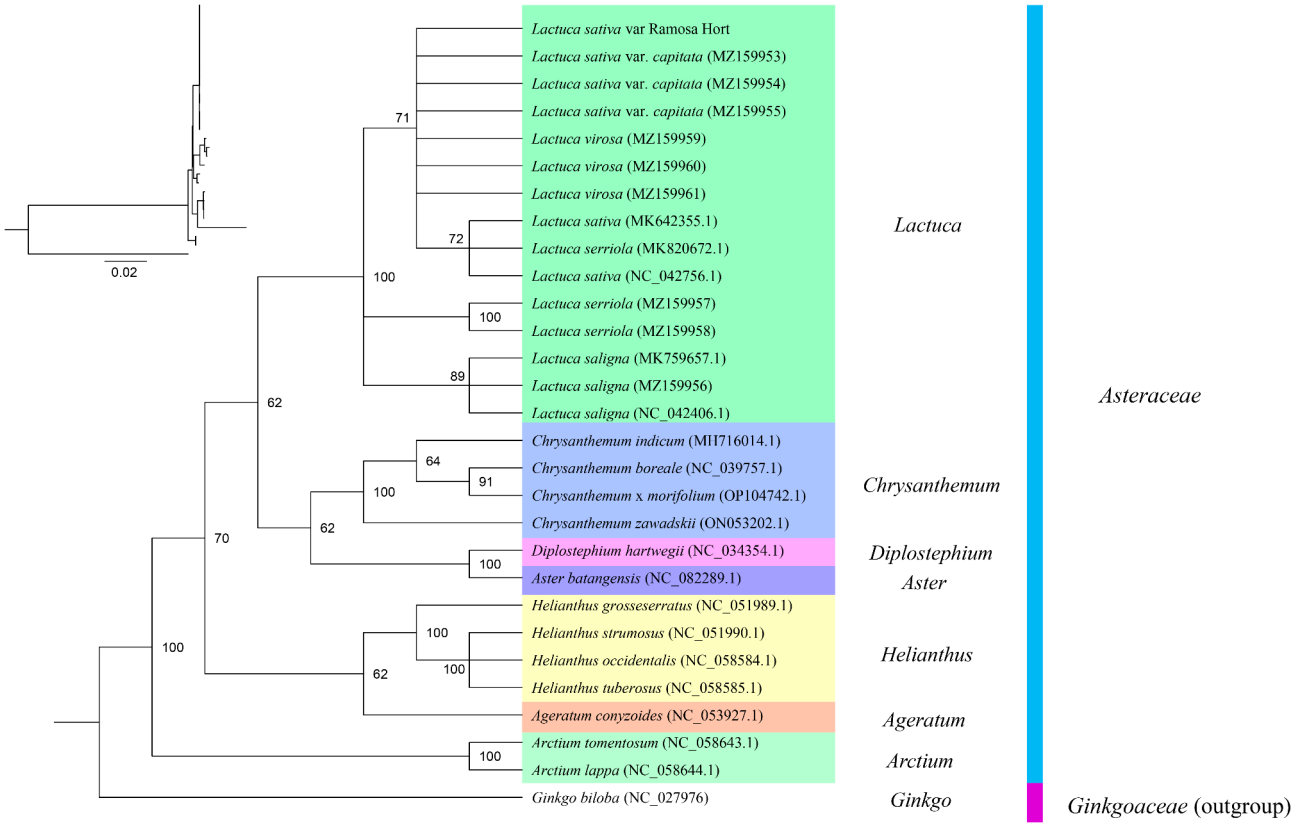

Fig. 6. A Bayesian phylogenetic tree was created based on 31 conserved PCGs among 27 Asteraceae species. *Ginkgo biloba* (NC_027976) was selected as an outgroup.

No	Identity%	Length	Mismatches	Gap	Cp start	Cp end	Mt start	Mt end	Gene(cp)	Gene(mt)
1	99.836	1,219	1	1	17,307	18,524	180,266	179,048	*rpoC1*(43.45%)	
2	77.929	1,101	171	42	65,615	66,684	277,254	278,313	petL;petG;trnW;trnP	*trnW-CCA*; *trnQ-TTG*
3	73.761	888	180	40	135,598	136,461	4,916	4,058	*rrn16*(57.99%)	*rrn18*(44.12%)
4	73.761	888	180	40	100,408	101,271	4,058	4,916	*rrn16*(57.99%)	*rrn18*(44.12%)
5	73.761	888	180	40	135,598	136,461	117,884	117,026	*rrn16*(57.99%)	*rrn18*(44.12%)
6	73.761	888	180	40	100,408	101,271	117,026	117,884	*rrn16*(57.99%)	*rrn18*(44.12%)
7	93.434	198	8	2	55,899	56,091	322,360	322,163	*rbcL*(13.46%)	
8	81.994	311	48	5	79,899	80,209	243,999	244,301	*infA*(6.41%); *rps8*(43.21%)	
9	75.726	585	68	38	43,338	43,875	183,889	184,446	*ycf3*(27.80%)	
10	88.55	131	12	2	11,596	11,723	99,070	99,200	*trnD-GUC*	*trnD-GTC*
11	88.618	123	11	2	11,599	11,718	250,740	250,618	*trnD-GUC*	*trnD-GTC*
12	97.5	80	2	0	12	91	337,495	337,574	*trnH-GUG*	*trnH-GTG*
13	94.048	84	4	1	108,267	108,349	172,674	172,591	*trnN*	*trnN-GTT*
14	94.048	84	4	1	128,520	128,602	172,591	172,674	*trnN*	*trnN-GTT*
15	94.937	79	4	0	52,115	52,193	87,323	87,401	*trnM*	*trnM-CAT*

Table 6. Fragments transferred from cp to mt in *L. sativa* var. *ramosa* Hort.

In *L. sativa* var. *ramosa* Hort, 29 high-frequency codons with RSCU > 1 were identified, of which 93.10% (27) codons preferred to end with A or U bases, which was agreed with previous studies[52–54] . Besides, the most frequently used amino acid was leucine in the *L. sativa* var. *ramosa* Hort mt genome, and the similar results were found in the *Conopomorpha sinensis*[55] and *Perilla frutescens* mt genomes[56].

RNA editing is widely existed in the mt genome of plants, which involved in plant development and stress response[57]. A total of 500 RNA-editing sites within all the 35 PCGs were predicted in the *L. sativa* var. *ramosa* Hort mt genome, which presented much higher than those in *Welwitschia* (226)[58], *Garcinia mangostana* L.variety Mesta (333)[59] and *Abelmoschus esculentus* (281)[60], and lower than those in Hypopitys monotropa

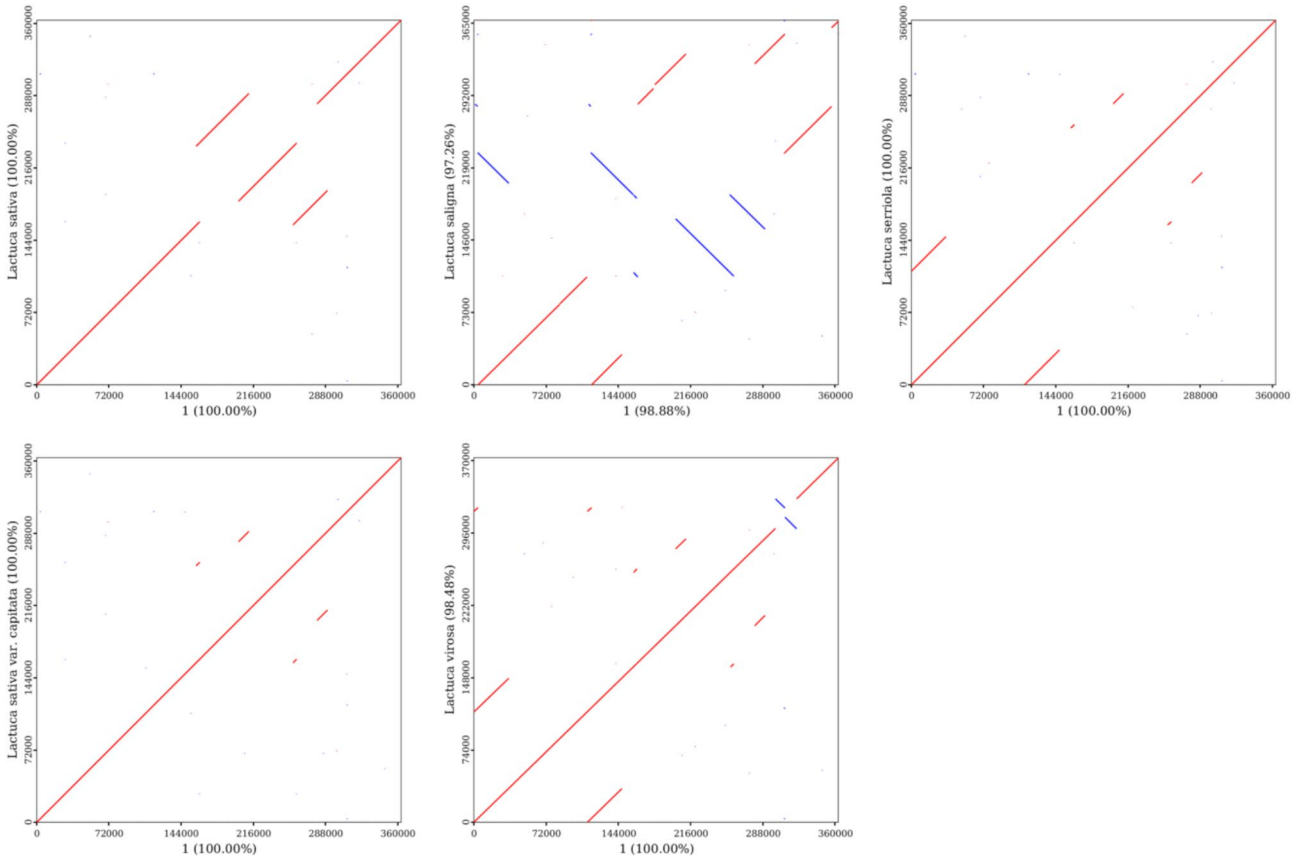

Fig. 7. Dot-plot graphs showing synteny sequences between mt genomes in *Lactuca* species compared to *L. sativa* var. *ramosa* Hort as the reference.

(545)[61] and *Pulsatilla patens* (902)[8]. Most of RNA editing sites has been found to be C-to-U conversion in plant mt genomes[62]. A total of 500 C-to-U edit sites were observed in 35 PCGs, while no U-to-C sites were found in the *L. sativa* var. *ramosa* Hort mt genome, being similar as in the Cycas[63] and Ginkgo mt genomes[58] . Most of RNA-editing sites generated at the first or second codon positions, and no RNA editing sites were observed at the third codon position in the *L. sativa* var. *ramosa* Hort mt genome. The similar results were obtained in the *Suaeda glauca*[32], *Macadamia integrifolia*[34] and *L.cylindrica*[45]. These identified RNA-editing sites provide necessary clues for exploring evolution and predicting gene function of new codons, which could help us better understand the gene expression of mt genomes in plants.

Repetitive sequences containing tandem, short and large repeats, are abundant distributed in the mt genome of higher plants, and vary from a few bp to tens kb, accounting for 6.84% ~ 58.34% of the entire mt genome[64–66]. Repetitive sequences are essential for intermolecular recombination, which can produce extreme mt genome sizes and structural variations[5] . SSR has an important function and are widely used for population diversity, genetic stability, species identification and phylogenetic analysis[67]. In *L. sativa* var. *ramosa* Hort, 110 SSRs were observed, of which 100% monomers being A or T, and 30.43% dimers being TA, resulting richness AT content (54.65%) in the *L. sativa* var. *ramosa* Hort mt genome. The abundant AT content were also found in the *Ilex metabaptista* mt genome[68]. Furthermore, the proportion of interspersed repeats in the *L. sativa* var. *ramosa* Hort mt genome (16%) was less than that of *Acer yangbiense* (17.20%) and *A. truncatum* (18.24%), and the largest interspersed repeats were 34,696 bp, 27,124 bp and 28,452 bp, respectively[69,70]. Besides, 15 tandem repeats were obtained in *L. sativa* var. *ramosa* Hort, which was much less than *Selenicereus monacanthus* (94)[71] and *Cyperus esculentus* L.(82)[72]. The repeats obtained in this study will provide valuable information for future study on developing potential molecular markers and genetic evolution in the *Lactuca* species.

Genetic diversity refers to the variation of genes within an organism, including genetic variations between significantly different populations within the same or different species[73]. Studying the genetic diversity of crop populations will help us to better understand the genetic structure, highly variability regions and genetic background[74]. Previous studies reported that the highly variable regions could be designed as potential molecular markers for population genetics[68,75]. In *L. sativa* var. *ramosa* Hort, the highest Pi value of all the PCGs was *atp1* gene, revealing that *atp1* gene might be developed as an available molecular marker for the *Lactuca* species. The *atp1* gene was widely identified in the plant mt genomes, and involved in the ATP synthase[76,77]. Whereas in *Ilex metabaptista*, *atp9* genes (Pi = 0.114) showed the largest variability, which also played an important in the ATP synthase[68]. In our study, five hotspots, namely, *atp1*, *nad2*, *nad6*, and *cox2*, were found and used as potentially molecular markers. Three highly variable regions, being *atp9, sdh3* and *cox2,* were selected as molecular marker

in the *Ilex metabaptista* mt *genome*[68]. while four hotspots, being *rpl5*, *atp8*, *rps3*, and *nad1*, were obtained in the *Piophila casei* mt genome, of which might potentially use as molecular markers[75]. Most PCGs with lower Pi values declared that the *L. sativa* var. *ramosa* Hort mt genome was highly conserved.

The genome-wide data was widely used to analyze the evolutionary relationship among different species[68,78,79]. It is not clear that different lettuce species are involved in the domestication and/or diversification of *L.sativa*. From the -perspective of nuclear genome, *L.serriola* is considered as one of the direct ancestors of *L.sativa* and the closest relationship with *L.sativa*[20,80]. The rapid development of sequencing technologies and the recent increase in sequenced genomes contributed to illustrating the relationship between *Lactuca* species. The *Lactuca* species were well clustered and subdivided into several clades including *L. sativa*, *L. serriola*, *L. virosa*, and *L. saligna*[81]. In this study, the Bayesian phylogenetic was conducted based on 27 mt genomes of Asteraceae species and an outgroup mt genome. *L. sativa* var. *ramosa* Hort was well clustered with the species of *Lactuca* genus, and was closely related to *L. sativa* var. *capitata* and *L. virosa* , implicating that *L. sativa* var. *ramosa* Hort belongs to the *Lactuca* genus in the Asteraceae family. The similar results were obtained in the analysis of the whole genome resequencing of 445 *Lactuca* species[81]. Additionally, the synteny analysis showed that *L. sativa* var. *ramosa* Hort has a similarity structure with *L. sativa* var. *capitata*. The sequence rearrangement events were observed in *L. sativa*, *L. saligna* and *L.virosa* compared to *L. sativa* var. *ramosa* Hort. Although *L. sativa* var. *ramosa* Hort has the same size with *L. sativa* var. *capitata*, it exhibits minor differences in gene content in comparison with *L. sativa* var. *capitata* (Table 2). Gene mutation, homologous sequence interference or sequencing artifacts might be caused for these differences between *L. sativa* var. *ramosa* Hort and *L. sativa* var. *capitata*. These minor differences might be also caused due to the different genetic characteristics of different *Lactuca* varieties. Intraspecifc variations in mt genome sequence and gene content have been identified in six *Lactuca* varieties, which helped to distinguish this genome from previously sequenced *L. sativa* mt genomes.

Sequences migrated from the cp genome can be found in the plant mt genome, usually accounting for 1–12% of the whole mt genome[82]. About 33.33% tRNA genes originated from cp genome and gradually migrated during evolution[83]. The total length of migrated sequences varies from 50 kb (*Arabidopsis thaliana*) to 1.1 Mb (*Oryza sativa* subsp.japonica) based on the plant species[84]. In our study, 15 fragments with the total length of 7,547 bp (2.08% of the *L. sativa* var. *ramosa* Hort mt genome) migrated from cp to mt genomes, implicating that these transferred fragments might play an important role in evolution. Seven genes, including six tRNA genes (*trnW-CCA*, *trnQ-TTG*, *trnD-GTC*, *trnH-GTG*, *trnN-GTT*, and *trnM-CAT*) and *rrn18*, were migrated between cp and mt genomes. According to previous studies on higher plants, about 42% of the cp genome fragments were integrated into the *Vitis vinifera* mt genome with a length of 773,279 bp, including more than 30 cp PCGs and 17 tRNA genes[85]. In addition, over than 113 kb cp migrated sequences were found in the *Cucurbita pepo* mt genome, and most of transferred genes were tRNA genes[5]. Combined with the above findings, tRNA genes are more conserved than PCGs in the mt genome of *L. sativa* var. *ramosa* Hort, which might be a character of a mt genome during the process of evolution in *Lactuca* species.

Conclusion

In this study, we sequenced and successfully drew the genome with a typical circular structure in the *L. sativa* var. *ramosa* Hort mt genome. Its genome has a length of 363,324 bp, consisting of 71 genes with 35 PCGs, 6 rRNAs, 28 tRNAs, and 2 pseudogenes, within 45.35% GC content. Subsequently, we carried out studies on codon preference, SSRs, tandem repeats and interspersed repeats in the *L. sativa* var. *ramosa* Hort mt genome. Additionally, 500 RNA-editing sites were detected in 35 PCGs, which is helpful to predict gene function by using new codons. Based on gene migration analysis, a total of 15 fragments, including six complete tRNA genes, were migrated from cp genome to mt genome. Most PCGs with low Pi values illustrated that the mt genome was conserved in *L. sativa* var. *ramosa* Hort. Phylogenetic analysis confirmed that *L. sativa* var. *ramosa* Hort is genetically closer to *L. sativa* var. *capitata* and and *L. virosa*, which belongs to the *Lactuca* genus in the Asteraceae family. In summary, *L. sativa* var. *ramosa* Hort has a similarity structure with *L. sativa* var. *capitata*, but displays minor differences in gene content compared to *L. sativa* var. *capitata*.

Data availability

The new obtained mt genome sequence was submitted in GenBank of NCBI (Accession Number: PP999685).

Received: 20 September 2024; Accepted: 10 March 2025
Published online: 18 March 2025

References

1. Ala, K. G. et al. Comparative analysis of mitochondrial genomes of two alpine medicinal plants of *Gentiana* (Gentianaceae). *PLoS ONE* **18**(1), e0281134 (2023).
2. Wang, J. et al. Mitochondrial functions in plant immunity. *Trends Plant Sci.* **27**(10), 1063–1076 (2022).
3. de Oca, B. P. M. Mitochondria-plasma membrane interactions and communication. *J. Biol. Chem.* **297**(4), 101164 (2021).
4. Richardson, A. O. et al. The "fossilized" mitochondrial genome of *Liriodendron tulipifera*: Ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biol.* **11**, 1–17 (2013).
5. Alverson, A. J. et al. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Mol. Biol. Evol.* **27**(6), 1436–1448 (2010).
6. Robles, P. & Quesada, V. Organelle genetics in plants. *Int. J. Mol. Sci.* **22**(4), 2104 (2021).
7. Xiong, A. S. et al. Gene duplication and transfer events in plant mitochondria genome. *Biochem. Biophys. Res. Commun.* **376**(1), 1–4 (2008).
8. Szandar, K. et al. Breaking the limits-multichromosomal structure of an early eudicot Pulsatilla patens mitogenome reveals extensive RNA-editing, longest repeats and chloroplast derived regions among sequenced land plant mitogenomes. *BMC Plant Biol.* **22**(1), 109 (2022).

9. Morley, S. A. & Nielsen, B. L. Plant mitochondrial DNA. *Molecules* **15**(17), 10–2741 (2017).
10. Mower, J. P. Variation in protein gene and intron content among land plant mitogenomes. *Mitochondrion* **53**, 203–213 (2020).
11. Shan, Y. et al. The complete mitochondrial genome of *Amorphophallus albus* and development of molecular markers for five *Amorphophallus* species based on mitochondrial DNA. *Front. Plant Sci.* **14**, 1180417 (2023).
12. Yang, J. et al. A comparative genomics approach for analysis of complete mitogenomes of five Actinidiaceae plants. *Genes* **13**(10), 1827 (2022).
13. Liu, Q. et al. Complete mitochondrial genome of the endangered *Prunus pedunculata* (Prunoideae, Rosaceae) in China: Characterization and phylogenetic analysis. *Front. Plant Sci.* **14**, 1266797 (2023).
14. Luo, X. L. et al. Impact of iron supplementation on media bed aquaponic system: Focusing on growth performance, fatty acid composition and health status of juvenile mirror carp (*Cyprinus carpio* var. *specularis*), production of lettuce (*Lactuca sativa var. ramosa* Hort), water quality and nitrogen utilization rate. *Aquaculture* **582**, 74050 (2024).
15. Yu, X. et al. Comparative analysis of Italian lettuce (*Lactuca sativa* L. var. *ramose*) transcriptome profiles reveals the molecular mechanism on exogenous melatonin preventing cadmium toxicity. *Genes* **13**(6), 955 (2022).
16. Lindqvist, K. On the origin of cultivated lettuce. *Hereditas* **46**, 319–350 (1960).
17. Fertet, A. et al. Sequence of the mitochondrial genome of *Lactuca virosa* suggests an unexpected role in *Lactuca sativa*'s evolution. *Front. Plant Sci.* **12**, 697136 (2021).
18. Davis, R. M. et al. *Compendium of Lettuce Diseases* (APS Press, 1997).
19 Kesseli, R., Ochoa, O. & Michelmore, R. Variation at RFLP loci in *Lactuca* spp. and origin of cultivated lettuce (L. sativa). *Genome* **34**(3), 430–436 (1991).
20 De Vries, I. M. & Van Raamsdonk, L. W. D. Numerical morphological analysis of lettuce cultivars and species (*Lactuca* sect. Lactuca, Asteraceae). *Plant Syst. Evol.* **193**, 125–141 (1994).
21. Wei, Z. et al. Phylogenetic relationships within *Lactuca* L. (Asteraceae), including African species, based on chloroplast DNA sequence comparisons. *Genet. Resour. Crop Evol.* **64**(1), 55–71 (2017).
22. Kozik, A. et al. The alternative reality of plant mitochondrial DNA: One ring does not rule them all. *PLoS Genet.* **15**(8), e1008373 (2019).
23. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**(18), 3094–3100 (2018).
24. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**(5), 722–736 (2017).
25 Chan, P. P. et al. tRNAscan-SE 2.0: Improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**(16), 9077–9096 (2021).
26. Lohse, M., Drechsel, O. & Bock, R. OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* **52**, 267–274 (2007).
27. Mower, J. PREP-Mt: Predictive RNA editor for plant mitochondrial genes. *BMC Bioinf.* **6**, 96 (2005).
28. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**(15), 1972–1973 (2009).
29. Kalyaanamoorthy, S. et al. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**(6), 587–589 (2017).
30. Qiao, Y. et al. Assembly and comparative analysis of the complete mitochondrial genome of *Bupleurum chinense* DC. *BMC Genom.* **23**(1), 1–17 (2022).
31. Alverson, A. J. et al. The mitochondrial genome of the legume *Vigna radiata* and the analysis of recombination across short mitochondrial repeats. *PLoS ONE* **6**(1), e16404 (2011).
32. Cruz Plancarte, D. & Solórzano, S. Structural and gene composition variation of the complete mitochondrial genome of *Mammillaria huitzilopochtli* (Cactaceae, Caryophyllales), revealed by de novo assembly. *BMC Genom.* **24**(1), 509 (2023).
33. Yan, C. et al. Assembly and comparative analysis of the complete mitochondrial genome of *Suaeda glauca*. *BMC Genom.* **22**(1), 15–25 (2021).
34. Wu, S. et al. Extensive genomic rearrangements mediated by repetitive sequences in plastomes of Medicago and its relatives. *BMC Plant Biol.* **21**, 1–16 (2021).
35 Niu, Y. et al. Assembly and comparative analysis of the complete mitochondrial genome of three Macadamia species (*M. integrifolia*, *M. ternifolia* and *M. tetraphylla*). *Plos ONE* **17**(5), e0263545 (2022).
36. Liu, Y. C. et al. Exploiting EST databases for the development and characterization of EST-SSR markers in blueberry (*Vaccinium*) and their cross-species transferability in *Vaccinium* spp.. *Sci. Hortic.* **176**, 319–329 (2014).
37. Huan, G. A. O. & Jie, K. O. N. G. Distribution characteristics and biological function of tandem repeat sequences in the genomes of different organisms. *Zool. Res.* **26**(5), 555–564 (2005).
38. Bi, C. et al. Characterization and analysis of the mitochondrial genome of common bean *(Phaseolus vulgaris*) by comparative genomic approaches. *Int. J. Mol. Sci.* **21**(11), 3778 (2020).
39. Møller, I. M., Rasmusson, A. G. & Van Aken, O. Plant mitochondria–past, present and future. *Plant J.* **108**(4), 912–959 (2021).
40. Zhang, Y. et al. Codon usage bias analysis of cultivated and wild lettuce mitochondrial genomes. *J. Henan Agric. Sci.* **51**(10), 114 (2022).
41. Gong, Y. et al. Assembly and comparative analysis of the complete mitochondrial genome of *Brassica rapa* var. *Purpuraria*. *BMC Genom.* **25**(1), 546 (2024).
42. Jiang, M. et al. Characterisation of the complete mitochondrial genome of *Taraxacum mongolicum* revealed five repeat-mediated recombinations. *Plant Cell Rep.* **42**(4), 775–789 (2023).
43. Liu, D. 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. *Front. Genet.* **13**, 1050040 (2023).
44. Tang, D. et al. Mitochondrial genome characteristics and phylogenetic analysis of the medicinal and edible plant *Mesona chinensis* Benth. *Front. Genet.* **13**, 1056389 (2023).
45. Gong, Y. et al. Assembly and comparative analysis of the complete mitochondrial genome of white towel gourd (*Luffa cylindrica*). *Genomics* **116**, 110859 (2024).
46. Xu, Y. et al. Characterization and phylogenetic analysis of the complete mitochondrial genome sequence of *Diospyros oleifera*, the first representative from the family Ebenaceae. *Heliyon* **8**(7), e09870 (2022).
47. De Oliveira, J. L. et al. Inferring adaptive codon preference to understand sources of selection shaping codon usage bias. *Mol. Biol. Evol.* **38**(8), 3247–3266 (2021).
48. Tang, D. et al. Codon usage bias and evolution analysis in the mitochondrial genome of *Mesona chinensis* Benth. *Acta Physiol. Plantarum* **44**(11), 118 (2022).
49. Li, Q. et al. Analysis of synonymous codon usage patterns in mitochondrial genomes of nine *Amanita* species. *Front. Microbiol.* **14**, 1134228 (2023).
50. Gao, W. et al. Intraspecific and interspecific variations in the synonymous codon usage in mitochondrial genomes of 8 pleurotus strains. *BMC Genom.* **25**(1), 456 (2024).
51. Li, J. et al. Complete mitochondrial genome of *Agrostis stolonifera*: insights into structure, Codon usage, repeats, and RNA editing. *BMC Genom.* **24**(1), 466 (2023).

52. Hao, Z. et al. Complete mitochondrial genome of *Melia azedarach* L., reveals two conformations generated by the repeat sequence mediated recombination. *BMC Plant Biol.* **24**(1), 645 (2024).
53. Yang, Y. et al. Structural characteristics and phylogenetic analysis of the mitochondrial genomes of four *Krisna* species (Hemiptera: Cicadellidae: Iassinae). *Genes* **14**(6), 1175 (2023).
54. Song, N., Geng, Y. & Li, X. The mitochondrial genome of the phytopathogenic fungus *Bipolaris sorokiniana* and the utility of mitochondrial genome to infer phylogeny of Dothideomycetes. *Front. Microbiol.* **11**, 863 (2020).
55. Chang, H. et al. Comparative genome and phylogenetic analysis revealed the complex mitochondrial genome and phylogenetic position of *Conopomorpha sinensis* Bradley. *Sci. Rep.* **13**(1), 4989 (2023).
56. Wang, R. et al. Insights into structure, codon usage, repeats, and RNA editing of the complete mitochondrial genome of *Perilla frutescens* (Lamiaceae). *Sci. Rep.* **14**(1), 13940 (2024).
57. Tang, W. & Luo, C. Molecular and functional diversity of RNA editing in plant mitochondria. *Mol. Biotechnol.* **60**(12), 935–945 (2018).
58. Guo, W. et al. *Ginkgo* and *Welwitschia* mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution. *Mol. Biol. Evol.* **33**(6), 1448–1460 (2016).
59. Wee, C. C. et al. Mitochondrial genome of *Garcinia mangostana* L. variety Mesta. *Sci. Rep.* **12**(1), 9480 (2022).
60. Li, J. et al. 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 Genom.* **23**(1), 481 (2022).
61. Shtratnikova, V. Y. et al. Mitochondrial genome of the nonphotosynthetic mycoheterotrophic plant *Hypopitys monotropa*, its structure, gene expression and RNA editing. *PeerJ* **8**, e9309 (2020).
62. Hao, W. et al. RNA editing and its roles in plant organelles. *Front. Genet.* **12**, 757109 (2021).
63. Salmans, M. L. et al. Editing site analysis in a gymnosperm mitochondrial genome reveals similarities with angiosperm mitochondrial genomes. *Curr. Genet.* **56**, 439–446 (2010).
64. Ogihara, Y. et al. Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. *Nucleic Acids Res.* **33**(19), 6235–6250 (2005).
65. Clifton, S. W. et al. Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant Physiol.* **136**(3), 3486–3503 (2004).
66 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.* **31**(20), 5907–5916 (2003).
67. Saski, C. et al. Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. *Plant Mol. Biol.* **59**, 309–322 (2005).
68. Zhou, P. et al. Assembly and comparative analysis of the complete mitochondrial genome of *Ilex metabaptista* (Aquifoliaceae), a Chinese endemic species with a narrow distribution. *BMC Plant Biol.* **23**(1), 393 (2023).
69. Yang, J. et al. De novo genome assembly of the endangered *Acer yangbiense*, a plant species with extremely small populations endemic to Yunnan Province, China. *Gigascience* **8**(7), giz085 (2019).
70. Ma, Q. et al. Assembly and comparative analysis of the first complete mitochondrial genome of *Acer truncatum* Bunge: a woody oil-tree species producing nervonic acid. *BMC Plant Biol.* **22**, 1–17 (2022).
71. Lu, G. et al. Complete mitogenome assembly of *Selenicereus monacanthus* revealed its molecular features, genome evolution, and phylogenetic implications. *BMC Plant Biol.* **23**(1), 541 (2023).
72. Niu, L. et al. Complete mitochondrial genome sequence and comparative analysis of the cultivated yellow nutsedge. *Plant Genome* **15**(3), e20239 (2022).
73. Carvalho, Y. G. S. et al. Recent trends in research on the genetic diversity of plants: Implications for conservation. *Diversity* **11**(4), 62 (2019).
74. Salgotra, R. K. & Chauhan, B. S. Genetic diversity, conservation, and utilization of plant genetic resources. *Genes* **14**(1), 174 (2023).
75. Bi, S. et al. Complete mitochondrial genome of *Piophila casei* (Diptera: Piophilidae): Genome description and phylogenetic implications. *Genes* **14**(4), 883 (2023).
76. Li, J. 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. *Planta* **254**, 1–16 (2021).
77. Han, F. et al. Unraveling the complex evolutionary features of the *Cinnamomum camphora* mitochondrial genome. *Plant Cell Rep.* **43**(7), 183 (2024).
78 Li, J. et al. Complete mitochondrial genome assembly and comparison of *Camellia sinensis* var. Assamica cv. Duntsa. *Front. Plant Sci.* **14**, 1117002 (2023).
79. Zhong, F. et al. Comprehensive analysis of the complete mitochondrial genomes of three *Coptis* species (*C. chinensis*, *C. deltoidea* and *C. omeiensis*): The important medicinal plants in China. *Front. Plant Sci.* **14**, 1166420 (2023).
80 De Vries, I. M. Crossing experiments of lettuce cultivars and species (Lactuca sect. Lactuca, Compositae). *Plant Syst. Evol.* **171**, 233–248 (1990).
81. Wei, T. et al. Whole-genome resequencing of 445 Lactuca accessions reveals the domestication history of cultivated lettuce. *Nat. Genet.* **53**(5), 752–760 (2021).
82. Mower, J. P., Sloan, D. B. & Alverson, A. J. Plant mitochondrial genome diversity: The genomics revolution. Plant genome diversity volume 1: Plant genomes, their residents, and their evolutionary dynamics 123–144 (2012).
83. Hoffmann, M. et al. The RNA world of plant mitochondria. *Prog. Nucleic Acid Res. Mol. Biol.* **70**(70), 19–54 (2001).
84. Straub, S. C. K. et al. Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (Apocynaceae). *Genome Biol. Evol.* **5**(10), 1872–1885 (2013).
85. Vadim, V. G. et al. Mitochondrial DNA of *vitis vinifera* and the issue of rampant horizontal gene transfer. *Mol. Biol. Evol.* **6**(1), 99–110 (2009).

Author contributions

Y.H. G. and G.H. Z. designed and conducted the experiment. Y.Y. W, H.T.L and P. L. contributed to the plant materials collection. Q. Y. L. and R. L. contributed to the data analysis. Y.H. G. and G.H. Z. wrote, reviewed, and edited the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by the construct program of plant protection applied characteristic discipline in Hunan Province and Hunan Natural Science Regional Joint Fund Project (Grant Number 2024JJ7235).

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-93762-3.

Correspondence and requests for materials should be addressed to Y.G. or G.Z.

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

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