Comparative Analysis of the Mitochondrial Genome Sequences of two Medicinal Plants: Arctium Lappa and A. Tomentosum

Dachuan Zhang  
Liaoning University of Traditional Chinese Medicine - Dalian Campus

Yanping Xing  
Liaoning University of Traditional Chinese Medicine - Dalian Campus

Liang Xu (✉️ 861364054@qq.com )  
Rong Zhao  
Liaoning University of Traditional Chinese Medicine - Dalian Campus

Yanyun Yang  
Liaoning University of Traditional Chinese Medicine - Dalian Campus

Tingting Zhang  
Liaoning University of Traditional Chinese Medicine - Dalian Campus

Shengnan Li  
Liaoning University of Traditional Chinese Medicine - Dalian Campus

Guihua Bao  
Inner Mongolia University for Nationalities

Wuliji Ao  
Inner Mongolia University for Nationalities

Tao Liu  
China Medical University

Tingguo Kang  
Liaoning University of Traditional Chinese Medicine - Dalian Campus

Research article

Keywords: Arctium lappa, Arctium tomentosum, Mitochondrial genome, Comparative analysis

DOI: https://doi.org/10.21203/rs.3.rs-28443/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License. 
Read Full License
Abstract

Background: Mitochondrial genome sequence analysis is of great significance for understanding the evolution and genome structure of different plant species. *Arctium lappa* and *A. tomentosum* are distributed in China and frequently used as medicinal plants. People usually think *A. tomentosum* is an adulterant or substitute of *A. lappa* as a traditional Chinese Medicine (TCM). It is therefore critically important to identify the different species that are utilized in medicinal applications. This study aims to determine and compare their mitochondrial genomes, gene structure, and phylogenetic relationship. These results may provide additional insights into development of genetic research.

Results: We determined the complete sequences of the mitochondrial genomes of *A. lappa* and *A. tomentosum* for the first time. The mitochondrial genomes of *A. lappa* and *A. tomentosum* were assembled into 2 single circular molecules of 312598 bp and 312609 bp, respectively. A total of 131 and 130 genes were annotated in two plants. 50 pairs of large repeat sequences were detected in *A. lappa* and *A. tomentosum*. The number of simple sequence repeats (SSRs) in both species was 192 while the total length of SSR was 2491 bp for *A. lappa* and 2489 bp for *A. tomentosum*. Only 51 single nucleotide polymorphisms (SNPs) and 3 insertion-deletions (InDels) were detected between the two plants. The two mitochondrial genome structures were highly similar and highly collinear. Both of the chloroplast genomes and mitochondrial genomes of the two plants had the phenomenon of gene exchange and transfer. Core genes and specific genes were analyzed for *A. lappa* and *A. tomentosum* and three closely related Asteraceae species, the specific gene of *A. lappa* was *orf115a*. In addition, a phylogenetic tree of the mitochondrial genomes were constructed, which laced the two *Arctium* species into one branch within Asteraceae.

Conclusions: We identified and analyzed the mitochondrial genome features of two species of *Arctium* in China with implications for species identification and phylogenetic analysis. The mitochondrial genomes of *A. lappa* and *A. tomentosum* were very similar in size and structure. The ORF genes of the two were different, which could provide a theoretical basis for the development of molecular markers.

Background

As a semi-autonomous organelle in eukaryotic cells, mitochondria are the main site of intracellular oxidative phosphorylation and synthesis of adenosine triphosphate (ATP), and its genome is independent of the nuclear genome. As sequencing technology matures, the mitochondrial genome of many crops has been sequenced. It is found that the mitochondrial genome of *Cucumis melo* L. (2500 kb) is the largest mitochondrial genome of all crops and plants that have been sequenced [1]. And a kind of medicinal plant, *Marchantia polymorpha* L. (183 kb), is the smallest [2]. The in-depth study also found that crop cytoplasmic male sterility (CMS) is closely related to the structural changes of the mitochondrial genome [3], and CMS is widely used in heterosis breeding to significantly increase crop yield [4–8]. Therefore, it is important to study the mitochondrial genome of crops and plants. How to develop high-yield and high-quality new varieties of crops is an important topic in agricultural research.
The mitochondrial genome of plants is usually similar to the chloroplast genome and has a circular molecular structure, which is complex and highly variable, with abundant non-coding regions and introns. Studies have shown that plant mitochondrial genome is a mixture of DNA molecules with different shapes [9]. The mitochondrial genomes of chrysanthemum, sunflower and *diplosephium hartweegii* are all circular [10–11]. In addition, there are also plant mitochondrial genomes with multiple rings, such as wheat, rape, etc. [12–13]. The complexity of genome structure is related to the number, location and direction of the recombinant active repeat region on mitochondria [1]. The secondary structure of mitochondrial genome is complex and changeable, but its sequence is relatively fixed, which has little influence on medicinal plant identification and research. The size, structure and gene sequence of mitochondrial genome in angiosperms vary greatly [14–16]. There are significant differences in narrow taxonomic span [17]. The mitochondrial genome of plants is much larger than that of animals, ranging from 200 to 2500 KB, with a variation of more than 10 times [15]. In most angiosperms, the size of mitochondrial genome is concentrated at 300–600 KB. At present, a large number of mitochondrial genomes of plants have been reported, which lays a foundation for comparative analysis of mitochondrial genomes among different plants. The mitochondrial genome of angiosperms can not only reveal the phylogenetic relationship between species, but also be used to study intraspecific differentiation [18]. It is necessary to study the differences between different species or even subspecies of mitochondrial genome of higher plants, the position of mitochondria in biological evolution, and better understand the evolution of higher plants and their differentiation time, and give full play to their practical application in genetic breeding, which is advantageous to species with economic benefits, representative species in biological evolution, and scientific research. Large scale mitochondrial genome sequencing of model species is very necessary and meaningful.

Most of traditional Chinese medicine comes from natural plants. Due to various reasons, there has been a variety of confusion and uneven quality of traditional Chinese medicine for a long time, which seriously affects the effectiveness, stability and safety of traditional Chinese medicine. Research and sort out the varieties of traditional Chinese medicine has always been one of the basic tasks of the research. It is limited to identify the original species of traditional Chinese medicine only based on traditional methods like morphological observation and microscopic observation. In recent years, the application of DNA barcode technology in the identification of traditional Chinese medicine has made rapid progress, and some achievements have been made. The molecular research on the original plants of traditional Chinese medicine has gradually become a hot topic. Researchers used mitochondrial genes nad1/b-c and nad5/d-e to study species identification of *Gentiana macrophylla* [19].

There are approximately 11 species of *Arctium* (Asteraceae) in the world, and they are widely distributed in temperate regions of Eurasia. *Arctium lappa* and *A. tomentosum* are distributed in China and used as medicinal plants as well as snacks [20–21]. *A. lappa*, known as burdock, is called “Niu Bang” in Chinese. Arctii Fructus is dried ripe fruit of burdock as a traditional Chinese medicine, which is included in Chinese Pharmacopoeia 2015 edition [22]. Its main effects include dispelling wind-heat, clearing the lungs, draining skin eruptions, detoxifying, and relieving a sore throat and its leaves and stems are eaten raw as a snack or stewed [23]. Arctiin and its aglucone obtained from the tap roots of burdock can interfere with
early stages of replication of the avian influenza virus and can also hamper progeny virus release in mammalian cells [24–25]. Although burdock has a large amount of cultivation in China, its yield is not ideal enough, and there is no good economic benefit. *A. tomentosum*, known as cotton/woolly burdock, is a biennial wild plant species and is distributed worldwide and used as food and rich sources of secondary metabolites for the pharmaceutical industry [24]. *A. tomentosum* (which is not included in Chinese Pharmacopoeia 2015 edition) is widely used in Xinjiang province, which belongs to the characteristic national medicine in China, but people usually think *A. tomentosum* is an adulterant or substitute of *A. lappa.* At present, the genetic research of *Arctium* plants in China is mainly focused on the identification of internal transcribed spacer sequence and the study of whole chloroplast genome sequence [20].

In this study, we determined the complete sequences of the mitochondrial genomes of *A. lappa* and *A. tomentosum* and compared their structures, and with that of three other Asteraceae plants at the two levels of genome and gene, including syntenic sequence analysis, SNP, InDel, SV detection, Core-pan analysis and species phylogenetic analysis. We also analyzed the transfer of mitochondrial genes and chloroplast genes between the two medicinal plants. Hope the results and conclusions can provide a theoretical basis for their identification, molecular markers and genetic breeding.

**Methods**

**Plant materials and mitochondrial DNA extraction**

The fruits of *A. lappa* and *A. tomentosum* were collected from the herb garden of Liaoning University of traditional Chinese medicine (E 121° 52′, N 39° 03′) and from Urumqi, Xinjiang (E 84° 33′, N 44° 07′), both the plant fruits samples had the permission for biological experiments. Professor Tingguo Kang at Liaoning University of Traditional Chinese Medicine, identified the certificate specimens (*A. lappa* number: 10162190625306LY, *A. tomentosum* number: 10162190625307LY), plant samples were deposited in the herbarium of Liaoning University of Traditional Chinese Medicine and the genomic DNA was stored in the Key Laboratory of Traditional Chinese Medicine in the University (Dalian, China, 116600), fruits were cultured in dark under the condition of 23 °C and 60% humidity. When the seedling weight was 3–5 g, the whole plant was collected, washed, frozen with liquid nitrogen, and then stored at -80 °C for standby. We used an improved extraction method [26] for the mitochondrial DNA isolation.

**Mitochondria Dna Sequencing And Genome Assembly**

After DNA isolation, 1 µg of purified DNA was fragmented to construct short-insert libraries (insert size 430 bp) according to the manufacturer's instructions (Illumina), then sequenced on the Illumina Hiseq 4000 [27] (Shanghai BIOZERON Co., Ltd). The high molecular weight DNA was purified and used for PacBio library prep, BluePippin size selection, then sequenced on the Sequel Sequencer.
Prior to assembly, Illumina raw reads were filtered firstly. This filtering step was performed in order to remove the reads with adaptors, the reads showing a quality score below 20 (Q < 20), the reads containing a percentage of uncalled based ("N" characters) equal or greater than 10% and the duplicated sequences. The mitochondria genome was reconstructed using a combination of the Pacbio Sequel data and the Illumina Hiseq data, and the following three steps were used to assemble mitochondria genomes. First, assemble the genome framework by the both Illumina and Pacbio data using SPAdes v3.10.1 [28]. Second, verifying the assembly and completing the circle or linear characteristic of the mitochondria genome, filling gaps if there were. Third, clean reads were mapped to the assembled mitochondria genome to correct the wrong bases, judge if there is any insertion and deletion.

**Genome Annotation**

The mitochondria genes were annotated using homology alignments and denovo prediction, and the EVidenceModeler v1.1.1 [29] was used to integrate gene set. tRNA genes and rRNA genes were predicted by tRNAscan-SE [30] and rRNAmer 1.2 [31]. A whole mitochondria genome Blast [32] search (E-value ≤ 1e-5, minimal alignment length percentage ≥ 40%) was performed against 5 databases. They are KEGG [33–35] (Kyoto Encyclopedia of Genes and Genomes), COG [36–37] (Clusters of Orthologous Groups), NR (Non-Redundant Protein Database databases), Swiss-Prot [38], and GO [39] (Gene Ontology). The circular of genome map was drawn using OrganellarGenomeDRAW v1.2 [40].

**Comparative Analysis Of The Mitochondrial Genomes**

**Large repeat sequence analysis**

We used the software REPuter (http://bibiserv.techpak.uni-bielefeld.de/computer/), the minimum sequence length was 30 bp, and the editing distance was 3. Use the following four repetition methods to find the long repetition sequence: F: forward, R: reverse, C: complexity, P: palindromic.

**Simple Sequence Repeats (ssr) Analysis**

The SSR software MicroSAtellite (MISA) (http://pgrc.ipk-gatersleben.de/misa/) was used to identify SSR sequences and tandem repeats of 1–6 nucleotides were considered microsatellites. The minimum numbers of repeats were set to 10, 6, 5, 5, 5, and 5 for mono-, di-, tri-, tetra-, penta-, and hexa-nucleotides, respectively. The minimum distance between two SSR was set to 100 bp. Finally, software primer 3 (http://www.simgene.com/primer3) was designed for SSR sequence obtained from MISA recognition.

**Snp And Indel Detection**
SNP mainly refers to DNA sequence polymorphism caused by single nucleotide variation at the genome level. InDel refers to the insertion and deletion sequence of small segments in genome. In this section, in order to identify sequence variations in the known genes as well as the ORFs between \textit{A. lappa} and \textit{A. tomentosum} mitochondrial sequences, SNPs and InDels were detected. The SNP annotation results (Additional file 10: Table S10) of \textit{A. lappa} with \textit{A. tomentosum} as reference sequence showed that the SNPs in coding region had no synonymous mutation at start codon and synonymous mutation at stop codon, nor nonsynonymous mutation at start codon and nonsynonymous mutation at stop codon, and no nonsense mutation at the same time. There were 6 synonymous mutations and 13 nonsynonymous mutations in the gene. There were 32 intergenic mutations. InDel annotation results (Additional file 11: Table S11) of \textit{A. lappa} with \textit{A. tomentosum} as reference sequence showed that only 3 InDels existed between genes, and the results of gene mutation type caused by InDel with \textit{A. tomentosum} as reference sequence showed that there were no gene mutation existed.

\textbf{Syntenic sequence analysis and SV analysis between mitochondrial genomes of} \textit{A. lappa} \textit{and} \textit{A. tomentosum}

Through the syntenic sequence between the genomes, we can observe the insertion and deletion of the sequence between the genome of the target species and the reference genome. From the statistical table (Table 3) of collinear comparison coverage of five Asteraceae plants, we could find \textit{A. lappa} \textit{and} \textit{A. tomentosum} mitochondrial genome sequence alignment regions accounted for 100\% of the whole genome, and Fig. 4 illustrated that the two showed a complete collinearity. We compared the mitochondrial genomes of \textit{A. lappa} \textit{and} \textit{A. tomentosum} with those of three other Asteraceae plants, and the results showed a very high similarity. \textit{Arctium} species differed greatly from other 3 Asteraceae plants in their mitochondrial genomes, in which the percentage of regions aligned with \textit{Helianthus annuus} mitochondrial genome sequences was only 51.7\% of the whole genome, and the number of comparison blocks was 96. Additional file 14–15
Table 3
Coverage statistics of collinearity comparison of 5 Asteraceae plants

| Species                  | Total length of sequence alignment region (bp) | Total Sequence Length (bp) | Sequence alignment regions as a percentage of the entire genome (%) | Total length of sequence alignment region (bp) | Total Sequence Length (bp) | Sequence alignment regions as a percentage of the entire genome (%) | Number of comparison blocks |
|--------------------------|-----------------------------------------------|----------------------------|------------------------------------------------------------------|-----------------------------------------------|----------------------------|------------------------------------------------------------------|-----------------------------|
| A. lappa VS A. tomentosum | 312609                                       | 312609                     | 100                                                              | 312598                                       | 312598                     | 100                                                              | 1                           |
| A. lappa VS Helianthus annuus | 156224                                       | 312598                     | 49.98                                                            | 155596                                       | 300945                     | 51.7                                                             | 96                          |
| A. lappa VS Diplostephium hartwegii | 168718                                       | 312598                     | 53.97                                                            | 168492                                       | 277718                     | 60.67                                                            | 81                          |
| A. lappa VS Chrysanthemum boreale | 157655                                       | 312598                     | 50.43                                                            | 157319                                       | 211002                     | 74.56                                                            | 70                          |
| A. tomentosum VS Helianthus annuus | 156244                                       | 312609                     | 49.98                                                            | 155596                                       | 300945                     | 51.7                                                             | 96                          |
| A. tomentosum VS Diplostephium hartwegii | 168723                                       | 312609                     | 53.97                                                            | 168492                                       | 277718                     | 60.67                                                            | 81                          |
| A. tomentosum VS Chrysanthemum boreale | 157599                                       | 312609                     | 50.41                                                            | 157254                                       | 211002                     | 74.53                                                            | 71                          |

Figure S1-2 showed that the mitochondrial genomes of *Arctium* plants had lower collinearity with that of *Helianthus annuus* and slightly higher collinearity with that of *Chrysanthemum boreale* (74.56% and
74.53%). This indicated that the mitochondrial genomes of plants in different genera varied greatly even in the one Asteraceae family. The microbial genome has dense functional genes, and the occurrence of structural variations will cause the loss or alteration of multiple gene functions, resulting in changes in microbial phenotypes, functional differences and pathogenicity. For further clarifying the difference between the two mitochondrial genomes, SV were investigated between *A. lappa* and *A. tomentosum* using *A. tomentosum* as reference (Fig. 5). No translocations, inversions, translocations and deletions were found in the mitochondrial genomes of *A. lappa* and *A. tomentosum*. The results indicated that the mitochondrial genome similarity between *A. lappa* and *A. tomentosum* was very high.

**Gene transfer between chloroplast genome and mitochondrial genome in two Arctium plants**

Biomass (information) exchanges occur between subcellular units or organelles in eukaryotic cells to coordinately regulate various life activities of cells [48]. Recent studies have shown that information exchange and transfer between chloroplasts and mitochondria exists in plants and induces the occurrence of PCD (programmed cell death), but the mechanism of action has not been fully analyzed [49]. Both of the chloroplast genomes and mitochondrial genomes of *A. lappa* and *A. tomentosum* had the phenomenon of gene exchange and transfer (Fig. 6–7). Among them, 48 transfer segments with a similarity of not less than 80% we calculated had a total length of 8229 bp, the shortest transfer segments length was 45 bp, and the longest transfer segments length was 2532 bp, respectively (Additional file 12–13: Table S12-13). This phenomenon was found in other plants such as *Ginkgo biloba* and *Salvia miltiorrhiza*.

**Core, Specific and Pan Gene analysis**

The homologous genes present in all samples are regarded as common genes (Core genes), after removing the common genes, the non-common genes (Dispensable genes) are obtained, and the specific genes are the only genes that are specifically owned by the sample. All non-shared genes are merged with shared genes as the Pan genes. Among them, the core gene and the specific gene are likely to correspond to the commonality and characteristics of samples, which can be used as the basis for the study of functional differences between samples. Core genes and specific genes were analyzed for *A. lappa* and *A. tomentosum* and three other Asteraceae plants (Fig. 8). There were 354 genes and 22 core genes for these five Asteraceae plants, and 1, 2, 0, 1 and 0 specific genes for *A. lappa*, *Helianthus annuus*, *Chrysanthemum boreale*, *Diplosteophium hartwegii* and *A. tomentosum*, respectively. There were 95 Dispensable genes. Among them, the specific genes of *A. lappa* were orf115a, orf873 and rps12 for *Helianthus annuus*, and rps19 for *Diplosteophium hartwegii*. The number of Pan genes was 117. These core and specific genes were likely to correspond to the commonality and characteristics of these five plants, which can provide a basis for the study of functional differences among different species.

**Syntenic Sequence And Structural Variation Analysis**
MUMmer software was used to compare the 5 genomes, including *A. lappa*, *A. tomentosum* genomes and 3 Asteraceae plants genomes downloaded from the NCBI (*Chrysanthemum Boreale*, genbank accession number: NC039757; *diplosephium hartwegii*, genbank accession number: NC034354; *Helianthus annuus*, genbank accession number: NC023337), and the large-range collinearity relationship between genomes was determined. Later, LASTZ was used to compare the regions, confirm the local position arrangement relationship, and find the regions of translocation (Translocation/Trans), inversion (Inversion/Inv), and translocation + inversion (Trans + Inv).

The mitochondrial genomes of *A. lappa* and *A. tomentosum* were compared using MUMmer software, and then the regions were compared using LASTZ to find the SV from the regional comparison results.

**Gene transfer between chloroplast genome and mitochondrial genome**

The chloroplasts of *A. lappa* and *A. tomentosum* were compared with their mitochondrial genomes by BLSTN, respectively. The selected parameter E value was less than $1 \times 10^{-10}$.

**Core, Specific and Pan Gene analysis**

Using cd-hit (v4.6.1, http://cd-hit.org) software, the protein sequences of multiple samples to be analyzed were clustered, and the screening parameters for Identity and alignment length (requiring identity > 50% and coverage > 50% of clustering) were set. The clustering of all protein sequences was obtained according to the results of software analysis. Core and Pan gene sets were constructed by comparing the protein sequences of *A. lappa*, *A. tomentosum* genomes and 3 Asteraceae plants genomes.

**Phylogenetic Analysis**

In order to determine the phylogenetic position between *Arctium* and other genera in Asteraceae, phylogenetic tree was constructed based on the whole mitochondrial genomes of 28 species, including *A. lappa* and *A. tomentosum*, 25 species of the other genera in Asteraceae and other Asterids plants, 1 species of the outgroup (*Ginkgo biloba*). All GenBank accession numbers were listed (Additional file 1: Table S1). The PhyML V3.0 software was used to construct a phylogenetic tree by maximum likelihood method (ML), bayes correction, 1000 bootstrap replicates to calculate bootstrap values [41].

**Results**

**Sequencing and assembly of the mitochondrial genome in *A. lappa* and *A. tomentosum***
Mitochondrial genomes of *A. tomentosum* and *A. lappa* were sequenced via Illumina PE and PacBio techniques. For the Illumina PE sequencing, 6363 Mb and 6850 Mb raw data; 5706 Mb and 6292 Mb clean data were obtained from *A. lappa* and *A. tomentosum*, respectively. The Q20 values for *A. lappa* and *A. tomentosum* were 98.04% and 97.98%, The Q30 values for *A. lappa* and *A. tomentosum* were 93.87% and 93.73%, respectively. In regard to the PacBio sequencing, subreads with N50 = 9836 bp and subreads with N50 = 9865 bp were generated from *A. lappa* and *A. tomentosum*, respectively. Both of the largest length of subreads were 26635 bp. Through de novo assembling, mitochondrial genomes of *A. lappa* and *A. tomentosum* were obtained, of which the first one was 312598 bp and second one was 312609 bp. The GC contents of the mitochondrial genomes of *A. lappa* and *A. tomentosum* were 45.43% and 45.44%, respectively. (Additional file 2–3: Table S2-3) And we made the data public and uploaded it to the GenBank nucleotide sequence database. Accession number: SRX7565441 (Illumina), SRX7565442 (pacbio) for *Arctium lappa*, SRX7565439 (Illumina), SRX7565440 (pacbio) for *Arctium tomentosum*.

**Annotation and general characteristics of *A. lappa* and *A. tomentosum* mitochondrial genome**

After the annotation, the master circles images were generated from the software OrganellarGenomeDRAW v1.2 [40], and 131 genes including 98 ORFs, 6 ribosomal RNAs and 25 transfer RNAs were identified in *A. lappa*. 130 genes including 97 ORFs, 6 ribosomal RNAs and 25 transfer RNAs were identified in *A. tomentosum*. (Fig. 1.2) Compared with *A. lappa*, the number of protein coding sequences in *A. tomentosum* annotated in NR and GO databases was more than 1, and the number of other databases annotated in *A. lappa* was the same. The mitochondrial genomes of *A. lappa* and *A. tomentosum* were assembled into 2 single circular molecules of 312598 bp and 312609 bp, respectively (Table 1). From the results, *A. lappa* had only one more gene than *A. tomentosum*. As to the known genes in *A. lappa* and *A. tomentosum* mitochondrial genomes, there was one-to-one correspondence with each other [3]. The ORF encoding proteins that were equal to or larger than 100 aas were identified from the assembled mitochondrial genomes. ORFs and other component of the two genomes were listed in the Additional file 4: Table S4. There were 9 introns in the mitochondrial protein coding genes of *A. lappa* and 21 introns in the whole mitochondrial genome in total. The numbers of that in *A. tomentosum* were 8 and 20. Among them, *rps3* gene of *A. lappa* contained 1 intron, while the gene of *A. tomentosum* had no introns. The number of introns and exons in other two genes was the same, and the length was the same or similar. *nad7*, *nad1* and *nad2* had 4 introns and 5 exons (Additional file 5: Table S5).
Table 1
Part of the gene component and general characteristics of the mitochondrial genomes of *A. lappa* and *A. tomentosum*.

| Gene category                          | *A. lappa*                                      | *A. tomentosum*                               |
|---------------------------------------|------------------------------------------------|-----------------------------------------------|
| complex I (NADH dehydrogenase)        | \textit{nad1}\textit{2}\textit{3}\textit{4}\textit{4L}\textit{5}\textit{6}\textit{6-D2}\textit{7}\textit{9} | \textit{nad1}\textit{2}\textit{3}\textit{4}\textit{4L}\textit{5}\textit{6}\textit{6-D2}\textit{7}\textit{9} |
| complex II (succinate dehydrogenase)  | \textit{sdh4}                                  | \textit{sdh4}                                 |
| complex III (ubichinol cytochrome c reductase) | \textit{cob}                                  | \textit{cob}                                 |
| complex IV (cytochrome c oxidase)     | \textit{cox1}\textit{2}\textit{3}             | \textit{cox1}\textit{2}\textit{3}            |
| ATP synthase                          | \textit{atp1}\textit{4}\textit{6}\textit{8}\textit{9} | \textit{atp1}\textit{4}\textit{6}\textit{8}\textit{9} |
| ribosomal proteins (SSU)              | \textit{rps12}\textit{rps13}\textit{rps14}\textit{rps3}\textit{rps4} | \textit{rps12}\textit{rps13}\textit{rps14}\textit{rps3}\textit{rps4} |
| ribosomal proteins (LSU)              | \textit{rpl10}\textit{rpl5}                   | \textit{rpl10}\textit{rpl5}                  |
| maturases                              | \textit{matR}                                  | \textit{matR}                                 |
| other genes                           | \textit{ccmB}\textit{ccmC}\textit{ccmFc}\textit{ccmFn}\textit{mttB} | \textit{ccmB}\textit{ccmC}\textit{ccmFc}\textit{ccmFn}\textit{mttB} |
| ORF number                            | 98                                             | 97                                            |
| transfer RNA number                   | 25                                             | 25                                            |
| ribosomal RNA number                  | 6                                              | 6                                             |
| Genome size (bp)                      | 312598                                         | 312609                                        |
| GC Content (%)                        | 45.43                                          | 45.44                                        |
| Gene number                           | 131                                            | 130                                           |
| Gene total length (bp)                | 81252                                          | 80865                                        |
| Gene length / Genome (%)              | 25.99                                          | 25.87                                        |

Comparative analysis of the mitochondrial genomes

Repeat sequence analysis

Large repeat analysis
Many repeats are present in gene deserts, although whole-genome sequencing has shown that they can occur in functional regions as well [42–43]. Repeat of more than 30 bases were considered as the large repeats. From the results (Additional file 6–7: Table S6-7), 50 pairs of large repeat sequences were detected in *A. lappa* and *A. tomentosum*, 49 of them in non-coding region and 51 of them in coding region, of which 36-21386 bp was the length of *A. lappa* long repeat sequence, with a total length of 44089 bp. The length of *A. tomentosum* large repeat was 36-21393 bp, and the total length was 44006 bp. There were two types of large repeat in *A. lappa* and *A. tomentosum*: Forward and Reverse.

**SSR analysis**

SSR is a PCR-based highly efficient molecular labeling technique [44]. Single parental inheritance, together with other characteristics, has been extensively used in species identification and genetic diversity analysis. The content of SSR in eukaryote genome is very rich, and it is often randomly distributed in nuclear DNA. SSR is also abundant in plants and evenly distributed in the whole plant genome, but the frequency of SSR in different plants varies greatly. Because of the characteristics of neutral markers, the highly variable numbers of repeats and the relative conservatism of flanking sequences of SSRs, it is widely distributed in the genome of organisms. The technique is easy to operate and has high repeatability and codominant inheritance among alleles. SSRs marker technique is the best choice for evaluating genetic diversity of crops [45–47]. In this analysis, the minimum distance between two SSRs was set to 100 bp. From the Table 2, SSRs of mitochondrial genomes of *A. lappa* and *A. tomentosum* had exactly the same types and numbers, and both of them had 5 types. Among them, the single nucleotide A/T SSR was the most, 111, but only one was longer than 20 bp. The second largest number is the TetraNucl SSR, 34. The minimum number of SSRs is the Pentanucl, only 3. The total length of SSR was 2491 bp for *A. lappa* and 2489 bp for *A. tomentosum* (Additional file 8–9: Table S8-9). From the SSR comparison of mitochondrial genomes of three other Asteraceae plants (Fig. 3) (*Chrysanthemum Boreale*, *Diplosephium hartwegii* and *Helianthus annuus*), the mitochondrial genomes of *A. lappa*, *A. tomentosum*, *Helianthus annuus*, *Diplosephium hartwegii*, *Chrysanthemum Boreale*, contained 192, 192, 134, 151, 127 SSRs, respectively. We could intuitively find that the proportion of single nucleotide SSR in mitochondrial genome of these five plants was the largest, followed by that of TetraNucl type SSR. Compared with the other three Asteraceae plants, *A. lappa* and *A. tomentosum* had no HexaNucl SSR, and the proportion of this SSR in the other three plants was smaller. It could be seen that the mitochondrial genome SSR types of the five Asteraceae plants were similar, and the SSR types of the same genus were more analogous.
Table 2
SSRs identified in the mitochondrial genomes of *A. lappa* and *A. tomentosum* and 3 other Asteraceae plants

| Length | Type    | A. lappa | A. tomentosum | Helianthus annuus | Diplosephium hartwegii | Chrysanthemum Boreale |
|--------|---------|----------|---------------|-------------------|------------------------|----------------------|
|        | Amount  | Amount(> = 20 bp) | Amount  | Amount(> = 20 bp) | Amount  | Amount(> = 20 bp) | Amount  | Amount(> = 20 bp) |
| 1      | A/T     | 111      | 1            | 111               | 1         | 53                  | 1       | 82                  | 0    |
|        | C/G     | 23       | 0            | 23                | 0         | 15                  | 0       | 10                  | 0 |
| 2      | AG/CT   | 11       | 0            | 11                | 0         | 9                   | 0       | 11                  | 0 |
|        | AT/AT   | 6        | 1            | 6                 | 1         | 8                   | 0       | 8                   | 0 |
| 3      | AA/G/C  | 4        | 0            | 4                 | 0         | 3                   | 0       | 5                   | 0 |
|        | TT      |          |              |                   |           |                     |         |                     |     |
|        | AA      | 0        | 0            | 1                 | 0         | 0                   | 0       |                     |     |
|        | C/G     |          |              |                   |           |                     |         |                     |     |
|        | TT      |          |              |                   |           |                     |         |                     |     |
|        | AAT     | 0        | 0            | 5                 | 0         | 0                   | 0       |                     |     |
|        | /AT     |          |              |                   |           |                     |         |                     |     |
|        | AC      | 0        | 0            | 2                 | 0         | 0                   | 0       |                     |     |
|        | T/AG    |          |              |                   |           |                     |         |                     |     |
|        | AG      | 0        | 0            | 2                 | 0         | 0                   | 0       |                     |     |
|        | C/C     |          |              |                   |           |                     |         |                     |     |
| 4      | Tetr     | 34       | 0            | 34                | 0         | 29                  | 0       | 0                   | 21  |
|        | anucleotide |      |              |                   |           |                     |         |                     |     |
| 5      | Pent     | 3        | 2            | 3                 | 2         | 5                   | 0       | 1                   | 0 |
|        | anucleotide |        |              |                   |           |                     |         |                     |     |
|        |          |          |              |                   |           |                     |         |                     |     |
### Phylogenetic Analysis

In this study, in order to determine the evolutionary position of the *A. lappa* and *A. tomentosum*, phylogenetic tree was constructed based on the mitochondrial genomes of 28 species, including 25 Asterids plants of other genera and 1 species of the outgroup (*Ginkgo biloba*). In the phylogenetic tree, 10 of 25 nodes exhibited bootstrap values of 100% and eight had bootstrap values ≥ 90%. The *A. lappa* and *A. tomentosum* species clustered closely together into one branch, which had a 99% bootstrap values. This result was consistent with earlier phylogenetic analyses based on the chloroplast DNA of the two samples [20], *Diplostephium hartwegii*, *Chrysanthemum boreale* and *Helianthus annuus* were clustered in one clade and the first two species had a closer relationship, implying that they are very closely related. At the same time, they were clustered in one large clade with *A. lappa* and *A. tomentosum*. In addition, other closely related species had similar evolutionary locations. Three species of *Nicotiana*, two species of *Solanum* were also clustered into one branch. *Bupleurum falcatum* and *Daucus carota*, two species belonging to the Umbelliferae family, did not cluster together. *Bupleurum falcatum* and *Vaccinium macrocarpon* clustered together, but the bootstrap value was only 79, while *Daucus carota* clustered with Asteraceae. Such data will provide certain help to subsequent botany for classifying plants.

### Discussion

**General characteristics of *A. lappa* and *A. tomentosum* mitochondrial genome**

By measuring the mitochondrial genomes of *A. lappa* and *A. tomentosum*, we found that the mitochondrial genome structures of the two species were very similar. The mitochondrial genome length of the two species was more than 30 kb, which was longer than that of the previously reported Asteraceae plants, *Chrysanthemum boreale* (211002 bp), *Helianthus annuus* (300104 bp) and *Diplostephium hartwegii* (277718 bp) [11], less than the medicinal plant *Salvia miltiorrhiza* (499236 bp) [50] etc, the mitochondrial genome lengths of plants in different families and genera differed greatly, but the mitochondrial genome sizes of the same genera (*Arctium*) differed little from those of *A. lappa* and *A.
tomentosum. In addition, the mitochondrial genome structure of A. lappa and A. tomentosum was very similar, including GC content, the composition of coding genes, and the composition of RNA. There was only one ORF gene difference between the two species. The similarity of mitochondrial genomes in tomatoes and their wild varieties could reach to 98%, but the mitochondrial genomes of 138A male sterile line and its maintainer line 138B in pepper were quite different [3]. This might be related to the low rate of variation in the mitochondrial genome and the occurrence of gene recombination and rearrangement [11], which we might also indicate that A. lappa and A. tomentosum originate from the same maternal line and later evolve into different two species. In addition, the annotated genes of A. lappa and A. tomentosum were very similar, A. lappa had only one orf115a gene than that in A. tomentosum, in addition, the conserved genes of both were similar to those of Lactuca plants, A. lappa and A. tomentosum lacked rpl16 gene, had one more gene, atp8, and they all had pseudogene sdh4.

The Open Reading Frame (ORF) starts at the start codon, is a sequence in DNA sequence with the potential to encode proteins, and ends at a sequence of bases that is contiguous with the stop codon. A. lappa and A. tomentosum were annotated to 98 and 97 ORF genes, respectively, which differed greatly from the number of some reported plants, such as melon with only 2 [51], these genes were the basis of our future research on the two Arctium plants. In addition, A. lappa and A. tomentosum had 25% GC content, which was lower than that of other plants, such as the mitochondrial genome GC content of Saussurea involucrata (29%) and Ginkgo biloba mitochondrial genome GC content (50.36%) [52–53], which might be related to the stability of the mitochondrial DNA. Introns in eukaryotes could be used to study phylogenetic evolution, evolutionary distance and gene expression regulation [54]. Introns-containing genes had been found in the mitochondrial genes of A. lappa and A. tomentosum. The main difference between them was that the rps3 gene, A. lappa had introns inside, while A. tomentosum genome did not have introns inside the rps3. This gene also had one intron in Ginkgo biloba, and the gene nad1 also had four introns [53], suggested that mitochondrial genes might lose introns during the evolution, and this phenomenon could be used to study the evolutionary relationship of species, as well as classification and identification.

Comparative analysis of the mitochondrial genomes

The presence of a large number of repetitive sequences in the mitochondrial genome could promote the occurrence of intramolecular sequence recombination. As a result, isomeric genomic sequences in mitochondria and sub-genomic sequences derived from some small circular molecular structures were generated [55]. SSRs sequences were abundant in plants with strong repeatability, simple genome structure, and relatively conserved, which were widely used in species identification and genetic analysis of individuals and populations [56–57]. As an example, some researchers had used mitochondrial SSR to study the genetic diversity of Salvia miltiorrhiza from different origins, and the research showed that Salvia miltiorrhiza from different origins had genetic diversity, and the average versatility was high [58]. Both A. lappa and A. tomentosum mitochondrial genomes contained a large number of SSRs, which could be used to analyze their genetic diversity and lay a foundation for their quality research. The mitochondrial genome repeats of A. lappa and A. tomentosum were basically identical, which indicated
that the mitochondrial genome could remain relatively static for a long time, but could also show diversity due to rearrangement, sequence loss and unknown sequence [11]. Therefore, the analysis of SNP (single nucleotide polymorphism) and InDel of *A. lappa* mitochondrial genome with the mitochondrial genome of *A. tomentosum* as reference could reveal the diversity of their mitochondrial genomes. SNPs were highly stable, while SNPs located in coding regions were more stable [59], which could be used to analyze the evolutionary relationship of species as well as identification and typing between or within species. SNPs existed between *A. lappa* and *A. tomentosum*, and we could use these SNPs to identify the interspecies between them. There were only 3 small fragments of InDel in *A. lappa* and *A. tomentosum*, and both of them were in the *nad1* gene. This was consistent with the fact that a large number of small fragments of SNPs had been found in both sterile and fertile lines of *Hibiscus cannabinus*, and most of them were different between genes [60], indicated that the differences of InDel among species of *Arctium* were relatively small.

**Conclusions**

The mitochondrial genomes of *A. lappa* and *A. tomentosum* were similar in size, 312598 bp and 312609 bp, respectively, and the number of annotated genes was 131 and 130, respectively. The ORF genes of the two were different, which could provide a theoretical basis for the development of molecular markers. Both mitochondrial genomes contained a certain number of SSRs and large repeat sequences, which could provide a basis for future genetic diversity and identification analysis of both. The mitochondrial genomes of the two species had gene exchange and transfer with their own chloroplast genomes, respectively. In addition, the mitochondrial genomes of the two species showed complete collinearity, but they differed greatly from the mitochondrial genomes of three plants of different genera in the same family, with translocation, inversion, translocation and inversion. The evolutionary position of *A. lappa* and *A. tomentosum* in evolutionary analysis was the same as that of traditional classification. In future studies, we can search for related genes of *Arctium* plants molecular identification barcode and molecular breeding based on their mitochondrial genomic characteristics.

**Abbreviations**

TCM
traditional Chinese Medicine
PCR
Polymerase chain reaction
ORF
Open reading frame
SNP
Single nucleotide polymorphism
tRNA
transfer RNA
Declarations

Acknowledgments

We thank the Shanghai BIOZERON Biotechnology Co., Ltd. for processing the raw sequencing data.

Authors’ contributions

LX and TK designed the research and conceptualization. DZ, YX and RZ conducted experiments. TZ, GB, TL, TZ, WA and DZ did the software job and analysis, DZ and YY wrote the paper. TK did the project administration job. All authors read and approved the final manuscript.

Funding

The study was supported by National Natural Science Foundation of China (General Program, Grant Numbers 81874338, 81773852), the sequencing costs were funded by 2019 Liaoning Provincial Department of Education Scientific Research Project, China (No. L201942), National Key Research and Development Plan, China (No. 2018YFC1708200). The publication costs are funded by Major special fund for science and technology of Inner Mongolia Autonomous Region (No. 2019ZD004).

Availability of data and materials

Sequence data from this article have been submitted to the GenBank Data Libraries under accession numbers SRX7565441 (Illumina), SRX7565442 (pacbio) for *Arctium lappa*, SRX7565439 (Illumina), SRX7565440 (pacbio) for *Arctium tomentosum*, respectively.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.
References

1. Janska H, Woloszynska M. The dynamic nature of plant mitochondrial genome organization. Acta BiochimPol. 1997;44:239 – 50.

2. Oda K, Kohchi T, Ohyama K. Mitochondrial-DNA of Marchantia Polymorpha as a single circular form with no Incorporation of foreign DNA. Biosci Biotechnol Biochem. 1992;56: 132 – 35.

3. Wang P. Lu Q, Ai Y. Wang Y, Li T. Wu L, Liu J. Cheng Q, Sun L. Shen H. Candidate Gene Selection for Cytoplasmic Male Sterility in Pepper (Capsicum annuum L.) through Whole Mitochondrial Genome Sequencing. Int J Mol Sci. 2019;20:578.

4. Chen G. Ye X, Zeng F. Wang J, Yuan L. Zhu S, Hou J. Cheng Y, Wang C. Characterization and utilization of a cytoplasmic male sterility line of Wucai (Brassica campestris L.). Horticulture Environment Biotechnology. 2019;60:373 – 82.

5. Li J. Nadeem M, Sun G. Wang X, Qiu L. Male sterility in soybean: Occurrence, molecular basis and utilization. Plant Breeding. 2019;00:1–18.

6. Khaimongkhon S. Chakhonkaen S, Pitngam K. Ditthab K, Sangarwut N. Panyawut N, Wasinanon T. Mongkolsiriwatana C, Chunwongse J. Muangporm A. Molecular Markers and Candidate Genes for Thermo-Sensitive Genic Male Sterile in Rice. Rice Sci. 2019; 26:147 – 56.

7. Shu J. Liu Y, Zhang L. Li Z, Fang Z. Yang L, Zhuang M. Zhang Y, Lv H. Evaluation and selection of sources of cytoplasmic male sterility in broccoli. Euphytica. 2019;215:125.

8. Yamagishi H. Tanaka Y, Shiiba S. Hashimoto A, Fukunaga A. Terachi T. Mitochondrial orf463 causing male sterility in radish is possessed by cultivars belonging to the 'Niger' group. Euphytica. 2019;215:109.

9. Kubo T, Newton K J. Angiosperm mitochondrial genomes and mutations. Mitochondrion. 2008;8:5–14.

10. Makarenko MS, Usatov AV, Tatarinova TV, Azarin KV, Logacheva MD, GavriloVA VA, Horn R. Characterization of the mitochondrial genome of the MAX1 type of cytoplasmic male-sterile sunflower. BMC Plant Biol. 2019;19:41–7.
11. Wynn EL, Christensen AC. Repeats of Unusual Size in Plant Mitochondrial Genomes: Identification, Incidence and Evolution. G3 Genesgenetics. 2019;9:549–59.

12. 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;31:5907–16.

13. Ogihara Y, Yamazaki Y, Murai K, Kanno A, Terachi T, Shiina T, Miyashita N, Nasuda S, Nakamura C, Mori N, et al. Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. Nucleic Acids Res. 2005;33:6235–50.

14. Palmer JD. Mitochondrial. DNA in Plant Systematics: Applications and Limitations. New York: Springer US; 1992. pp. 36–49.

15. Levings CS. Brown GG. Molecular biology of plant mitochondria. Cell. 1989;56:171 – 79.

16. Adams KL. Qiu YL, Stoutemyer M. Palmer JD. Inaugural Article: Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc Natl Acad Sci USA. 2002;99:9905-12.

17. Knoop V. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. Curr Genet. 2004;46:123 – 39.

18. Fujii S. Kazama T, Yamada M. Toriyama K. Discovery of global genomic re-organization based on comparison of two newly sequenced rice mitochondrial genomes with cytoplasmic male sterility-related genes. BMC Genom. 2010;11:209 – 10.

19. Lu JN, Zhao ZL, Ni LH, Ga W, Mi M. The identification of Sect. Crucia (Gentiana) species using mtDNA nad1/b-c and nad5/d-e fragments. Acta Pharmaceutica Sinica. 2019;54:166–72.

20. Xing YP, Xu L, Chen SY, Liang YM, Wang JH, Liu CS, Liu T, Kang TG. Comparative analysis of complete chloroplast genomes sequences of Arctium lappa and A. tomentosum. Biol Plant. 2019;63:565 – 74.

21. Chan YS, Cheng LN, Wu JH, Chan E, Kwan YW, Lee SMW, Leung GPH, Yu PHF, Chan SW. A review of the pharmacological effects of Arctium lappa (burdock). Inflammopharmacology. 2011;19:245 – 54.

22. China Pharmacopoeia Committee. Pharmacopoeia of People’s Republic of China. China Medical Science Press. 2015; Vol.1:72.

23. Tardio J, Pascual H, Morales R. Wild food plants traditionally used in the province of Madrid. Economic Botany. 2005;59:122 – 36.

24. Yaqi B, Tugume AK, Valkonen JPT. Small-RNA Deep Sequencing Reveals Arctium tomentosum as a Natural Host of Alstroemeria virus X and a New Putative Emaravirus. PLoS One. 2012;7:e42758.

25. Hayashi K, Narutaki K, Nagaoka Y, Hayashi T, Uesato S. Therapeutic effect of arctiiin and arctigenin in immunocompetent and immunocompromised mice infected with influenza A virus. Biol Pharm Bull. 2010;33:1199–205.

26. Chen J, Guan R, Chang S, Du T, Zhang H, Xing H. Substoichiometrically different mitotypes coexist in mitochondrial genomes of Brassica napus L. PLoS One. 2011;6:e17662.
27. Borgstrom E, Lundin S, Lundeberg J. Large scale library generation for high throughput sequencing. PLoS One. 2011;6:e19119.

28. Dmitry A, Anton K, Jeffrey SM, Pavel AP. HYBRIDSPADES: an algorithm for hybrid assembly of short and long reads. Bioinformatics. 2016;32:7.

29. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 2008;9:R7.

30. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic acids research. 1997;25:955–64.

31. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery D. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic acids research. 2007;35:3100–8.

32. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of molecular biology. 1990;215:403–10.

33. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. Nucleic acids research. 2004;32:D277-80.

34. Kanehisa M. A database for post-genome analysis. Trends Genet. 1997;13:375.

35. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. From genomics to chemical genomics: new developments in KEGG. Nucleic acids research. 2006;34:D354-7.

36. Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. Science. 1997;278:631–7.

37. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, et al. The COG database: an updated version includes eukaryotes. BMC Bioinformatics. 2003;4:41.

38. Magrane M, Mercier PL. UniProt Knowledgebase: a hub of integrated protein data. Database: The Journal of Biological Databases & Curation. 2011;bar009.

39. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM. Gene Ontology: tool for the unification of biology. Nat Genet. 2000;25:25–9.

40. 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. 2007;52:267–74.

41. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59:307–21.

42. Duitama J, Zablotskaya A, Gemayel R, Jansen A, Belet S, Vermeesch JR, Froyen G. Large-scale analysis of tandem repeat variability in the human genome. Nucleic Acids Res. 2014;42:5728–41.
43. Zhang TT, Xing YP, Xu L, Bao GH, Zhan ZL, Yang YY, Wang JH, Li SN, Zhang DC, Kang TG. Comparative analysis of the complete chloroplast genome sequences of six species of *Pulsatilla* Miller, Ranunculaceae. Chin Med. 2019;14:53.

44. Akkaya MS, Shoemaker RC, Specht JE, Bhagwat AA, Cregan PB. Integration of Simple Sequence Repeat DNA Markers into a Soybean Linkage Map. Crop Sci. 1995;35:1439–45.

45. Bindler G, van der Hoeven R, Gunduz I, Plieske J, Ganal M, Rossi L, Gadani F, Donini P. A microsatellite marker based linkage map of tobacco. Theor Appl Genet. 2007;114:341–9.

46. Rajput SG, Santra DK. Evaluation of Genetic Diversity of Proso Millet Germplasm Available in the United States using Simple-Sequence Repeat Markers. Crop Sci. 2016;56:2401–9.

47. Li LJ, Yang KC, Pan GT, Rong TZ. Genetic Diversity of Maize Populations Developed by Two Kinds of Recurrent Selection Methods Investigated with SSR Markers. Agr. Sci. China. 2008;7:1037-45.

48. Schuldiner M, Guo W. Editorial overview: cell organelles: organelle communication: new means and new views. Curr Opin Cell Biol. 2015;35:v–vi.

49. Zhao YN, Luo LL, Xu JS, Xin PY, Guo HY, Wu J, Bai L, Wang GD, Chu JF, Zuo JR, et al. Malate transported from chloroplast to mitochondrion triggers production of ROS and PCD in *Arabidopsis thaliana*. Cell Res. 2018;28:448–61.

50. Wu B. Chen H, Shao J. Identification of Symmetrical RNA Editing Events in the Mitochondria of *Salvia miltiorrhiza* by Strand-specific RNA Sequencing. Sci Rep. 2017;7:42250.

51. Celik Altunoglu Y, Baloglu MC, Baloglu P. Genome-wide identification and comparative expression analysis of LEA genes in watermelon and melon genomes. Physiology Molecular Biology of Plants. 2017;23:5–21.

52. Li J, Yan PY, Qian FJ, Qiu BY, Xia WW, Mu XW, Qiu LJ, Lin ZP, Chen M, Zhu JB, et al. The Sasusserd involucrata Transcriptome Knowledge Base. Chinese Bulletin of Botany. 2017;52:530–8.

53. Guo WH, Felix G, Fan WS, Gregory JY, Volker K, Jeffrey DP, Jeffrey PM. Ginkgo and Welwitschia Mitogenomes Reveal Extreme Contrasts in Gymnosperm Mitochondrial Evolution. Mol Biol Evol. 2016;6:6.

54. Cao J. Research Progress in Eukaryotic Intron. Bioprocess. 2011;1:9–12.

55. Fauron C. Casper M, Gao Y. Moore B. The maize mitochondrial genome: dynamic, yet functional. Trends Genet. 1995;11:228 – 35.

56. Choi KS, Park S. The complete chloroplast genome sequence of *Aster spathulifolius* (Asteraceae); genomic features and relationship with Asteraceae. Gene. 2015;572:214–21.

57. Khakhlova O, Bock R. Elimination of deleterious mutations in plastid genomes by gene conversion. Plant J. 2006;46:85–94.

58. Sha XF, Peng F, Tao S, Wu Y, Liu JM, Zhang C, Li Q. Development and diversity analysis of mitochondrial and chloroplast microsatellite markers in *Salvia miltiorrhiza*. Acta Bot Boreal Occident Sin. 2018;38:2215-23.
59. Liu GZ, Gong ZH, Huang W, Du Y, Ji JJ. Acquisition of different gene of pepper Cytoplasmic Male Sterile and maintainer line and Single Nucleotide Polymorphism analysis. Journal of China Agricultural University. 2010;15:19–24.

60. Liao XF, Diao Y, Qiu AH, Zhao YH, Zhou BJ, Chen P, Zhou RY. Cloning and expression analysis of cox1 incytoplasmic male sterility line UG93A andits maintainer line UG93B of kenaf. Journal of China Agricultural University. 2016;21:36–45.

Figures
Figure 1

Mitochondrial genome map of A. lappa.
Figure 2

Mitochondrial genome map of A. tomentosum, the outer circle was the position coordinates of the genome components such as genes and non-coding RNAs, and had corresponding gene names; the inner circle was the genomic GC content.
**Figure 3**

SSR characteristics of mitochondrial genomes of five Asteraceae plants, MonoNucl represents mononucleotide repeats, DiNucl represents dinucleotide repeats, TriNucl represents trinucleotide repeats, TetraNucl represents tetranucleotide repeats, PentaNucl represents pentanucleotide repeats, and HexaNucl represents hexanucleotide repeats.
Figure 4

Collinear relationship between the mitochondrial genomes of A. lappa and A. tomentosum. The upper and lower bars in each figure represented the mitochondrial genomes. Dark-orange and blue regions in each bar represented the forward and reverse direction of the aligned genome, respectively. White regions in each bar represent the sequences that could not be aligned to the other genome. Lines between the two bars indicated the syntenic types and locations: magenta, blue-green, dark yellow, and light-green represent for collinear, translocation, inversion, and tran + inver, respectively.
Figure 5

Structural variation map of mitochondrial genomes of A. lappa and A. tomentosum. The inner circle is the A. tomentosum genome and the outer circle is the A. lappa genome. Collinear: homologous region; Translocation: translocation region; Inversion: inverted region; Tran+Inver: translocation and inversion region; Insertion: insertion region with length greater than or equal to 50 bp; Deletion: deletion region with length greater than or equal to 50 bp; ComplexInDel: region without alignment but corresponding to location; Forward_chain: forward chain of genomic sequence, at which time the gene sits Markers increase clockwise; Reverse_chain: the reverse chain of a genomic sequence, where the gene coordinates increase counterclockwise
Figure 6

Gene exchange in chloroplast genome and mitochondrial genome of A. lappa, the color of the connecting line is selected according to ratio = score/max, and the score is differently colored in the corresponding range blue <= 0.25, green <= 0.50, orange <= 0.75, red > 0.75. The warmer the color, the higher the similarity. chl is the abbreviation of chloroplast genome and mito is the abbreviation of mitochondrial genome.
Figure 7

Gene exchange in chloroplast genome and mitochondrial genome of A. tomentosum, the color of the connecting line is selected according to ratio = score/max, and the score is differently colored in the corresponding range blue <= 0.25, green <= 0.50, orange <= 0.75, red > 0.75. The warmer the color, the higher the similarity. chl is the abbreviation of chloroplast genome and mito is the abbreviation of mitochondrial genome.
Figure 8

Core, Specific and Pan genes of 5 Asteraceae plants
Figure 9

Molecular phylogenetic tree of 28 species based on the whole mitochondrial genomes. The tree was constructed via a maximum likelihood analysis using PhyML v3.0 with 1000 bootstrap replications.
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additonalfile1.docx
- Additonalfile3.docx
- Additonalfile2.docx
- Additonalfile4.docx
- Additonalfile5.docx
- Additonalfile6.docx
- Additonalfile7.docx
- Additonalfile8.docx
- Additonalfile9.docx
- Additonalfile10.docx
- Additonalfile11.docx
- Additonalfile12.docx
- Additonalfile13.docx
- Additonalfile14.docx
- Additonalfile15.docx