Article

Complete Chloroplast Genome Sequence and Phylogenetic Analysis of *Aster tataricus*

Xiaofeng Shen [1,†], Shuai Guo [2,†], Yu Yin [3], Jingjing Zhang [4], Xianmei Yin [5], Conglian Liang [6], Zhangwei Wang [7], Bingfeng Huang [7], Yanhong Liu [7], Shuiming Xiao [1,8,*] and Guangwei Zhu [1,8,*]

1   Artemisinin Research Center, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China; xfshen1030@163.com
2   College of Agricultural (College of Tree Peony), Henan University of Science and Technology, Luoyang 471023, China; shuai93guo@163.com
3   School of Computer Science and Technology, Shandong University, Jinan 250101, China; 18384253631@163.com
4   College of Pharmacy, Hubei University of Chinese Medicine, Wuhan 430065, China; zhangjingjing851@126.com
5   College Pharmacy, Chengdu University of Chinese Medicine, Chengdu 611137, China; xianmeiyinCM@163.com
6   College of Pharmacy, Shandong University of Traditional Chinese Medicine, Jinan 250355, China; conglian_liang@163.com
7   State Key Laboratory of Innovative Natural Medicine and TCM Injections, Ganzhou 341000, China; jxwzw1@qfyy.com.cn (Z.W.); jxhbf1@qfyy.com.cn (B.H.); jxlyh1@qfyy.com.cn (Y.L.)
8   State Key Laboratory of Innovative Natural Medicine and TCM Injections, Jiangxi qingfeng pharmaceutical co. LTD, Ganzhou 341000, China
*   Correspondence: smxiao@icmm.ac.cn (S.X.); gwzhu@icmm.ac.cn (G.Z.)
†   These authors contributed equally to this work.

check for updates

Abstract: We sequenced and analyzed the complete chloroplast genome of *Aster tataricus* (family Asteraceae), a Chinese herb used medicinally to relieve coughs and reduce sputum. The *A. tataricus* chloroplast genome was 152,992 bp in size, and harbored a pair of inverted repeat regions (IRa and IRb, each 24,850 bp) divided into a large single-copy (LSC, 84,698 bp) and a small single-copy (SSC, 18,250 bp) region. Our annotation revealed that the *A. tataricus* chloroplast genome contained 115 genes, including 81 protein-coding genes, 4 ribosomal RNA genes, and 30 transfer RNA genes. In addition, 70 simple sequence repeats (SSRs) were detected in the *A. tataricus* chloroplast genome, including mononucleotides (36), dinucleotides (1), trinucleotides (23), tetranucleotides (1), pentanucleotides (8), and hexanucleotides (1). Comparative chloroplast genome analysis of three *Aster* species indicated that a higher similarity was preserved in the IR regions than in the LSC and SSC regions, and that the differences in the degree of preservation were slighter between *A. tataricus* and *A. altaicus* than between *A. tataricus* and *A. spathulifolius*. Phylogenetic analysis revealed that *A. tataricus* was more closely related to *A. altaicus* than to *A. spathulifolius*. Our findings offer valuable information for future research on *Aster* species identification and selective breeding.

Keywords: *Aster tataricus*; chloroplast genome; phylogeny

1. Introduction

Aster tataricus is a tall perennial herb of the genus *Aster* (family Asteraceae). It has been used as a therapeutic traditional medicine for eliminating phlegm and relieving cough for thousands of years [1,2], and cultivated as a high-value medicinal plant. A number of bioactive compounds have

been isolated from *A. tataricus*, such as shionone, epifriedelinol, quercetin, emodin, caffeoylquinic acid, kaempferol, and some other triterpenes or saponins [3,4]. Modern pharmacological studies [5–7] have shown that *A. tataricus* exhibits diverse pharmacological effects, including antibacterial, antiviral, antitumor, and diuretic activities; consequently, it is recorded as a basic clinical herb in the Chinese Pharmacopoeia. Furthermore, as a late-blooming aster, *A. tataricus* is an appealing ornamental plant in some East Asian countries. However, genomic research on *A. tataricus*—both nuclear genome sequencing and plastid genome sequencing—has been relatively scarce. This lack of genetic information has hindered the basic research into, and applications of, this valuable plant, such as molecular authentication, breeding, cultivation, and bioactive compound biosynthesis.

The chloroplast, the photosynthetic organelle of most green plants, is involved in both developmental processes and secondary metabolic activities [8], as well as coordination of gene expression between organelle and nuclear genome [9]. The chloroplast (cp) genome is considered to have originated from an ancestral endosymbiotic cyanobacteria [10] and is organized into large clusters of polycistronic transcribed genes [11]. Therefore, it has a highly conserved tetrad structure, with a large single-copy (LSC) region, a small single-copy (SSC) region, and two inverse repeats (IRs) regions. As conserved matrilineal genetic information, chloroplasts play a part in cytoplasmic male sterility (CMS) [12], genetic diversity in populations [13–15], and especially, phylogenetic analysis [16,17]. In addition, the comparison of chloroplast sequences has been successfully used to study the evolution [18] and athentication markers to molecular identification at genus levels [19]. Evolutionary events, such as IR contraction and expansion, and re-inversion of SSC, have been found in the comparison of chloroplast genome [20,21]. Moreover, the chloroplast DNA sequences offers adequate information to study the phylogenetics and phylogeography of angiosperms at lower taxonomic levels [8]. In past decades, there have been great advances in our understanding of chloroplasts [22–25], in terms of their origin, structure, evolution, forward and reverse genetics, and genetic engineering. Moreover, the emergence of second-generation sequencing technology means that there is now less demand for research into chloroplast genomics [26,27]. Therefore, a draft cp genome assembly for *A. tataricus* is of great significance for exploring molecular identification, phylogenetic relationships, and evolutionary events studies among Asteraceae species.

Here, we reported the first whole chloroplast genome sequence for *A. tataricus*, together with the characterization of its gene annotations and repeat compositions. We also compared this chloroplast genome with that of other *Aster* species, which revealed significant variation in genome size and highly divergent regions in intergenic spacers. This comprehensive cp genomic analysis would provide a basis for molecular identification and further understanding of the evolutionary history of *Aster* species.

2. Results and Discussion

2.1. Features of A. tataricus cpDNA

The complete cp genome sequence of *A. tataricus* was 152,992 bp (GenBank accession number MH669275). The structure of the *A. tataricus* cp genome was analogous to those from other *Aster* species [28], and included an LSC region (84,698 bp; covering 55.4%), an SSC region (18,250 bp; covering 11.9%), and a pair of inverted repeats (IRA/IRB, 25,022 bp; covering 16.4%) (Table 1). The content of DNA G + C in the LSC, SSC, and IR regions, and the whole genome, was 35.2%, 31.3%, 43%, and 37.3%, respectively. The DNA G + C content is a very significant indicator when evaluating species affinity, and the cpDNA G + C content of *A. tataricus* is identical to that of other *Aster* species [28]. The DNA G + C content of the IR regions in *A. tataricus* was greater than that of other regions (LSC, SSC); this phenomenon is very common in other plants too [16,17]. The relatively high DNA G + C content of the IR regions is generally attributable to the rRNA genes and tRNA genes [29–31].

Table 1. Summary of complete chloroplast genomes for three *Aster* species.

Species	*Aster altaicus*	*Aster spathulifolius*	*Aster tataricus*
Large single-copy (LSC)			
Length (bp)	84,240	81,998	84,698
G + C (%)	35.3	31.4	35.2
Length (%)	55.3	54.9	55.4
Small single-copy (SSC)			
Length (bp)	18,196	17,973	18,250
G + C (%)	31.3	35.8	31.3
Length (%)	11.9	12.0	11.9
IR			
Length (bp)	25,005	24,751	25,022
G + C (%)	43.0	43.2	43.0
Length (%)	16.4	16.6	16.4
Total			
Length (bp)	152,446	149,473	152,992
G + C (%)	37.3	37.7	37.3
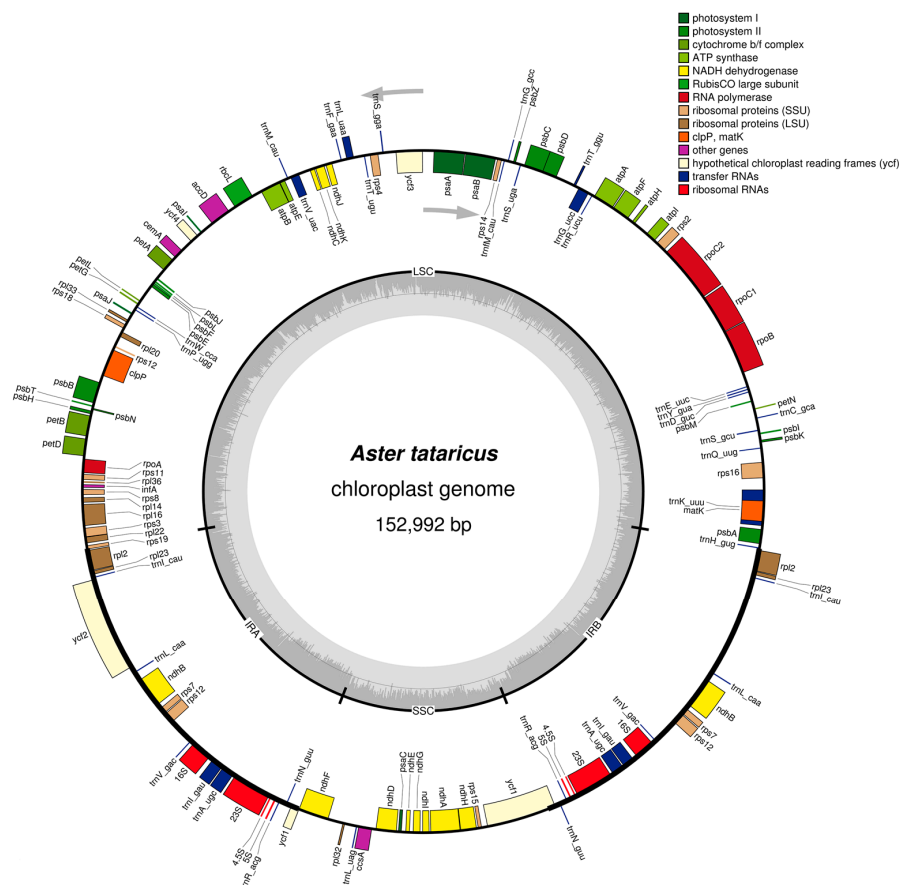
Figure 1. Gene map of the *A. tataricus* chloroplast genome. Genes drawn inside the circle are transcribed clockwise, and those outside are transcribed counterclockwise. Genes belonging to different functional groups are color-coded. The darker gray in the inner circle corresponds to DNA G + C content, while the lighter gray corresponds to A + T content.

In the *A. tataricus* cp genome, 115 functional genes were observed, including four rRNA genes, 30 tRNA genes, and 81 protein-coding genes (Table 2). Furthermore, 18 genes—seven tRNA, all four rRNA, and seven protein-coding genes—were repeated in the IR regions (Figure 1). The LSC region contained 62 protein-coding and 22 tRNA genes, while the SSC region comprised one tRNA gene and 12 protein-coding genes.

The sequences of the tRNA and protein-coding genes were studied, and the frequency of codon usage was inferred and summarized for the *A. tataricus* chloroplast genome. Our study revealed that 23,441 codons characterize the coding capacity of 81 protein-coding and 30 tRNA genes in *A. tataricus* (Table 3). Of these codons, 4065 (17.34%) were found to code for leucine and 204 (0.87%) for tryptophan, which represented the maximum and minimum prevalent number of amino acids in the *A. tataricus* chloroplast genome, respectively. A- and U-ending codons were ordinary.

Table 2. Genes in the *A. tataricus* chloroplast genome.

Category	Gene Group	Gene Names
Self-replication	Large subunit of ribosomal proteins	*rpl2* **,a, *14*, *16* **, *20*, *22*, *23* a, *32*, *33*, *36*
	Small subunit of ribosomal proteins	*rps2*, *3*, *4*, *7* a, *8*, *11*, *12* **,a, *14*, *16* **, *18*, *19*
	DNA-dependent RNA polymerase	*rpoA*, *B*, *C1* **, *C2*
	rRNA genes	*rrn16S* a, *rrn23S* a, *rrn4.5S* a, *rrn5S* a
	tRNA genes	*trnA-UGC* **,a, *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnfM-CAU*, *trnG-UCC* **, *trnG-GCC*, *trnH-GUG*, *trnI-CAU*, *trnI-GAU* **,a, *trnK-UUU* **, *trnL-CAA*, *trnL-UAA* **, *trnL-UAG*, *trnM-CAU*, *trnN-GUU*, *trnP-UGG*, *trnQ-UUG*, *trnR-ACG*, *trnR-UCU*, *trnS-GCU*, *trnS-GGA*, *trnS-UGA*, *trnT-GGU*, *trnT-UGU*, *trnV-GAC*, *trnV-UAC* **, *trnW-CCA*, *trnY-GUA*
Photosynthesis	Photosystem I	*psaA*, *B*, *C*, *I*, *J*
	Photosystem II	*psbA*, *B*, *C*, *D*, *E*, *F*, *H*, *I*, *J*, *K*, *L*, *M*, *N*, *T*, *Z*,
	NADH oxidoreductase	*ndhA* **, *B* **,a, *C*, *D*, *E*, *F*, *G*, *H*, *I*, *J*, *K*
	Cytochrome b6/f complex	*petA*, *B* **, *D* **, *G*, *L*, *N*
	ATP synthase	*atpA*, *B*, *E*, *F* **, *H*, *I*
	Rubisco	*rbcL*
Other genes	Maturase	*matK*
	Protease	*clpP* **
	Envelope membrane protein	*cemA*
	Subunit acetyl-CoA-carboxylase	*AccD*
	c-Type cytochrome synthesis gene	*CcsA*
	Conserved open reading frames	*ycf1*, *2* a, *3* **, *4*, *15*

** Genes containing introns. a Duplicated gene (genes present in the inverted repeat (IR) regions).

There were 18 intron-containing genes in total: 12 protein-coding genes and six tRNA genes (Table 4). Fifteen genes (nine protein-coding and six tRNA genes) comprised one intron, and two genes (*ycf3* and *clpP*) comprised two introns (Table 4). The intron of the *trnK-UUU* gene contains the *matK* gene, and the size of the intron was 2497 bp. The *rps12* gene was a trans-spliced gene, with the 5′ end located in the LSC region and the copied 3′ end located in the IR regions. Earlier studies have reported that *ycf3* is essential for the constant accumulation of the photosystem I compound [32]. Therefore, we suppose that the intron gain in *ycf3* of *A. tataricus* may be valuable information in terms of further studies of the mechanism of photosynthesis evolution. We compared the length of exons and introns in genes with introns in the *A. tataricus* and *A. spathulifolius* chloroplast genomes (Table 4). While these were found to be broadly similar, some differences were noted: (1) in the *A. spathulifolius* chloroplast genome, the *rpl16* gene had no intron; (2) there was significant variation in *rps12* and *rps16* intron length between the two species; and (3) in *A. spathulifolius*, the *rpl12* gene had no intron.

Table 3. Codon–anticodon recognition patterns and codon usage for the *A. tataricus* chloroplast genome.

Amino Acid	Codon	No.	RSCU *	tRNA	Amino Acid	Codon	No.	RSCU *	tRNA
Phe	UUU	1064	1.34		Tyr	UAU	793	1.39	
Phe	UUC	528	0.66	*trnF-GAA*	Tyr	UAC	348	0.61	*trnY-GUA*
Leu	UUA	554	1.55	*trnL-UAA*	Stop	UAA	508	1.09	
Leu	UUG	492	1.37	*trnL-CAA*	Stop	UAG	368	0.79	
Leu	CUU	459	1.28		His	CAU	374	1.36	
Leu	CUC	205	0.57		His	CAC	175	0.64	*trnH-GUG*
Leu	CUA	272	0.76	*trnL-UAG*	Gln	CAA	482	1.4	*trnQ-UUG*
Leu	CUG	169	0.47		Gln	CAG	206	0.6	
Ile	AUU	798	1.46		Asn	AAU	792	1.39	
Ile	AUC	397	0.73	*trnI-GAU*	Asn	AAC	347	0.61	*trnN-GUU*
Ile	AUA	441	0.81	*trnI-UAU*	Lys	AAA	914	1.42	*trnK-UUU*
Met	AUG	415	1	*trn(f)M-CAU*	Lys	AAG	376	0.58	
Val	GUU	364	1.45		Asp	GAU	471	1.43	
Val	GUC	176	0.7	*trnV-GAC*	Asp	GAC	187	0.57	*trnD-GUC*
Val	GUA	318	1.26	*trnV-UAC*	Glu	GAA	550	1.39	*trnE-UUC*
Val	GUG	148	0.59		Glu	GAG	241	0.61	
Ser	UCU	509	1.37		Cys	UGU	345	1.15	
Ser	UCC	329	0.89	*trnS-GGA*	Cys	UGC	253	0.85	*trnC-GCA*
Ser	UCA	493	1.33	*trnS-UGA*	Stop	UGA	524	1.12	
Ser	UCG	292	0.79		Trp	UGG	491	1	*trnW-CCA*
Pro	CCU	259	1.29		Arg	CGU	204	0.73	*trnR-ACG*
Pro	CCC	156	0.78	*trnP-GGG*	Arg	CGC	108	0.39	
Pro	CCA	224	1.12	*trnP-UGG*	Arg	CGA	282	1.01	
Pro	CCG	164	0.82		Arg	CGG	173	0.62	
Thr	ACU	321	1.22		Arg	AGA	329	0.89	*trnR-UCU*
Thr	ACC	246	0.93	*trnT-GGU*	Arg	AGG	277	0.75	
Thr	ACA	314	1.19	*trnT-UGU*	Ser	AGU	578	2.06	
Thr	ACG	174	0.66		Ser	AGC	337	1.2	*trnS-GCU*
Ala	GCU	250	1.25		Gly	GGU	315	0.95	
Ala	GCC	169	0.85		Gly	GGC	205	0.62	*trnG-GCC*
Ala	GCA	242	1.21	*trnA-UGC*	Gly	GGA	466	1.4	*trnG-UCC*
Ala	GCG	138	0.69		Gly	GGG	342	1.03	

* RSCU: relative synonymous codon usage.

Advances in phylogenetic research have revealed that chloroplast genome evolution encompasses both structural changes and nucleotide substitutions [33–35]. A few examples of these changes, including intron or gene losses [21,36], have been discovered in chloroplast genomes. Introns play an important role in regulating gene expression. They can increase gene expression at a particular position and at a specific time [37]. Intron regulation mechanisms have been reported in other species. More experimental work is required to study the relationship between intron loss and gene expression introns in *A. tataricus*.

Table 4. A comparison of exon and intron length in genes with introns in the *A. tataricus* and *A. spathulifolius* chloroplast genomes.

Gene	Location	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
trnK-UUU	LSC	37	2497	38		
		37	2502	35		
trnG-UCC	LSC	23	732	48		
		23	723	47		
trnL-UAA	LSC	34	441	50		
		37	423	50		
trnV-UAC	LSC	36	575	37		
		38	573	37		
trnI-GAU	IR	38	781	35		
		43	776	35		

Table 4. *Cont.*

Gene	Location	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
trnA-UGC	IR	38	820	35		
		38	820	35		
rps12 *	LSC	234	535	25	——	114
		114	——	243	——	243
rps16	LSC	234	820	40		
		39	826	216		
rpl16	LSC	402	1008	10		
		——	——	——		
rpl2	IR	391	671	434		
		393	668	435		
rpoC1	LSC	431	709	1639		
		429	721	1641		
ndhA	SSC	552	1055	540		
		553	1105	540		
ndhB	IR	777	674	756		
		777	670	756		
ycf3	LSC	124	690	228	739	155
		124	697	230	739	153
petB	LSC	6	754	658		
		6	745	642		
atpF	LSC	144	718	411		
		145	699	410		
clpP	LSC	71	812	291	614	229
		71	800	291	623	229
petD	LSC	9	645	526		
		9	724	474		

Exon and intron lengths in genes with introns in the *A. tataricus* chloroplast genome (gray background), and in the *A. spathulifolius* chloroplast genome (normal background) * The *rps12* gene is a trans-spliced gene with the 5′ end located in the LSC region and the duplicated 3′ ends located in the IR regions.

2.2. Simple Sequence Repeat (SSR) Analysis

Simple sequence repeats (SSRs) of 10 bp or longer are inclined toward slipped-strand mispairing, which is known to be the main mutational mechanism utilized in SSR polymorphisms. SSRs in the chloroplast genome can be extremely variable at the intra-specific level and are often used as genetic markers in population genetics and evolutionary studies [38–41]. In this research, we investigated the SSRs in the chloroplast genome of *A. tataricus* and in that of two other *Aster* species (Figure 2). The cp genome of *Aster tataricus*, *Aster altaicus*, and *Aster spathulifolius* contained 70, 58, and 36 SSRs, respectively. The level of mononucleotide repeat content was high (*A. tataricus*, 51.4%; *A. altaicus*, 62.1%; *A. spathulifolius*, 77.8%) in all the above species. These results will provide chloroplast SSR markers that can be used to study genetics, select germplasm for breeding, and facilitate the molecular identification of species.

2.3. Comparative Chloroplast Genomic Analysis

Comparative analysis of genomes is a tremendously important step in genomics [42,43]. Comparing the structural changes amongst *Aster* chloroplast genomes revealed that the chloroplast genome *A. spathulifolius* was the smallest of the three whole *Aster* chloroplast genomes (Table 1). *A. spathulifolius* had the fewest IR regions (17,973 bp) among these sequenced *Aster* chloroplast genomes. We supposed that the dissimilar length of the IR regions was the principal reason for the change in sequence length. To explicate the level of genome differences, the sequence identity of the *Aster* chloroplast DNAs was computed using mVISTA software, with *A. tataricus* as a reference (Figure 3). The results of this comparison showed that the IR (A/B) regions exhibited fewer differences than the LSC and SSC regions. Moreover, the non-coding regions showed more variability than the

coding regions, and the marked differences in regions among the three chloroplast genomes were evident in the intergenic spacers. Of the three *Aster* chloroplast genomes, *A. tataricus* and *A. altaicus* exhibited the fewest differences.
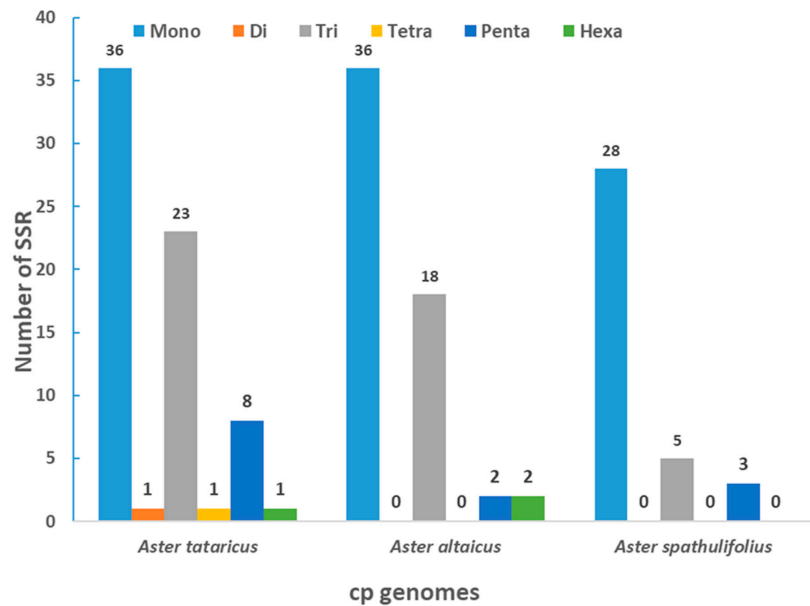

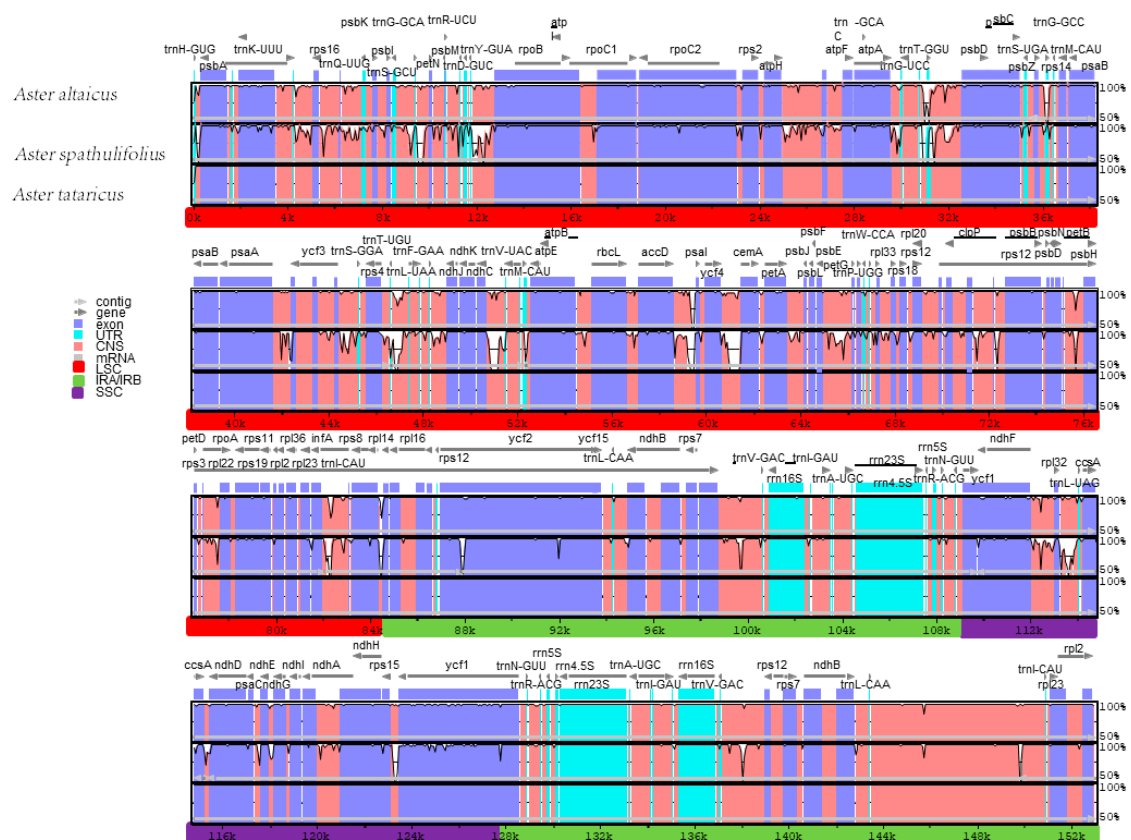**Figure 2.** Analysis of simple sequence repeats (SSRs) in the three *Aster* chloroplast genomes.

Figure 3. Comparison of three chloroplast genomes using mVISTA. Gray arrows and thick black lines above the alignment indicate gene orientation. Purple bars represent exons, blue bars represent untranslated regions (UTRs), pink bars represent conserved non-coding sequences (CNS), and gray bars represent mRNA. The y-axis represents the percentage identity (shown: 50–100%).

2.4. Inverted Repeat (IR) Contraction and Expansion in the A. tataricus Chloroplast Genome

Contractions and expansions of the IR regions at the borders are ordinary evolutionary events and represent the main reasons for changes in the size of chloroplast genomes; they play a significant role in evolution [44–46]. For *A. altaicus*, *A. spathulifolius* and *A. tataricus*, we conducted an exhaustive comparison of four junctions, LSC-IRA (JLA), LSC-IRB (JLB), SSC-IRA (JSA), and SSC-IRB (JSB), between the two IRs (IRA and IRB) and the two single-copy regions (LSC and SSC) (Figure 4). The JSA junction was placed in the *ycf1* pseudogene region in all the *Aster* species chloroplast genomes and outspread to different lengths (*A. altaicus*, 563 bp; *A. spathulifolius*, 608 bp; *A. tataricus*, 563 bp) within the IRA region of all the genomes; the IRB region contained 563, 567, and 565 bp of the *ycf1* gene, respectively. Recently, it was reported that *ycf1* is required for plant viability and codes Tic214, a significant component of the *Arabidopsis*Tic complex member [47,48]. Correspondingly, the *trnH* gene was placed in the LSC region, 3, 22, and 3 bp away from the IRB/LSC border in the three *Aster* chloroplast genomes, respectively. The JLA in the *Aster* species was overlapped by *rps19*. The *ndhF* gene was found to be 22, 5, and 22 bp away from the IRA/SSC border in the *Aster* species.

Although the gene order in chloroplasts is generally conserved in most green plants, it has been reported that many sequences are rearranged in chloroplast genomes from an extensive variety of different plant species, including inversions in the LSC region, IR contractions or expansions with inversions, and re-inversion in the SSC region [49–53]. Sequence rearrangements that convert chloroplast genome structure in related species may also reveal information about genetic diversity that could be used for molecular classification and evolution studies.
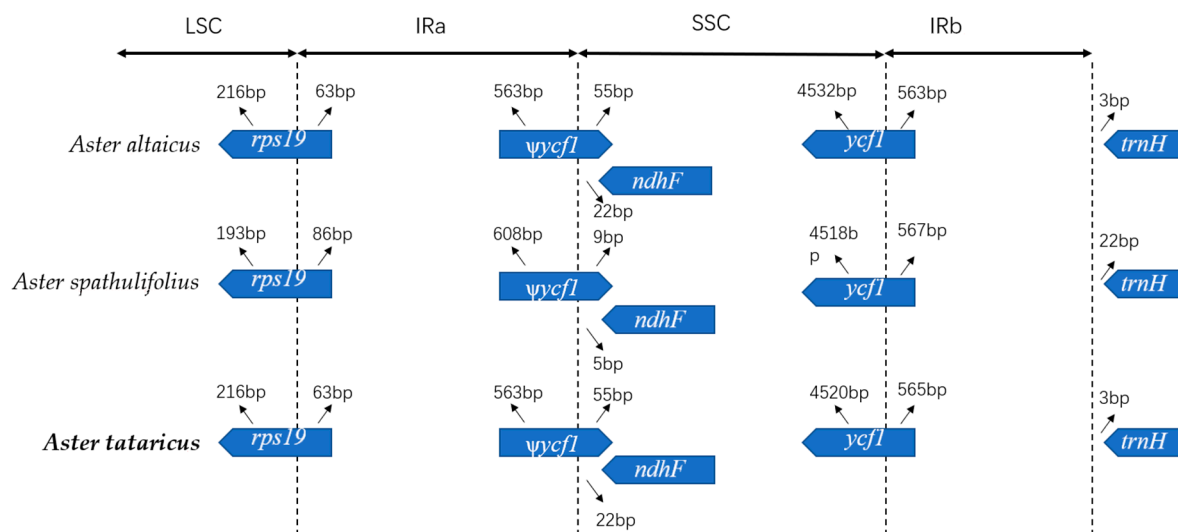

Figure 4. Comparison of border distance between adjacent genes and junctions of the LSC, SSC, and two IR regions among the chloroplast genomes of three *Aster* species. Boxes below the main line indicate the adjacent border genes. The figure is not to scale with respect to sequence length and only shows relative changes at or near the IR/SC borders.

2.5. Phylogenetic Analysis

The availability of a completed *A. tataricus* cp genome provided us with sequence information that can be used to study the phylogeny of *A. tataricus* among Asteraceae. We performed multiple sequence alignments using the whole cp genome sequences in 16 Asteraceae species. One additional cp genome, *Paeonia ostii* (Paeoniaceae), was included as an outgroup (Figure 5). The method of maximum likelihood (ML) was used to construct a phylogenetic tree. The results strongly supported the finding that *A. altaicus* and *A. tataricus* are sister species, and *A. tataricus* is closer to *A. altaicus* than to *A. spathulifolius*.
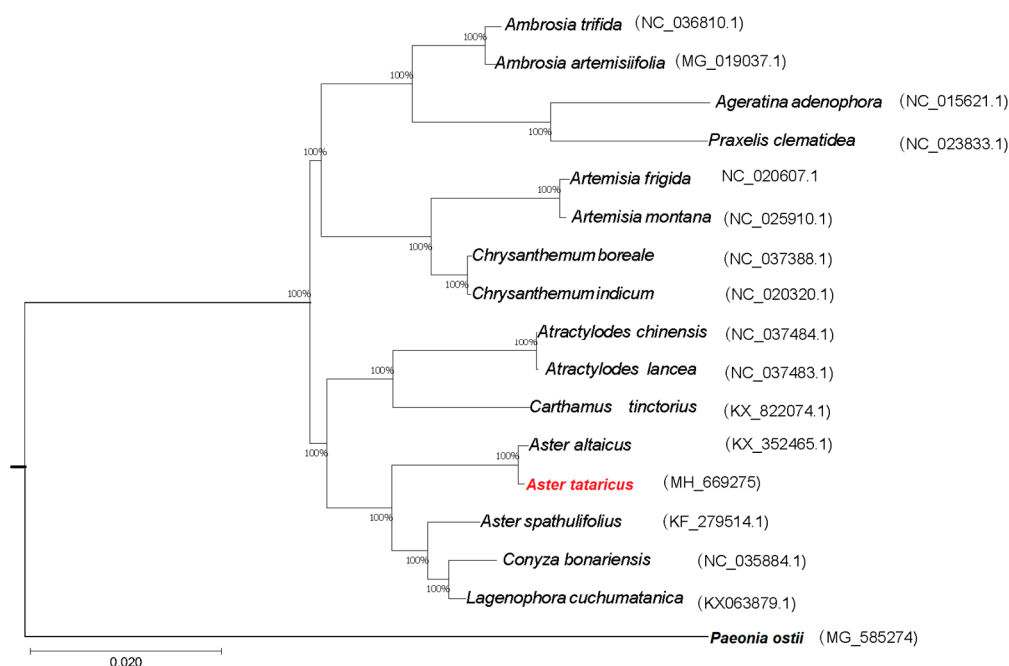
Figure 5. Maximum likelihood (ML) phylogenetic tree reconstruction including 17 species based on concatenated sequences from all chloroplast genomes. The position of *A. tataricus* is indicated in red text. *Paeonia ostii* was used as the outgroup.

3. Materials and Methods

3.1. DNA Sequencing, Chloroplast Genome Assembly, and Validation

The *A. tataricus* was planted in the China Academy of Chinese Medical Sciences (N 39°56′, E 116°25′, Beijing, China). Fresh leaves were gathered and covered with tin foil, frozen in liquid nitrogen, and maintained at −80 °C. An improved cetyltrimethylammonium bromide (CTAB) method was used to obtain the whole genomic DNA of *A. tataricus* [54]. The concentration of DNA was estimated using an ND-2000 spectrometer (Nanodrop Technologies, Wilmington, DE, USA) [55]. A 250-bp shotgun library was constructed according to the manufacturer's instructions (Vazyme Biotech Co. Ltd., Nanjing, China). The library was sequenced using an Illumina X Ten platform (Illumina, San Diego, CA, USA) double terminal sequencing method (150 pair-ends). The sample contained 5 G of raw data, and over 34 million paired-end reads (SRA accession: SRP154896) were obtained.

The raw data was filtered using Skewer-0.2.2 (Institute of Plant Quarantine Research, Chinese Academy of Inspection and Quarantine, Beijing, China, https://sourceforge.net/projects/skewer/) [56]. BLAST searches were used to abstract chloroplast-like reads from clean-reads in comparison with reference sequences (*A. altaicus*). Lastly, we used the chloroplast-like reads to assemble sequences using SOAPdenovo-2.04 (BGI·tech, Shenzhen, China, https://sourceforge.net/projects/soapdenovo2/files/SOAPdenovo2/) [57]. SSPACE-3.0 (Leiden University, Leiden, The Netherlands https://www.baseclear.com/services/bioinformatics/basetools/sspace-standard/) [58] and GapCloser-1.12 (BGI·tech, Shenzhen, China, https://sourceforge.net/projects/oapdenovo2/files/GapCloser/) [59] were used to outspread sequences and fill gaps. PCR amplification and Sanger sequencing were used to confirm the four junction regions between the IR regions and the LSC/SSC regions, to confirm the assembly (Table S1).

3.2. Gene Annotation and Sequence Analyses

CpGAVAS [60] was used to annotate the sequences; DOGMA [61] and BLAST were used to check the annotation findings. tRNAscanSEv1.21 [62], with default settings, was used to identify all tRNA genes. OGDRAWv1.2 [63] was used to show the structural features of the chloroplast genomes. Relative synonymous codon usage (RSCU) values were defined using MEGA5.2 [64].

3.3. Genome Comparison

mVISTA [65] (Shuffle-LAGAN mode) was used to compare the whole chloroplast genome of *A. tataricus*, *A. altaicus* (KX352465), and *A. spathulifolius* (KF279514), with the annotation of *A. tataricus* as the reference. Phobos version 3.3.12 [66] was employed to detect SSRs within the cp genome, with the search parameters set at 10 repeat units for mononucleotides, _8 repeat units for dinucleotides,_4 repeat units for trinucleotides and tetranucleotides, and _3 repeat units for pentanucleotide and hexanucleotide SSRs.

3.4. Phylogenetic Analysis

We downloaded 16 whole chloroplast genome sequences of Asteraceae species from the National Center for Biotechnology Information (NCBI) Organelle Genome and Nucleotide Resources database. The whole chloroplast genome sequences were used to analyze the phylogenetics. The software clustalw2 (The Conway Institute of Biomolecular and Biomedical Research, Dublin, Ireland) was used to align sequences. MEGA5.2 was used to analyze and plot the phylogenetic tree with ML (maximum likelihood). We used 1000 replicates and TBR (tree bisection and reconnection) branch exchange to complete the bootstrap analysis. Furthermore, *Paeonia ostii* was set as the outgroup.

4. Conclusions

To our knowledge, we were the first to complete the sequencing and analysis of the whole chloroplast genome of *A. tataricus*, showing that the quadruple structure, gene order, DNA G + C content, and codon usage features were similar to those of the other *Aster* chloroplast genomes studied. Compared with the chloroplast genomes of the other two *Aster* species, the chloroplast genome of *A. tataricus* was the largest, while the genome structure and composition were found to be similar. Of the three *Aster* chloroplast genomes, *A. tataricus* and *A. altaicus* exhibited the fewest differences. Examination of the phylogenetic relationships among the three *Aster* species revealed that *A. tataricus* was more closely related to *A. altaicus* than to *A. spathulifolius*. The findings of this study offer an assembly of a whole chloroplast genome of *A. tataricus*, which would be valuable for molecular identification, breeding, and further biological discoveries.

References

1.   Zhao, D.X.; Hu, B.Q.; Zhang, M.; Zhang, C.F.; Xu, X.H. Simultaneous separation and determination of phenolic acids, pentapeptides, and triterpenoid saponins in the root of *Aster tataricus* by high-performance liquid chromatography coupled with electrospray ionization quadrupole time-of-flight mass spectrometry. *J. Sep. Sci.* **2015**, *38*, 571–575. [CrossRef] [PubMed]

2. Yu, P.; Cheng, S.; Xiang, J.; Yu, B.; Zhang, M.; Zhang, C.; Xu, X. Expectorant, antitussive, anti-inflammatory activities and compositional analysis of *Aster tataricus*. *J. Ethnopharmacol.* **2015**, *164*, 328–333. [CrossRef] [PubMed]

3. Fang, H.; Shan, G.; Qin, G.; Zhen, L.N.; Li, M.H.; Hao, J.L. Advances on chemical components and pharmacological actions of *Aster tataricus*. *Med. Res. Educ.* **2012**, *29*, 73–77.

4. Zhou, W.B.; Tao, J.Y.; Xu, H.M.; Tan, N.H. Three new antiviral triterpenes from *Aster tataricus*. *Z. Naturforschung B* **2010**, *65*, 1393–1396. [CrossRef]

5. Tang, X.W.; Liu, X.X.; Tang, Y.L.; Liu, Y.L.; Xu, K.H. Analysis of effective constituents from *Aster tataricus* L. and extracting of alkaloid and its antibacterial test in vitro. *J. Tradit. Chin. Vet. Med.* **2006**, *1*, 16–18.

6. Du, L.; Mei, H.F.; Yin, X.; Xing, Y.Q. Delayed growth of glioma by a polysaccharide from *Aster tataricus*, involve upregulation of Bax/Bcl-2 ratio, activation of caspase-3/8/9, and downregulation of the Akt. *Tumour Biol.* **2014**, *35*, 1819–1825. [CrossRef] [PubMed]

7. Morita, H.; Nagashima, S.; Takeya, K.; Itokawa, H. Solution forms of antitumor cyclic pentapeptides with 3,4-dichlorinated proline residues, astins a and c, from *Aster tataricus*. *Chem. Pharm. Bull.* **1995**, *43*, 1395–1397. [CrossRef] [PubMed]

8. Wicke, S.; Schneeweiss, G.M.; Depamphilis, C.W.; Kai, F.M.; Quandt, D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol.* **2011**, *76*, 273–297. [CrossRef] [PubMed]

9. Woodson, J.D.; Chory, J. Coordination of gene expression between organellar and nuclear genomes. *Nat. Rev. Gene* **2008**, *9*, 383–395. [CrossRef] [PubMed]

10. Yoon, H.S.; Hackett, J.D.; Bhattacharya, D. A genomic and phylogenetic perspective on endosymbiosis and algal origin. *J. Appl. Phycol.* **2006**, *18*, 475–481. [CrossRef]

11. Kanno, A.; Hirai, A. A transcription map of the chloroplast genome from rice (*Oryza sativa*). *Curr. Genet.* **1993**, *23*, 166–174. [CrossRef] [PubMed]

12. Nielsen, A.Z.; Ziersen, B.; Jensen, K.; Lassen, L.M.; Olsen, C.E.; Moller, B.L.; Jensen, P.E. Redirecting photosynthetic reducing power toward bioactive natural product synthesis. *ACS Synth. Biol.* **2013**, *2*, 308–315. [CrossRef] [PubMed]

13. Echt, C.S.; Deverno, L.L.; Anzidei, M.; Vendramin, G.G. Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. *Mol. Ecol.* **1998**, *7*, 307–316. [CrossRef]

14. Wang, Y.; Ghouri, F.; Shahid, M.Q.; Naeem, M.; Baloch, F.S. The genetic diversity and population structure of wild soybean evaluated by chloroplast and nuclear gene sequences. *Biochem. Syst. Ecol.* **2017**, *71*, 170–178. [CrossRef]

15. Alvespereira, A.; Clement, C.R.; Picanço-Rodrigues, D.; Veasey, E.A.; Dequigiovanni, G.; Ramos, S.L.F.; Pinheiro, J.B.; Zucchi, M.I. Patterns of nuclear and chloroplast genetic diversity and structure of manioc along major Brazilian Amazonian rivers. *Ann. Bot.* **2018**, *121*, 625–639. [CrossRef] [PubMed]

16. Shen, X.; Wu, M.; Liao, B.; Liu, Z.; Bai, R.; Xiao, S.; Li, X.; Zhang, B.; Xu, J.; Chen, S. Complete chloroplast genome sequence and phylogenetic analysis of the medicinal plant *Artemisia annua*. *Molecules* **2017**, *22*, 1330. [CrossRef] [PubMed]

17. Guo, S.; Guo, L.; Zhao, W.; Xu, J.; Li, Y.; Zhang, X.; Shen, X.; Wu, M.; Hou, X. Complete chloroplast genome sequence and phylogenetic analysis of *Paeonia ostii*. *Molecules* **2018**, *23*, 246. [CrossRef] [PubMed]

18. Zhang, Y.J.; Du, L.W.; Liu, A.; Chen, J.J.; Wu, L.; Hu, W.M.; Zhang, W.; Kim, K.H.; Lee, S.C.; Yang, T.J.; et al. The complete chloroplast genome sequences of five epimedium species: Lights into phylogenetic and taxonomic analyses. *Front. Plant Sci.* **2016**, *7*, 696. [CrossRef] [PubMed]

19. Nguyen, V.B.; Park, H.S.; Lee, S.C.; Lee, J.; Park, J.Y.; Yang, T.J. Authentication markers for five major Panax species developed via comparative analysis of complete chloroplast genome sequences. *J. Agric. Food Chem.* **2017**, *65*, 6298–6306. [CrossRef] [PubMed]

20. Li, Y.; Huo, N.; Dong, L.; Wang, Y.; Zhang, S.; Young, H.A.; Feng, X.; Gu, Y.Q. Complete chloroplast genome sequences of Mongolia medicine *Artemisia frigida* and phylogenetic relationships with other plants. *PLoS ONE* **2013**, *8*, e57533. [CrossRef] [PubMed]

21. He, L.; Qian, J.; Li, X.; Sun, Z.; Xu, X.; Chen, S. Complete chloroplast genome of medicinal plant lonicera japonica: Genome rearrangement, intron gain and loss, and implications for phylogenetic studies. *Molecules* **2017**, *22*, 249. [CrossRef] [PubMed]

22. Daniell, H.; Khan, M.S.; Allison, L. Milestones in chloroplast genetic engineering: An environmentally friendly era in biotechnology. *Trends Plant. Sci.* **2002**, *7*, 84–91. [CrossRef]

23. Daniell, H.; Kumar, S.; Dufourmantel, N. Breakthrough in chloroplast genetic engineering of agronomically important crops. *Trends Biotechnol.* **2005**, *23*, 238–245. [CrossRef] [PubMed]

24. Sang, T.; Crawford, D.J.; Stuessy, T.F. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of Paeonia (Paeoniaceae). *Am. J. Bot.* **1997**, *84*, 1120–1136. [CrossRef] [PubMed]

25. Grosche, C.; Funk, H.T.; Maier, U.G.; Zauner, S. The chloroplast genome of *Pellia endiviifolia*: Gene content, RNA-editing pattern, and the origin of chloroplast editing. *Genome. Biol. Evol.* **2012**, *4*, 1349–1357. [CrossRef] [PubMed]

26. Yang, J.B.; Li, D.Z.; Li, H.T. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Mol. Ecol. Res.* **2015**, *14*, 1024–1031. [CrossRef] [PubMed]

27. Li, R.; Ma, P.F.; Wen, J.; Yi, T.S. Complete sequencing of five Araliaceae chloroplast genomes and the phylogenetic implications. *PLoS ONE* **2013**, *8*, e78568. [CrossRef] [PubMed]

28. Choi, K.S.; Park, S.J. The complete chloroplast genome sequence of *Aster spathulifolius* (Asteraceae); genomic features and relationship with Asteraceae. *Gene* **2015**, *572*, 214–221. [CrossRef] [PubMed]

29. Liu, H.Y.; Yu, Y.; Deng, Y.Q.; Li, J.; Huang, Z.X.; Zhou, S.D. The Chloroplast Genome of *Lilium henrici*: Genome Structure and Comparative Analysis. *Molecules* **2018**, *23*, 1276. [CrossRef] [PubMed]

30. Zhou, J.; Cui, Y.; Chen, X.; Li, Y.; Xu, Z.; Duan, B.; Li, Y.; Song, J.; Yao, H. Complete Chloroplast Genomes of *Papaver rhoeas* and *Papaver orientale*: Molecular Structures, Comparative Analysis, and Phylogenetic Analysis. *Molecules* **2018**, *23*, 437. [CrossRef] [PubMed]

31. Meng, J.; Li, X.; Li, H.; Yang, J.; Wang, H.; He, J. Comparative Analysis of the Complete Chloroplast Genomes of Four Aconitum Medicinal Species. *Molecules* **2018**, *23*, 1015. [CrossRef] [PubMed]

32. Boudreau, E.; Takahashi, Y.; Lemieux, C.; Turmel, M.; Rochaix, J.D. The chloroplast ycf3 and ycf4 open reading frames of *Chlamydomonas reinhardtii* are required for the accumulation of the photosystem I complex. *Embo. J.* **1997**, *16*, 6095–6104. [CrossRef] [PubMed]

33. Carbonell-Caballero, J.; Alonso, R.; Ibañez, V.; Terol, J.; Talon, M.; Dopazo, J. A Phylogenetic Analysis of 34 Chloroplast Genomes Elucidates the Relationships between Wild and Domestic Species within the Genus Citrus. *Mol. Biol. Evol.* **2015**, *32*, 2015–2035. [CrossRef] [PubMed]

34. Li, P.; Zhang, S.; Li, F.; Zhang, S.; Zhang, H.; Wang, X.; Sun, R.; Bonnema, G.; Borm, T.J. A Phylogenetic Analysis of Chloroplast Genomes Elucidates the Relationships of the Six Economically Important Brassica Species Comprising the Triangle of U. *Front. Plant Sci.* **2017**, *8*, 1–13. [CrossRef] [PubMed]

35. Kong, W.; Yang, J. The complete chloroplast genome sequence of *Morus mongolica*, and a comparative analysis within the Fabidae clade. *Curr. Genet.* **2016**, *62*, 165–172. [CrossRef] [PubMed]

36. Doyle, J.J. Multiple Independent Losses of Two Genes and One Intron from Legume Chloroplast Genomes. *Syst. Bot.* **1995**, *20*, 272–294. [CrossRef]

37. Nguyen, D.S.; Sai, T.Z.; Nawaz, G.; Lee, K.; Kang, H. Abiotic stresses affect differently the intron splicing and expression of chloroplast genes in coffee plants (*Coffea arabica*) and rice (*Oryza sativa*). *J. Plant Physiol.* **2016**, *201*, 85–94. [CrossRef] [PubMed]

38. Mohammad-Panah, N.; Shabanian, N.; Khadivi, A.; Rahmani, M.-S.; Emami, A. Genetic structure of gall oak (*Quercus infectoria*) characterized by nuclear and chloroplast SSR markers. *Tree Genet. Genomes* **2017**, *13*, 70–82. [CrossRef]

39. Park, S.H.; Sang, I.P.; Gil, J.; Hwangbo, K.; Um, Y.; Kim, H.B.; Jung, C.S.; Kim, S.C.; Lee, Y. Development of Chloroplast SSR Markers to Distinguish Codonopsis Species. *Korean Soc. Hortic. Sci.* **2017**, *5*, 207–208.

40. Zeng, J.; Chen, X.; Wu, X.F.; Jiao, F.C.; Xiao, B.G.; Li, Y.P.; Tong, Z.J. Genetic diversity analysis of genus Nicotiana based on SSR markers in chloroplast genome and mitochondria genome. *Acta Tab. Sin.* **2016**, *22*, 89–97.

41. Park, S.; Sang, I.P.; Gil, J.; Um, Y.; Jung, C.S.; Lee, J.H.; Kim, S.C.; Kim, H.B.; Lee, Y. Development of Simple Sequence Repeat SSR Markers Based on Chloroplast DNA to Distinguish 3 Angelica Species. *Korean Soc. Hortic. Sci.* **2016**, *10*, 226.

42. Zhihai, H.; Jiang, X.; Shuiming, X.; Baosheng, L.; Yuan, G.; Chaochao, Z.; Xiaohui, Q.; Wen, X.; Shilin, C. Comparative optical genome analysis of two pangolin species: *Manis pentadactyla* and *Manis javanica*. *Gigascience* **2016**, *5*, 1–5. [CrossRef] [PubMed]

43. Xu, J.; Chu, Y.; Liao, B.; Xiao, S.; Yin, Q.; Bai, R.; Su, H.; Dong, L.; Li, X.; Qian, J.; et al. *Panax ginseng* genome examination for ginsenoside biosynthesis. *Gigascience* **2017**, *6*, 1–15. [CrossRef] [PubMed]

44. Kode, V.; Mudd, E.A.; Iamtham, S.; Day, A. The tobacco plastid accD gene is essential and is required for leaf development. *Plant J. Cell Mol. Biol.* **2005**, *44*, 237–244. [CrossRef] [PubMed]

45. Raubeson, L.A.; Peery, R.; Chumley, T.W.; Dziubek, C.; Fourcade, H.M.; Boore, J.L.; Jansen, R.K. Comparative chloroplast genomics: Analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics* **2007**, *8*, 174–201. [CrossRef] [PubMed]

46. Yao, X.; Tang, P.; Li, Z.; Li, D.; Liu, Y.; Huang, H. The first complete chloroplast genome sequences in actinidiaceae: Genome structure and comparative analysis. *PLoS ONE* **2015**, *10*, e0129347. [CrossRef] [PubMed]

47. Dong, W.; Xu, C.; Li, C.; Sun, J.; Zuo, Y.; Shi, S.; Cheng, T.; Guo, J.; Zhou, S. Ycf1, the most promising plastid DNA barcode of land plants. *Sci. Rep.* **2015**, *5*, 1–5. [CrossRef] [PubMed]

48. Kikuchi, S.; Bédard, J.; Hirano, M.; Hirabayashi, Y.; Oishi, M.; Imai, M.; Takase, M.; Ide, T.; Nakai, M. Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science* **2013**, *339*, 571–574. [CrossRef] [PubMed]

49. Doyle, J.J.; Davis, J.I.; Soreng, R.J.; Garvin, D.; Anderson, M.J. Chloroplast DNA inversions and the origin of the grass family (Poaceae). *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7722–7726. [CrossRef] [PubMed]

50. Jansen, R.K.; Palmer, J.D. A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 5818–5822. [CrossRef] [PubMed]

51. Kumar, S.; Hahn, F.M.; McMahan, C.M.; Cornish, K.; Whalen, M.C. Comparative analysis of the complete sequence of the plastid genome of Parthenium argentatum and identification of DNA barcodes to differentiate Parthenium species and lines. *BMC Plant Biol.* **2009**, *9*, 131–143. [CrossRef] [PubMed]

52. Palmer, J.D.; Nugent, J.M.; Herbon, L.A. Unusual structure of geranium chloroplast DNA: A triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 769–773. [CrossRef] [PubMed]

53. Raubeson, L.A.; Jansen, R.K. Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. *Science* **1992**, *255*, 1697–1699. [CrossRef] [PubMed]

54. Shams, S.S.; Vahed, S.Z.; Soltanzad, F.; Kafil, V.; Barzegari, A.; Atashpaz, S.; Barar, J. Highly effective DNA extraction method from fresh, frozen, dried and clotted blood samples. *Bioimpacts* **2011**, *1*, 183–187.

55. Simbolo, M.; Gottardi, M.; Corbo, V.; Fassan, M.; Mafficini, A.; Malpeli, G.; Lawlor, R.T.; Scarpa, A. DNA Qualification Workflow for Next Generation Sequencing of Histopathological Samples. *PLoS ONE* **2013**, *8*, e62692. [CrossRef] [PubMed]

56. Jiang, H.; Lei, R.; Ding, S.W.; Zhu, S. Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinform.* **2014**, *15*, 182–194. [CrossRef] [PubMed]

57. Luo, R.; Liu, B.; Xie, Y.; Li, Z.; Huang, W.; Yuan, J.; He, G.; Chen, Y.; Pan, Q.; Liu, Y.; et al. SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience* **2012**, *1*, 18–24. [CrossRef] [PubMed]

58. Boetzer, M.; Henkel, C.V.; Jansen, H.J.; Butler, D.; Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **2011**, *27*, 578–579. [CrossRef] [PubMed]

59. Acemel, R.D.; Tena, J.J.; Irastorzaazcarate, I.; Marlétaz, F.; Gómez-Marín, C.; de la Calle-Mustienes, E.; Bertrand, S.; Diaz, S.G.; Aldea, D.; Aury, J.M.; et al. A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation. *Nat. Genet.* **2016**, *48*, 336–341. [CrossRef] [PubMed]

60. Liu, C.; Shi, L.; Zhu, Y.; Chen, H.; Zhang, J.; Lin, X.; Guan, X. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics* **2012**, *13*, 715–722. [CrossRef] [PubMed]

61. Wyman, S.K.; Jansen, R.K.; Boore, J.L. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **2004**, *20*, 3252–3255. [CrossRef] [PubMed]

62. Schattner, P.; Brooks, A.N.; Lowe, T.M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **2005**, *33*, 686–689. [CrossRef] [PubMed]

63. Lohse, M.; Drechsel, O.; Bock, R. Organellar Genome DRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* **2007**, *52*, 267–274. [CrossRef] [PubMed]

64. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **2011**, *28*, 2731–2739. [CrossRef] [PubMed]

65. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **2004**, *32*, 273–279. [CrossRef] [PubMed]

66. Kraemer, L.; Beszteri, B.; Gäbler-Schwarz, S.; Held, C.; Leese, F.; Mayer, C.; Pöhlmann, K.; Frickenhau, S. STAMP: Extensions to the STADEN sequence analysis package for high throughput interactive microsatellite marker design. *BMC Bioinform.* **2009**, *10*, 41. [CrossRef] [PubMed]

Sample Availability: Sequence data of *Aster tataricus* are available from the authors.