


Article

Sequencing and Analysis of *Chrysanthemum carinatum* Schousb and *Kalimeris indica*. The Complete Chloroplast Genomes Reveal Two Inversions and *rbcL* as Barcoding of the Vegetable

Xia Liu * , Boyang Zhou, Hongyuan Yang, Yuan Li, Qian Yang, Yuzhuo Lu and Yu Gao

State Key Laboratory of Food Nutrition and Safety, Key Laboratory of Food Nutrition and Safety, Ministry of Education of China, College of Food Engineering and Biotechnology, Tianjin University of Science & Technology, Tianjin 300457, China; zhouboyang456@163.com (B.Z.); yanghongyuan1218@sina.com (H.Y.); 13820282933@163.com (Y.L.); yangqian832402@163.com (Q.Y.); 15822109131@163.com (Y.L.); gaoyu@mail.tust.edu.cn (Y.G.)

* Correspondence: liuxia831930@163.com; Tel.: +86-022-6091-2406

Received: 20 April 2018; Accepted: 31 May 2018; Published: 5 June 2018



Abstract: *Chrysanthemum carinatum* Schousb and *Kalimeris indica* are widely distributed edible vegetables and the sources of the Chinese medicine *Asteraceae*. The complete chloroplast (cp) genome of *Asteraceae* usually occurs in the inversions of two regions. Hence, the cp genome sequences and structures of *Asteraceae* species are crucial for the cp genome genetic diversity and evolutionary studies. Hence, in this paper, we have sequenced and analyzed for the first time the cp genome size of *C. carinatum* Schousb and *K. indica*, which are 149,752 bp and 152,885 bp, with a pair of inverted repeats (IRs) (24,523 bp and 25,003) separated by a large single copy (LSC) region (82,290 bp and 84,610) and a small single copy (SSC) region (18,416 bp and 18,269), respectively. In total, 79 protein-coding genes, 30 distinct transfer RNA (tRNA) genes, four distinct rRNA genes and two pseudogenes were found not only in *C. carinatum* Schousb but also in the *K. indica* cp genome. Fifty-two (52) and fifty-nine (59) repeats, and seventy (70) and ninety (90) simple sequence repeats (SSRs) were found in the *C. carinatum* Schousb and *K. indica* cp genomes, respectively. Codon usage analysis showed that leucine, isoleucine, and serine are the most frequent amino acids and that the UAA stop codon was the significantly favorite stop codon in both cp genomes. The two inversions, the LSC region ranging from *trnC-GCA* to *trnG-UCC* and the whole SSC region were found in both of them. The complete cp genome comparison with other *Asteraceae* species showed that the coding area is more conservative than the non-coding area. The phylogenetic analysis revealed that the *rbcL* gene is a good barcoding marker for identifying different vegetables. These results give an insight into the identification, the barcoding, and the understanding of the evolutionary model of the *Asteraceae* cp genome.

Keywords: *C. carinatum* Schousb; *K. indica*; chloroplast genome; *Asteraceae*; barcoding; inversion

1. Introduction

Chloroplasts are crucial for sustaining life on Earth. The chloroplast (cp) genome encodes many key proteins for photosynthesis and other important metabolic processes of plants' interactions with their environment, such as drought, salt, light, and so on, which give us insights to understand the plant biology, diversity, evolution and climatic adaptation, DNA barcoding and genetic engineering [1–6]. Although a relatively conserved architecture and core gene set of cp genomes are shown across higher plants due to the absence of sexual recombination or occurrence of cp capture, considerable variation occurs during evolution [7–11].

A lot of research on the evolution and barcoding of the cp genome in *Asteraceae* was reported [12]. Particularly interestingly, two cp genome regions have a higher inversion frequency in the sunflower family (*Asteraceae*) compared to most eudicots. One region locates the SSC (small single copy) and the other region locates between the *trnC-GCA* to *trnG-UCC* in the LSC (large single copy) region, which maybe help shape our understanding of Compositae evolution and the adaptive versus non-adaptive processes for cellular and genomic complexity [13–25]. To date, over 2400 sequenced cp genomes (<http://www.ncbi.nlm.nih.gov/genomes/>) are available. For the *Asteraceae* family, 129 whole cp genomes were sequenced, of which 16 have no inversions in the SSC compared with most other land plants [24–26]. However, all 16 species maintain inversion regions in LSC. If there does turn out to be a relationship between inversion and certain evolutionary factors in certain systems, it would depend on the acquisition of more genome sequence information, and the relationship between the cp genome structure/content and the complexity evolutionary factor [27].

The *Asteraceae* (Compositae), or sunflower family is the largest clade in the Asterales and the largest family of flowering plants on Earth. They are angiosperms that appeared about 140 million years ago, comprising 1911 genera and 32,913 species. Nearly one-fourth of all the species of flowering plants belong to the *Asteraceae*, *Fabaceae*, and *Orchidaceae* families [28–30]. The *Asteraceae* family includes a great diversity of species, including annuals, perennials, stem succulents, vines, shrubs, and trees. Commercially important plants in *Asteraceae* include food crops, ornamental plants for their flowers, and species with medicinal properties. *Kalimeris indica* (L.) and *Chrysanthemum carinatum* Schousb both belong to the Asterales family. *Kalimeris indica* (L.) is a member of *Kalimeris* in the sunflower family *Asteraceae* and is variously named as Ma Lan, Ji Er Chang, and Tian Bian Ju in China. *Kalimeris indica* (L.) is not only a traditional Chinese medicine and Miaos' medicinal plant employed in China to treat colds, diarrhea, gastric ulcers, acute gastric abscess, conjunctivitis, acute orchitis, blood vomiting, and injuries, but it is also a popular vegetable [31–35]. As well as *Kalimeris indica*, *Chrysanthemum carinatum* Schousb, an important species of *Chrysanthemum* L. in the *Asteraceae* family, is a popular annual herb foodstuff in China, because of its fragrance, flavor, and abundant nutritional value. Moreover, the aerial parts of *Chrysanthemum* L. have been used for the protection or remedy of several diseases in oriental medicinal systems [36–40].

Here, we report the complete cp genome sequence of *Kalimeris indica* (L.) and *Chrysanthemum carinatum* Schousb for the first time. Meanwhile, their gene structure and organization were analyzed. Furthermore, the whole cp genome sequences were compared with other genii of the *Asteraceae* family, especially the inversions, which were found and compared. Phylogenetic analyses of DNA sequences for *rcbL* of 24 plant species indicated that *rcbL* would be a good molecular barcoding target.

2. Results

2.1. Features of the *C. carinatum* Schousb and *K. indica* cp Genome

The complete cp genome of *C. carinatum* Schousb and *K. indica* have a typical quadripartite structure and are 149,752 bp and 152,885 bp in size, respectively (Figure 1). *C. carinatum* Schousb has an LSC region of 82,290 bp and an SSC region of 18,416 bp which are separated by a pair of IR regions of 24,523 bp (Table 1 and Figure 1A). *K. indica* has an LSC region of 84,610 bp ranging from *trnH-GUG* to *rps19* (Ribosomal protein S19), a SSC region of 18,269 bp from *ycf1* (hypothetical protein 1 gene) to *ndhF* (NAD(P)H dehydrogenase), a pair of IR regions of 25,003 bp from *rps19* to *ycf1* and ranging from pseudogene *ycf1* to pseudogene *rps19*, respectively (Table 1 and Figure 1B). Both of them have two inversions like most of the sunflower family species, one inversion occurs in the whole SSC region and the other inversion region is located in the LSC from *trnC-GCA* to *trnG-UCC* [13,15,41].

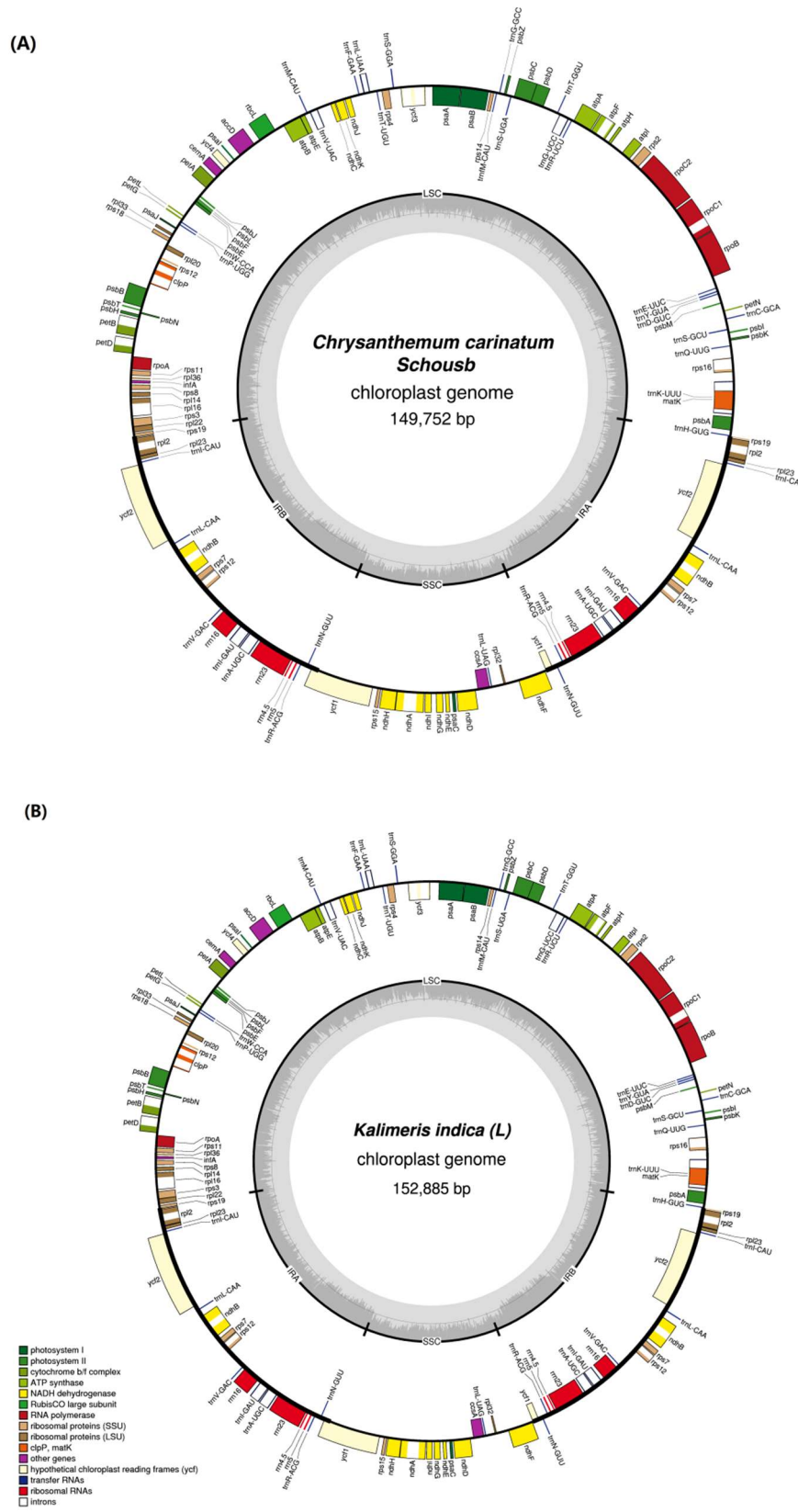


Figure 1. The complete cp genome map of *C. carinatum* Schousb and *K. indica*. **(A)** *C. carinatum* Schousb; **(B)** *K. indica*. The genes marked inside the circle are transcribed clockwise, and those outside are counterclockwise. Genes are color-coded according to their function.

From Table 1, the overall GC content of the *C. carinatum* Schousb and *K. indica* cp genome is the same (37.5%), with a higher GC content (43.1% and 43%) in the IR regions than in the LSC (35.7% and 35.2%) and SSC regions (30.8% and 31.2%), which is similar to other *Asteraceae* [11–15,17,18,25,26,42].

The AT content in the third, second and first codon position of the *C. carinatum* Schousb cp genome are 70.4%, 61.9%, and 54.3%, and the AT content in the third, second, and first codon position of *K. indica* are 69.8%, 61.9%, and 54.5% (Table 1). The results showed that the third codon position and the second codon position have significantly higher AT representation, which is a common feature of the cp genome [43–46].

Table 1. The base composition in the *C. carinatum* Schousb and *K. indica* cp genomes.

Region	<i>C. carinatum</i> Schousb					<i>K. indica</i>				
	T(U) (%)	C (%)	A (%)	G (%)	Length (bp)	T(U) (%)	C (%)	A (%)	G (%)	Length (bp)
LSC	32.4	17.6	32.0	18.1	82,290	32.6	17.3	32.2	17.9	84,610
SSC	35.1	14.7	34.1	16.1	18,416	34.9	14.8	33.9	16.4	18,269
IRa	28.3	22.3	28.6	20.8	24,523	28.3	22.2	28.7	20.8	25,003
IRb	28.6	20.8	28.3	22.3	24,523	28.7	20.8	28.3	22.2	25,003
Total	31.8	19.0	30.7	18.5	149,752	31.8	19.0	30.7	18.5	152,885
CDS	31.5	17.7	30.7	20.1	77,289	31.5	17.7	30.6	20.3	78,372
1st position	23.7	18.9	30.6	26.7	25,763	23.9	18.8	30.6	26.8	26,124
2nd position	32.6	20.4	29.3	17.7	25,763	32.6	20.3	29.3	17.8	26,124
3rd position	38.3	13.7	32.1	15.9	25,763	38.0	13.9	31.8	16.3	26,124

2.2. Functional Genes of the *C. carinatum* Schousb and *K. indica* cp Genome

There are 113 unique functional genes and two pseudogenes in the *C. carinatum* Schousb or *K. indica* cp genome (Table 2). Among the 113 functional genes, 79 protein-coding genes, 30 distinct transfer RNA (tRNA) genes, and four distinct rRNA genes were found not only in the *C. carinatum* Schousb genome but also in the *K. indica* cp genome (Table 2). Notably, seven protein-coding, seven tRNA, and all the rRNA genes are duplicated in the IR regions, which is common in most cp genomes [47,48]. Meanwhile, the coding regions constitute 51.6% and 51.3% of the genome in the *C. carinatum* Schousb and *K. indica* cp genome, respectively. The non-coding regions constitute 48.4% and 48.7% including the introns, pseudogenes, and intergenic spacers, respectively. Two pseudogenes (*ycf1* and *rps19*) are both found in both *C. carinatum* Schousb and *K. indica* cp genomes, a common feature shared with cp genomes from other plants [25,47,48].

Among the *C. carinatum* Schousb or *K. indica* cp genomes, a total of 18 genes (six tRNA genes and 12 protein-coding genes) contain introns (Table 3). They are mostly located in the LSC region (13 genes), with only one located in the SSC region, and four genes locate in the IR regions. Like other angiosperms, *clpP*, *rps12*, and *ycf3* contain two introns [48]. Interestingly, *rps12* was found to be a trans-spliced gene, whose 5' end is located in the LSC region and the duplicated 3' end is located in the IR region. Consistent with many research results, the largest intron is in the *trnK-UUUU* gene, which contains the *matk* gene in the intron. Moreover, *trnL-UAA* has the smallest intron in both of them. Comparing these 18 introns between *C. carinatum* Schousb and *K. indica*, most of them are in *K. indica* and are longer than those in *C. carinatum* Schousb, whereas the introns of *rpl16*, *trnK-UUUU*, and the introII *rps12* are a bit shorter, and the *trnV-UAC* intron is the same size, which may be one reason why the whole cp genome of *K. indica* is larger than *C. carinatum* Schousb.

Table 2. The genes present in the *C. carinatum* Schousb and *K. indica* cp genomes.

Group of Genes	<i>C. carinatum</i> Schousb Gene Names	<i>K. indica</i> Gene Names
Photosystem I	<i>psaA, psaB, psaC, psal, psaj</i>	<i>psaA, psaB, psaC, psal, psaj</i>
Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
Cytochrome b/f complex	<i>petA, petB *, petD *, petG, petL, petN</i>	<i>petA, petB *, petD *, petG, petL, petN</i>
ATP synthase	<i>atpA, atpB, atpE, atpF *, atpH, atpI</i>	<i>atpA, atpB, atpE, atpF *, atpH, atpI</i>
NADH dehydrogenase	<i>ndhA *, ndhB * (×2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>	<i>ndhA *, ndhB * (×2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
RuBisCO large subunit	<i>rbcL</i>	<i>rbcL</i>
RNA polymerase	<i>rpoA, rpoB, rpoC1 *, rpoC2</i>	<i>rpoA, rpoB, rpoC1 *, rpoC2</i>
Ribosomal proteins (SSU)	<i>rps2, rps3, rps4, rps7 (×2), rps8, rps11, rps12 ** (×2), rps14, rps15, rps16 *, rps18, rps19</i>	<i>rps2, rps3, rps4, rps7 (×2), rps8, rps11, rps12 ** (×2), rps14, rps15, rps16 *, rps18, rps19</i>
Ribosomal proteins (LSU)	<i>rpl2 * (×2), rpl14, rpl16 *, rpl20, rpl22, rpl23 (×2), rpl32, rpl33, rpl36</i>	<i>rpl2 * (×2), rpl14, rpl16 *, rpl20, rpl22, rpl23 (×2), rpl32, rpl33, rpl36</i>
Miscellaneous proteins	<i>accD, clpP **, matK, ccsA, cemA, infA</i>	<i>accD, clpP **, matK, ccsA, cemA, infA</i>
Hypothetical chloroplast reading frames (ycf)	<i>ycf1, ycf2 (×2), ycf3 **, ycf4</i>	<i>ycf1, ycf2 (×2), ycf3 **, ycf4</i>
Transfer RNAs	<i>trnA-UGC (×2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnG-UCC, trnH-GUG, trnI-CAU (×2), trnI-GAU (×2), trnK-UUU, trnL-CAA(×2), trnL-UAA, trnL-UAG, trnM-CAU, trnM-CAU, trnN-GUU (×2), trnP-UGG, trnQ-UUG, trnR-ACG (×2), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC (×2), trnV-UAC, trnW-CCA, trnY-GUA</i>	<i>trnA-UGC (×2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnG-UCC, trnH-GUG, trnI-CAU (×2), trnI-GAU (×2), trnK-UUU, trnL-CAA(×2), trnL-UAA, trnL-UAG, trnM-CAU, trnM-CAU, trnN-GUU (×2), trnP-UGG, trnQ-UUG, trnR-ACG (×2), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC (×2), trnV-UAC, trnW-CCA, trnY-GUA</i>
Ribosomal RNAs	<i>rrn4.5 (×2), rrn5 (×2), rrn16 (×2), rrn23 (×2)</i>	<i>rrn4.5 (×2), rrn5 (×2), rrn16 (×2), rrn23 (×2)</i>
Pseudogenes	<i>ycf1, rps19</i>	<i>ycf1, rps19</i>
Total	132	132

(×2) indicates a duplicated gene; * represents the introns that the gene contains; ** indicates there are two introns that the gene contains.

2.3. Codon Usage of the *K. indica* and *C. carinatum* Schousb cp Genome

As shown in Tables 4 and 5, a total of 25,415 and 26,124 codons are involved in the protein-coding in *C. carinatum* Schousb and *K. indica*, respectively. Of these codons, leucine, isoleucine, and serine are the most frequent amino acids in *C. carinatum* Schousb, which encode in 2778 (10.93%), 2195 (8.64%), and 2080 (8.18%) codons, respectively. Leucine, isoleucine, and serine are the most frequent amino acids in *K. indica*, which encode in 2795 (10.70%), 2190 (8.38%) and 2131 (8.16%) codons, respectively. Both of them contain 85 stop codons, which is the least frequent codon. The cysteine codons are the least frequent universal amino acid, which has 286 (1.13%) and 294 (1.13%) codons in *C. carinatum* Schousb and *K. indica*, respectively.

Usually, relative synonymous codon usage (RSCU) can be divided into four types, including lack of bias (RSCU < 1.0), low bias (1.0 < RSCU < 1.2), moderately biased (1.2 < RSCU < 1.3) and highly biased (RSCU > 1.3) [49,50]. As shown in Tables 4 and 5, it is uncannily similar that there are 32 lack of bias codons, two no-bias codons with RSCU = 1 (tryptophan and methionine), and 21 highly biased codons, with the exception of five low bias codons in *C. carinatum* Schousb but two low bias codons in *K. indica*, and three moderately biased codons in *C. carinatum* Schousb but six moderately biased codons in *K. indica*, respectively. The UAA stop codon was the significantly favorite stop codon in the cp genomes. The results showed that the RSCU was significantly biased except for tryptophan and methionine in *C. carinatum* Schousb and *K. indica* and that the A/T ending is very rich, which is popular in the cp genomes of higher plants [51,52].

Table 3. The intron-containing genes of the *C. carinatum Schousb* and *K. indica* cp genomes and the lengths of the introns and exons.

<i>C. Gene</i>	Location	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)	<i>K. Gene</i>	Location	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
<i>atpF</i>	LSC	145	699	410			<i>atpF</i>	LSC	145	709	410		
<i>clpP</i>	LSC	71	796	291	611	229	<i>clpP</i>	LSC	71	814	291	615	229
<i>ndhA</i>	SSC	539	1045	553			<i>ndhA</i>	SSC	553	1064	539		
<i>ndhB</i>	IR	777	670	756			<i>ndhB</i>	IR	777	674	756		
<i>petB</i>	LSC	6	751	642			<i>petB</i>	LSC	6	823	642		
<i>petD</i>	LSC	8	675	475			<i>petD</i>	LSC	8	809	475		
<i>rpl16</i>	LSC	9	1029	399			<i>rpl16</i>	LSC	9	1098	399		
<i>rpl2</i>	IR	391	664	434			<i>rpl2</i>	IR	391	671	434		
<i>rpoC1</i>	LSC	432	733	1641			<i>rpoC1</i>	LSC	432	742	1638		
<i>rps12</i> *	LSC	114	-	232	536	26	<i>rps12</i> *	LSC	114		232	535	26
<i>rps16</i>	LSC	41	891	184			<i>rps16</i>	LSC	41	876	226		
<i>trnA-UGC</i>	IR	38	812	35			<i>trnA-UGC</i>	IR	35	820	38		
<i>trnG-UCC</i>	LSC	23	722	47			<i>trnG-UCC</i>	LSC	23	732	48		
<i>trnI-GAU</i>	IR	43	775	35			<i>trnI-GAU</i>	IR	38	780	35		
<i>trnK-UUU</i>	LSC	37	2562	30			<i>trnK-UUU</i>	LSC	37	2539	35		
<i>trnL-UAA</i>	LSC	37	422	50			<i>trnL-UAA</i>	LSC	37	438	50		
<i>trnV-UAC</i>	LSC	38	573	37			<i>trnV-UAC</i>	LSC	38	573	37		
<i>ycf3</i>	LSC	125	698	229	735	153	<i>ycf3</i>	LSC	125	702	229	739	153

* The *rps12* gene is a trans-spliced gene with the 5' end located in the LSC region and a duplicated 3' end in the IR region.

Table 4. The codon-anticodon recognition pattern and codon usage for the *C. carinatum* Schousb cp genome.

Amino Acid	Codon	No.	RSCU	tRNA	Amino Acid	Codon	No.	RSCU	tRNA
Phe	UUU	956	1.32		Stop	UAA	49	1.73	
Phe	UUC	496	0.68	<i>trnF-GAA</i>	Stop	UAG	21	0.74	
Leu	UUA	874	1.89	<i>trnL-UAA</i>	Stop	UGA	15	0.53	
Leu	UUG	553	1.19	<i>trnL-CAA</i>	His	CAU	452	1.52	
Leu	CUU	620	1.34		His	CAC	143	0.48	<i>trnH-GUG</i>
Leu	CUC	183	0.40		Gln	CAA	703	1.51	<i>trnQ-UUG</i>
Leu	CUA	358	0.77	<i>trnL-UAG</i>	Gln	CAG	227	0.49	
Leu	CUG	190	0.41		Asn	AAU	978	1.56	
Ile	AUU	1075	1.47		Asn	AAC	277	0.44	<i>trnN-GUU</i>
Ile	AUC	428	0.58	<i>trnI-GAU</i>	Lys	AAA	1026	1.50	<i>trnK-UUU</i>
Ile	AUA	692	0.95	<i>trnI-CAU</i>	Lys	AAG	345	0.50	
Met	AUG	613	1.00	<i>trn(f)M-CAU</i>	Asp	GAU	831	1.59	
Val	GUU	490	1.44		Asp	GAC	214	0.41	<i>trnD-GUC</i>
Val	GUC	166	0.49	<i>trnV-GAC</i>	Glu	GAA	986	1.50	<i>trnE-UUC</i>
Val	GUA	523	1.54	<i>trnV-UAC</i>	Glu	GAG	331	0.50	
Val	GUG	178	0.52		Cys	UGU	198	1.38	
Ser	UCU	580	1.77		Cys	UGC	88	0.62	<i>trnC-GCA</i>
Ser	UCC	313	0.96	<i>trnS-GGA</i>	Trp	UGG	448	1.00	<i>trnW-CCA</i>
Ser	UCA	394	1.20	<i>trnS-UGA</i>	Arg	CGU	336	1.32	<i>trnR-ACG</i>
Ser	UCG	154	0.47		Arg	CGC	100	0.39	
Ser	AGA	474	1.86		Arg	CGA	339	1.33	
Ser	AGG	165	0.65	<i>trnS-GCU</i>	Arg	CGG	118	0.46	
Tyr	UAU	799	1.64		Arg	AGU	406	1.24	<i>trnR-UCU</i>
Tyr	UAC	175	0.36	<i>trnY-GUA</i>	Arg	AGC	118	0.36	
Pro	CCA	322	1.17	<i>trnP-UGG</i>	Gly	GGU	581	1.32	
Pro	CCG	159	0.58		Gly	GGC	188	0.43	<i>trnG-GCC</i>
Pro	CCU	434	1.58		Gly	GGA	686	1.56	<i>trnG-UCC</i>
Pro	CCC	184	0.67		Gly	GGG	305	0.69	
Thr	ACU	529	1.64		Ala	GCA	416	1.18	<i>trnA-UGC</i>
Thr	ACC	236	0.73	<i>trnT-GGU</i>	Ala	GCG	162	0.46	
Thr	ACA	404	1.25	<i>trnT-UGU</i>	Ala	GCU	611	1.73	
Thr	ACG	125	0.39		Ala	GCC	223	0.63	

RSCU: Relative synonymous codon usage. RSCU > 1 are highlighted in bold.

Table 5. The codon-anticodon recognition pattern and codon usage for the *K. indica* cp genome.

Amino Acid	Codon	No.	RSCU	tRNA	Amino Acid	Codon	No.	RSCU	tRNA
Phe	UUU	982	1.31		Stop	UAA	49	1.73	
Phe	UUC	515	0.69	<i>trnF-GAA</i>	Stop	UAG	21	0.74	
Leu	UUA	870	1.87	<i>trnL-UAA</i>	Stop	UGA	15	0.53	
Leu	UUG	578	1.24	<i>trnL-CAA</i>	His	CAU	452	1.51	
Leu	CUU	607	1.30		His	CAC	148	0.49	<i>trnH-GUG</i>
Leu	CUC	186	0.4		Gln	CAA	723	1.53	<i>trnQ-UUG</i>
Leu	CUA	375	0.81	<i>trnL-UAG</i>	Gln	CAG	220	0.47	
Leu	CUG	179	0.38		Asn	AAU	969	1.53	
Ile	AUU	1072	1.47		Asn	AAC	295	0.47	<i>trnN-GUU</i>
Ile	AUC	427	0.58	<i>trnI-GAU</i>	Lys	AAA	1036	1.48	<i>trnK-UUU</i>
Ile	AUA	691	0.95	<i>trnI-CAU</i>	Lys	AAG	360	0.52	
Met	AUG	633	1	<i>trn(f)M-CAU</i>	Asp	GAU	847	1.61	
Val	GUU	503	1.45		Asp	GAC	205	0.39	<i>trnD-GUC</i>
Val	GUC	180	0.52	<i>trnV-GAC</i>	Glu	GAA	987	1.47	<i>trnE-UUC</i>
Val	GUA	517	1.49	<i>trnV-UAC</i>	Glu	GAG	359	0.53	
Val	GUG	190	0.55		Cys	UGU	209	1.42	
Ser	UCU	585	1.76		Cys	UGC	85	0.58	<i>trnC-GCA</i>
Ser	UCC	308	0.93	<i>trnS-GGA</i>	Trp	UGG	463	1	<i>trnW-CCA</i>
Ser	UCA	401	1.21	<i>trnS-UGA</i>	Arg	CGU	351	1.33	<i>trnR-ACG</i>
Ser	AGG	177	0.67	<i>trnS-GCU</i>	Arg	CGC	108	0.41	
Ser	UCG	172	0.52		Arg	CGA	346	1.31	
Tyr	UAU	804	1.64		Arg	CGG	114	0.43	
Tyr	UAC	175	0.36	<i>trnY-GUA</i>	Arg	AGU	405	1.22	<i>trnR-UCU</i>
Pro	CCA	314	1.14	<i>trnP-UGG</i>	Arg	AGC	121	0.36	
Pro	CCG	167	0.61		Gly	GGU	571	1.28	
Pro	CCU	209	0.76		Gly	GGC	199	0.45	<i>trnG-GCC</i>
Pro	CCC	184	0.67		Gly	GGA	681	1.53	<i>trnG-UCC</i>
Thr	ACU	531	1.62		Gly	GGG	328	0.74	
Thr	ACC	239	0.73	<i>trnT-GGU</i>	Ala	GCU	622	1.74	
Thr	ACA	405	1.23	<i>trnT-UGU</i>	Ala	GCC	233	0.65	
Thr	ACG	137	0.42		Ala	GCA	410	1.15	<i>trnA-UGC</i>
					Ala	GCG	161	0.45	

2.4. Repeats Structure and SSRs

Analysis of the repeat structure showed that the total of 42 repeats of *C. carinatum* Schousb is significantly less than *K. indica* (59), which may be one reason for the larger cp genome of *K. indica*. There are 18 forward repeats, 20 palindromic repeats, three reverse repeats and one complement repeats in the cp genome of the *C. carinatum* Schousb (Table 6) and 17 forward repeats, 28 palindromic repeats, 11 reverse repeats, and eight complement repeats in the *K. indica* cp genome (Table 7). All of them range from 30 to 60 bp in length and are mostly located in the intergenic spacer (IGS) and intron sequences. Comparison analysis of the repeats of six other species of *Asteraceae* showed that *H. annuus* contained the most repeats (572) and *T. mongolicum* contained the fewest repetitive sequences (28) (Figure 2). Among the eight species of *Asteraceae*, we also found that the base fragment with the most repeats was between 30–39 bp in length. Complement repeats are relatively richer than in other families, although *A. annua*, *E. paradoxa*, *T. mongolicum* and *C. indicum* do not contain complement repeats [48].

There are 70 and 90 simple sequence repeats (SSRs) in the cp genome of *C. carinatum* Schousb and *K. indica*, respectively (Tables 8 and 9). Most of them are mononuclear repeats; 43 in *C. carinatum* Schousb and 30 in *K. indica*, respectively. Eight dinucleotide repeats, five trinucleotide repeats, thirteen tetranucleotide repeats, and one pentanucleotide repeats were also found in the *C. carinatum* Schousb cp genome. Eighteen dinucleotide repeats, twenty-eight trinucleotide repeats, seven tetranucleotide repeats, and seven pentanucleotide repeats were also discovered in *K. indica*. The majority of SSRs are located in the LSC region, 81.43% (*C. carinatum* Schousb) and 78.89% (*K. indica*) of SSRs are located in the LSC (Tables 8 and 9) whereas, only twelve (*C. carinatum* Schousb) and 9 (*K. indica*) SSRs are located in the CDSs (Table 10). Comparing with the six other species of *Asteraceae*, the most SSRs and the least SSRs located in CDS were found in the *K. indica* cp genome. Furthermore, most of the SSRs are AT repeats, which is consistent with the AT richness [48,51].

Table 6. The repeats of the *C. carinatum* Schousb cp genome and their distribution.

No.	Size (bp)	Type	Repeat 1 Location	Repeat 2 Location	Region	No.	Size (bp)	Type	Repeat 1 Location	Repeat 2 Location	Region
1	60	F	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRb	22	31	R	IGS (<i>trnT-GGU, psbD</i>)	IGS (<i>trnT-GGU, psbD</i>)	LSC
2	60	P	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRb, IRa	23	30	P	IGS (<i>psbI, trnS-GUC</i>)	<i>trnS-GGA</i> (CDS)	LSC
3	60	P	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRb, IRa	24	30	F	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRb
4	60	F	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRa, IRa	25	30	P	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRb, IRa
5	51	F	IGS (<i>rps11, rpl36</i>)	IGS (<i>rps11, rpl36</i>)	LSC	26	30	P	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRb, IRa
6	48	P	IGS (<i>psbT, psbN</i>)	IGS (<i>psbT, psbN</i>)	LSC	27	35	F	<i>psaB</i> (CDS)	<i>psaA</i> (CDS)	LSC
7	46	F	IGS (<i>accD, psal</i>)	IGS (<i>accD, psal</i>)	LSC	28	35	F	<i>ycf3</i> (intron2)	<i>ndhB</i> (intron)	LSC, IRb
8	45	F	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRb	29	35	P	<i>ycf3</i> (intron2)	<i>ndhB</i> (intron)	LSC, IRa
9	45	P	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRb, IRa	30	32	P	IGS (<i>trnT-GGU, psbD</i>)	IGS (<i>trnT-GGU, psbD</i>)	LSC
10	45	P	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRb, IRa	31	31	F	IGS (<i>psbI, trnS-GUC</i>)	IGS (<i>psbI, trnS-GUC</i>)	LSC
11	46	P	IGS (<i>petN, psbM</i>)	IGS (<i>petN, psbM</i>)	LSC	32	30	P	IGS (<i>psbC, trnS-UGA</i>)	<i>trnS-GGA</i> (CDS)	LSC
12	39	P	IGS (<i>rps12, trnV-GAC</i>)	<i>ndhA</i> (intron)	IRb, SSC	33	30	F	<i>rbcl</i> (CDS)	IGS (<i>rbcl, accD</i>)	LSC
13	39	F	<i>ndhA</i> (intron)	IGS (<i>trnV-GAC, rps12</i>)	SSC, IRa	34	32	F	IGS (<i>psbI, trnS-GUC</i>)	IGS (<i>psbC, trnS-UGA</i>)	LSC
14	41	F	<i>ycf3</i> (intron2)	IGS (<i>rps12, trnV-GAC</i>)	LSC, IRb	35	31	R	IGS (<i>trnL-UAG, rpl32</i>)	IGS (<i>trnL-UAG, rpl32</i>)	SSC
15	41	P	<i>ycf3</i> (intron2)	IGS (<i>trnV-GAC, rps12</i>)	LSC, IRa	36	30	C	IGS (<i>trnT-GGU, psbD</i>)	IGS (<i>trnT-GGU, psbD</i>)	LSC
16	39	P	<i>ycf3</i> (intron2)	<i>ndhA</i> (intron)	LSC, SSC	37	30	F	<i>psaB</i> (CDS)	<i>psaA</i> (CDS)	LSC
17	42	F	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRb	38	30	F	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRb, IRa
18	42	P	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRb, IRa	39	30	P	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRb
19	42	P	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRb, IRa	40	30	F	IGS (<i>trnL-UAG, rpl32</i>)	IGS (<i>trnL-UAG, rpl32</i>)	SSC
20	42	F	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRa	41	30	R	IGS (<i>trnL-UAG, rpl32</i>)	IGS (<i>trnL-UAG, rpl32</i>)	SSC
21	41	P	IGS (<i>ndhE, psac</i>)	IGS (<i>psac, ndhD</i>)	SSC	42	30	P	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRa

F = forward, P = palindrome, R = reverse, C = complement, IGS = intergenic spacer.

Table 7. The repeats of the *K. indica* cp genome and their distribution.

No.	Size (bp)	Type	Repeat 1 Location	Repeat 2 Location	Region	No.	Size (bp)	Type	Repeat 1 Location	Repeat 2 Location	Region
1	48	P	IGS (<i>psbT</i> , <i>psbN</i>)	IGS (<i>psbT</i> , <i>psbN</i>)	LSC	31	32	F	IGS (<i>rrn5</i> , <i>rrn4.5</i>)	IGS (<i>rrn5</i> , <i>rrn4.5</i>)	IRa
2	39	P	<i>ycf3</i> (intron2)	<i>ndhA</i> (intron)	LSC, SSC	32	34	R	IGS (<i>trnT</i> -GGU, <i>psbD</i>)	IGS (<i>trnT</i> -GGU, <i>psbD</i>)	LSC
3	41	F	<i>ycf3</i> (intron2)	IGS (<i>rps12</i> , <i>trnV</i> -GAC)	LSC, IRb	33	31	R	IGS (<i>ndhC</i> , <i>trnV</i> -UAC)	<i>petB</i> (intron)	LSC
4	41	P	<i>ycf3</i> (intron2)	IGS (<i>trnV</i> -GAC, <i>rps12</i>)	LSC, IRa	34	33	C	IGS (<i>accD</i> , <i>psaI</i>)	<i>petD</i> (CDS)	LSC
5	39	P	IGS (<i>rps12</i> , <i>trnV</i> -GAC)	<i>ndhA</i> (intron)	IRb, SSC	35	33	C	IGS (<i>accD</i> , <i>psaI</i>)	IGS (<i>accD</i> , <i>psaI</i>)	LSC
6	39	F	<i>ndhA</i> (intron)	IGS (<i>trnV</i> -GAC, <i>rps12</i>)	SSC, IRa	36	30	P	IGS (<i>psbC</i> , <i>trnS</i> -UGA)	<i>trnS</i> -GGA (CDS)	LSC
7	42	F	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRb	37	32	C	IGS (<i>trnH</i> -GUG, <i>psbA</i>)	IGS (<i>accD</i> , <i>psaI</i>)	LSC
8	42	P	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRb, IRa	38	32	F	IGS (<i>psbI</i> , <i>trnS</i> -GCU)	IGS (<i>psbC</i> , <i>trnS</i> -UGA)	LSC
9	42	P	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRb, IRa	39	32	P	IGS (<i>petN</i> , <i>psbM</i>)	IGS (<i>petN</i> , <i>psbM</i>)	LSC
10	42	F	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRa	40	32	C	IGS (<i>trnG</i> -UCC, <i>trnT</i> -GGU)	<i>petD</i> (intron)	LSC
11	41	R	<i>petD</i> (intron)	<i>petD</i> (intron)	LSC	41	32	F	<i>psaB</i> (CDS)	<i>psaA</i> (CDS)	LSC
12	43	P	IGS (<i>psaC</i> , <i>ndhD</i>)	IGS (<i>psaC</i> , <i>ndhD</i>)	SSC	42	32	R	IGS (<i>accD</i> , <i>psaI</i>)	IGS (<i>accD</i> , <i>psaI</i>)	LSC
13	39	R	IGS (<i>accD</i> , <i>psaI</i>)	IGS (<i>accD</i> , <i>psaI</i>)	LSC	43	31	P	IGS (<i>trnG</i> -UCC, <i>trnT</i> -GGU)	IGS (<i>ndhC</i> , <i>trnV</i> -UAC)	LSC
14	31	F	IGS (<i>rps12</i> , <i>trnV</i> -GAC)	IGS (<i>rps12</i> , <i>trnV</i> -GAC)	IRb	44	31	P	IGS (<i>trnG</i> -UCC, <i>trnT</i> -GGU)	IGS (<i>trnG</i> -UCC, <i>trnT</i> -GGU)	LSC
15	31	P	IGS (<i>rps12</i> , <i>trnV</i> -GAC)	IGS (<i>trnV</i> -GAC, <i>rps12</i>)	IRb, IRa	45	31	P	IGS (<i>trnT</i> -GGU, <i>psbD</i>)	IGS (<i>trnT</i> -UGU, <i>trnL</i> -UAA)	LSC
16	31	P	IGS (<i>rps12</i> , <i>trnV</i> -GAC)	IGS (<i>trnV</i> -GAC, <i>rps12</i>)	IRb, IRa	46	31	F	IGS (<i>psaA</i> , <i>ycf3</i>)	IGS (<i>psaA</i> , <i>ycf3</i>)	LSC
17	31	F	IGS (<i>trnV</i> -GAC, <i>rps12</i>)	IGS (<i>trnV</i> -GAC, <i>rps12</i>)	IRa	47	31	R	IGS (<i>accD</i> , <i>psaI</i>)	IGS (<i>accD</i> , <i>psaI</i>)	LSC
18	37	P	IGS (<i>rpl22</i> , <i>rps19</i>)	IGS (<i>rpl22</i> , <i>rps19</i>)	LSC	48	31	F	IGS (<i>accD</i> , <i>psaI</i>)	<i>petD</i> (intron)	LSC
19	31	R	<i>petD</i> (intron)	<i>petD</i> (intron)	LSC	49	31	C	<i>petD</i> (intron)	<i>petD</i> (intron)	LSC
20	30	P	IGS (<i>psbI</i> , <i>trnS</i> -GCU)	<i>trnS</i> -GGA (CDS)	LSC	50	31	F	<i>ndhF</i> (CDS)	IGS (<i>ndhF</i> , <i>ycf1</i>)	SSC
21	30	F	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRb	51	30	C	IGS (<i>trnG</i> -UCC, <i>trnT</i> -GGU)	IGS (<i>trnT</i> -GGU, <i>psbD</i>)	LSC
22	30	P	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRb, IRa	52	30	R	IGS (<i>trnT</i> -GGU, <i>psbD</i>)	IGS (<i>trnT</i> -GGU, <i>psbD</i>)	LSC
23	30	P	<i>ycf2</i> (CDS)	<i>ycf2</i> (CDS)	IRb, IRa	53	30	C	<i>ycf3</i> (intron2)	IGS (<i>ndhC</i> , <i>trnV</i> -UAC)	LSC
24	32	P	IGS (<i>trnG</i> -UCC, <i>trnT</i> -GGU)	IGS (<i>trnG</i> -UCC, <i>trnT</i> -GGU)	LSC	54	30	C	IGS (<i>ndhC</i> , <i>trnV</i> -UAC)	IGS (<i>accD</i> , <i>psaI</i>)	LSC
25	32	P	IGS (<i>psbZ</i> , <i>trnG</i> -GCC)	IGS (<i>psbZ</i> , <i>trnG</i> -GCC)	LSC	55	30	F	IGS (<i>ndhC</i> , <i>trnV</i> -UAC)	IGS (<i>ndhC</i> , <i>trnV</i> -UAC)	LSC
26	32	P	IGS (<i>accD</i> , <i>psaI</i>)	<i>petD</i> (CDS)	LSC	56	30	F	IGS (<i>accD</i> , <i>psaI</i>)	IGS (<i>accD</i> , <i>psaI</i>)	LSC
27	32	R	<i>petD</i> (CDS)	<i>petD</i> (CDS)	LSC	57	30	R	IGS (<i>accD</i> , <i>psaI</i>)	IGS (<i>accD</i> , <i>psaI</i>)	LSC
28	32	F	IGS (<i>rrn4.5</i> , <i>rrn5</i>)	IGS (<i>rrn4.5</i> , <i>rrn5</i>)	IRb	58	30	R	IGS (<i>accD</i> , <i>psaI</i>)	<i>petD</i> (intron)	LSC
29	32	P	IGS (<i>rrn4.5</i> , <i>rrn5</i>)	IGS (<i>rrn5</i> , <i>rrn4.5</i>)	IRb, IRa	59	30	F	<i>petD</i> (intron)	<i>petD</i> (intron)	LSC
30	32	P	IGS (<i>rrn4.5</i> , <i>rrn5</i>)	IGS (<i>rrn5</i> , <i>rrn4.5</i>)	IRb, IRa						

F = forward, P = palindrome, R = reverse, C = complement, IGS = intergenic spacer.

Table 8. The simple sequence repeats in the *C. carinatum* Schousb cp genome.

Unit	Length	No.	Location	Region	Unit	Length	No.	Location	Region	Unit	Length	No.	Location	Region	
A	19	1	IGS (<i>rpl32, ndhF</i>)	SSC	T	22	1	IGS (<i>atpF, atpA</i>)	LSC	TA	14	1	IGS (<i>trnH-GUG, psbA</i>)	LSC	
	13	1	IGS (<i>psbE, petL</i>)	LSC		16	1	<i>rps16</i> (intron)	LSC		12	2	IGS (<i>trnT-GGU, psbD</i>)	LSC	
	12	1	IGS (<i>ycf1, rps15</i>)	SSC		14	3	IGS (<i>trnE-UUUC, rpoB</i>)	LSC					IGS (<i>rpl33, rps18</i>)	LSC
	11	6	IGS (<i>trnK-UUUU, rps16</i>)	LSC				IGS (<i>atpB, rbcL</i>)	LSC		10	1	<i>rpoC1</i> (exonII)	LSC	
			IGS (<i>trnR-UUCU, trnG-UCC</i>)	LSC				IGS (<i>rps8, rpl14</i>)	LSC	TG	10	1	IGS (<i>rps16, trnQ-UUUG</i>)	LSC	
			IGS (<i>psbZ, trnG-GCC</i>)	LSC		13	3	IGS (<i>ndhC, trnV-UAC</i>)	LSC	ATT	12	2	IGS (<i>cemA, petA</i>)	LSC	
			IGS (<i>psbE, petL</i>)	LSC				IGS (<i>rbcL, accD</i>)	LSC					IGS (<i>trnL-UAG, rpl32</i>)	SSC
			IGS (<i>rps18, rpl20</i>)	LSC				IGS (<i>psbB, psbT</i>)	LSC	GAA	15	1	<i>ycf1</i>	SSC	
			IGS (<i>petD, rpoA</i>)	LSC		12	5	IGS (<i>trnH-GUG, psbA</i>)	LSC	TTA	12	1	<i>ndhA</i> (intron)	SSC	
		10	10	<i>trnk-UUUU</i> (intron)	LSC			IGS (<i>trnD-GUC, trnY-GUA</i>)	LSC	TTC	12	1	<i>psbC</i>	LSC	
				<i>rpoB</i>	LSC			<i>ycf3</i> (intronII)	LSC	AATA	12	2	IGS (<i>trnK-UUUU, rps16</i>)	LSC	
				<i>rpoC1</i> (intron)	LSC			<i>clpP</i> (intronI)	LSC					<i>ndhA</i> (intron)	SSC
				<i>rpoC1</i> (exonII)	LSC			<i>ycf1</i>	LSC	ATAC	12	1	IGS (<i>trnF-GAA, ndhJ</i>)	LSC	
				<i>rpoC1</i> (exonII)	LSC		11	1	IGS (<i>petA, psbJ</i>)	LSC	ATTG	12	1	<i>ycf2</i>	IRa
				IGS (<i>psaA, ycf3</i>)	LSC		10	9	<i>trnk-UUUU</i> (intron)	LSC	ATTT	22	1	IGS (<i>atpF, atpA</i>)	LSC
				IGS (<i>ycf3, trnS-GGA</i>)	LSC				IGS (<i>psbM, trnD-GUC</i>)	LSC	CAAT	12	1	IGS (<i>ycf2, trnL-CAA</i>)	IRb
				IGS (<i>trnT-UGU, rnl-UAA</i>)	LSC				<i>rpoC1</i> (intron)	LSC	GATT	12	1	<i>ndhA</i> (intron)	SSC
				<i>psbT</i>	LSC				IGS (<i>atpI, atpH</i>)	LSC	TAGA	12	1	IGS (<i>rbcL, accD</i>)	LSC
				IGS (<i>rrn5, trnR-ACG</i>)	IRb				IGS (<i>atpI, atpH</i>)	LSC	TAAA	12	1	IGS (<i>atpI, atpH</i>)	LSC
	C	12	1	<i>rps16</i> (intron)	LSC				IGS (<i>atpA, trnR-UUCU</i>)	LSC	TAAT	12	1	IGS (<i>petN, psbM</i>)	LSC
	AT	14	1	<i>rps16</i> (intron)	LSC				<i>rpoA</i>	LSC	TATT	12	2	IGS (<i>rpl33, rps18</i>)	LSC
		12	1	IGS (<i>psbZ, trnG-GCC</i>)	LSC				IGS (<i>rpl14, rpl16</i>)	LSC				<i>ndhD</i>	SSC
		10	1	<i>rpoC2</i>	LSC				IGS (<i>trnR-ACG, rrn5</i>)	IRa	TTTC	16	1	<i>rpl16</i> (intron)	LSC
											AATTT	15	1	IGS (<i>ccsA, trnL-UAG</i>)	SSC

Table 9. The simple sequence repeats in the *K. indica* cp genome.

Unit	Length	No.	Location	Region	Unit	Length	No.	Location	Region	Unit	Length	No.	Location	Region	
A	18	1	IGS (<i>trnS</i> -UGA, <i>psbZ</i>)	LSC	AT	16	1	IGS (<i>psbZ</i> , <i>trnG</i> -GCC)	LSC	GAA	15	1	<i>ycf1</i>	SSC	
	12	1	IGS (<i>psbZ</i> , <i>trnG</i> -GCC)	LSC		10	7	<i>rps16</i> (intron)	LSC		12	1	<i>ycf1</i>	SSC	
	11	3	IGS (<i>trnQ</i> -UUIG, <i>psbK</i>)	LSC		10		<i>rpoC2</i>	LSC	TAT	21	1	IGS (<i>rps12</i> , <i>trnV</i> -GAC)	IRb	
	11		IGS (<i>psbM</i> , <i>trnD</i> -GUIC)	LSC		10		IGS (<i>psbZ</i> , <i>trnG</i> -GCC)	LSC		12	2	IGS (<i>trnT</i> -UGU, <i>trnL</i> -UAA)	LSC	
	11		IGS (<i>ycf4</i> , <i>cemA</i>)	LSC		10		IGS (<i>psbZ</i> , <i>trnG</i> -GCC)	LSC		12		IGS (<i>trnF</i> -GAA, <i>ndhJ</i>)	LSC	
	10	7	<i>rpoB</i>	LSC		10		IGS (<i>psbE</i> , <i>petL</i>)	LSC	TTA	15	2	IGS (<i>ndhC</i> , <i>trnV</i> -UAC)	LSC	
	10		<i>rpoC1</i> (exonII)	LSC		10		IGS (<i>petD</i> , <i>rpoA</i>)	LSC		15		IGS (<i>clpP</i> , <i>psbB</i>)	LSC	
	10		IGS (<i>petB</i> , <i>petD</i>)	LSC		10		<i>rpl16</i> (intron)	LSC		12	2	<i>petB</i> (intron)	LSC	
	10		<i>ndhB</i> (intron)	IRb	TA	23	1	<i>petD</i> (intron)	LSC		12		<i>petD</i> (intron)	LSC	
	10		<i>ndhA</i> (intron)	SSC		20	1	IGS (<i>accD</i> , <i>psaI</i>)	LSC		12	1	<i>psbC</i>	LSC	
	10		IGS (<i>trnL</i> -UAG, <i>rpl32</i>)	SSC		18	1	<i>petD</i> (intron)	LSC		12	1	IGS (<i>trnK</i> -UUU, <i>rps16</i>)	LSC	
	10		IGS (<i>rpl2</i> , <i>rps19</i>)	IRa		14	1	IGS (<i>accD</i> , <i>psaI</i>)	LSC		12	1	IGS (<i>trnG</i> -UCC, <i>trnT</i> -GGU)	LSC	
	T	18	1	IGS (<i>trnE</i> -UUIC, <i>rpoB</i>)	LSC		12	3	IGS (<i>trnT</i> -GGU, <i>psbD</i>)	LSC		12	1	IGS (<i>trnE</i> -UUIC, <i>rpoB</i>)	LSC
		17	1	<i>rpoA</i>	LSC		12		<i>petD</i> (intron)	LSC		12	1	IGS (<i>clpP</i> , <i>psbB</i>)	LSC
		14	2	IGS (<i>rpl20</i> , <i>rps12</i>)	LSC		12		IGS (<i>rpl22</i> , <i>rps19</i>)	LSC		12	1	IGS (<i>trnS</i> -GCU, <i>trnC</i> -GCA)	LSC
		14		<i>ndhA</i> (intron)	SSC		10	3	<i>rpoC1</i> (exonII)	LSC		12	1	IGS (<i>ndhI</i> , <i>ndhG</i>)	SSC
		13	1	IGS (<i>psbE</i> , <i>petL</i>)	LSC		10		IGS (<i>psaJ</i> , <i>rpl33</i>)	LSC	TATC	12	1	IGS (<i>psbA</i> , <i>trnK</i> -UUUU)	LSC
12		2	IGS (<i>atpF</i> , <i>atpA</i>)	LSC		10		<i>ndhA</i> (intron)	SSC	TATT	12	2	IGS (<i>rpl33</i> , <i>rps18</i>)	LSC	
12			<i>clpP</i> (intronII)	LSC	AAT	12	2	IGS (<i>trnR</i> -UCU, <i>trnG</i> -UCC)	LSC		12		IGS (<i>psaC</i> , <i>ndhD</i>)	SSC	
11		2	IGS (<i>atpB</i> , <i>rbcl</i>)	LSC		12		IGS (<i>trnT</i> -GGU, <i>psbD</i>)	LSC	TCTA	12	1	<i>ndhA</i> (intron)	SSC	
11			IGS (<i>psaI</i> , <i>ycf4</i>)	LSC	ATA	21	1	IGS (<i>trnV</i> -GAC, <i>rps12</i>)	IRa	TTCT	12	1	IGS (<i>trnG</i> -UCC, <i>trnT</i> -GGU)	LSC	
10		9	IGS (<i>trnC</i> -GCA, <i>petN</i>)	LSC		15	3	<i>ycf3</i> (intron)	LSC	TTTA	12	1	<i>petD</i> (intron)	LSC	
10			IGS (<i>rpoC2</i> , <i>rps2</i>)	LSC		15		IGS (<i>trnP</i> -UGG, <i>psaJ</i>)	LSC	TTTC	12	1	<i>trnK</i> -UUU (intron)	LSC	
10			IGS (<i>atpI</i> , <i>atpH</i>)	LSC		15		IGS (<i>trnL</i> -UAG, <i>rpl32</i>)	SSC	ATTAG	15	1	IGS (<i>rbcl</i> , <i>accD</i>)	LSC	
10			IGS (<i>ndhC</i> , <i>trnV</i> -UAC)	LSC		12	2	IGS (<i>trnG</i> -UCC, <i>trnT</i> -GGU)	LSC	TATAT	23	1	<i>petD</i> (intron)	LSC	
10			<i>clpP</i> (intronI)	LSC		12		<i>rpl16</i> (intron)	LSC		20	1	IGS (<i>accD</i> , <i>psaI</i>)	LSC	
10			IGS (<i>rps8</i> , <i>rpl14</i>)	LSC	ATT	15	1	IGS (<i>trnG</i> -UCC, <i>trnT</i> -GGU)	LSC		15	1	IGS (<i>accD</i> , <i>psaI</i>)	LSC	
10			IGS (<i>rps19</i> , <i>rpl2</i>)	IRb	TAA	15	1	IGS (<i>accD</i> , <i>psaI</i>)	LSC	TATTA	21	2	IGS (<i>rps12</i> , <i>trnV</i> -GAC)	IRb	
10			IGS (<i>psaC</i> , <i>ndhD</i>)	SSC		12	2	IGS (<i>trnE</i> -UUU, <i>rpoB</i>)	LSC		21		IGS (<i>trnV</i> -GAC, <i>rps12</i>)	IRa	
10		<i>ndhB</i> (intron)	IRa		12		IGS (<i>trnT</i> -GGU, <i>psbD</i>)	LSC	TCCTA	15	1	IGS (<i>rps4</i> , <i>trnT</i> -UGU)	LSC		

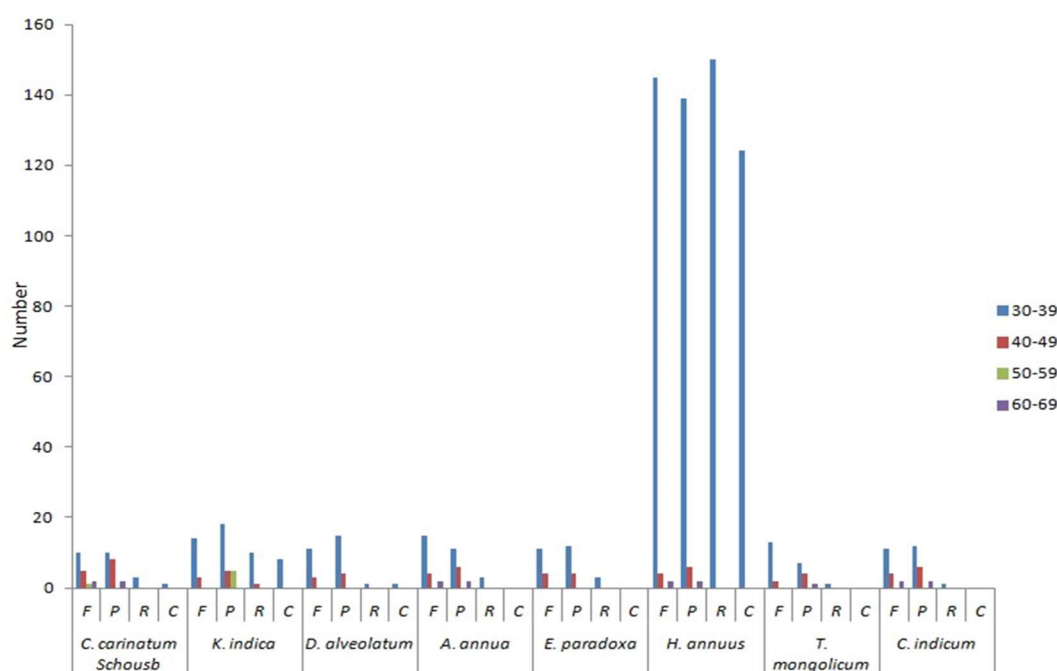


Figure 2. The repeat sequences of eight *Asteraceae* cp genomes. F (forward), P (palindrome), R (reverse), and C (complement) represent the repeat types. Different colours represent the repeats in different lengths.

Table 10. The distribution of SSRs present in the *Asteraceae* cp genomes.

Taxon	Genome Size (bp)	GC (%)	SSR Type						CDS			
			Mono	Di	Tri	Tetra	Penta	Hexa	Total	% ^a	No. ^b	% ^c
<i>C. carinatum</i> Schousb	149,752	37.47	43	8	5	13	1	0	70	51.6	12	17.1
<i>K. indica</i>	152,885	37.25	30	18	22	13	7	0	90	51.3	9	10
<i>D. alveolatum</i>	152,265	37.37	34	14	11	14	2	0	75	51.4	8	10.7
<i>A. annua</i>	150,952	37.48	39	10	4	15	3	0	71	52.1	10	14.1
<i>E. paradoxa</i>	151,837	37.59	37	4	4	5	0	1	51	51.5	14	27.5
<i>H. annuus</i>	151,104	37.62	40	4	4	4	0	0	52	51.2	14	26.9
<i>T. mongolicum</i>	151,451	37.67	14	5	3	6	1	1	30	51.3	11	36.7
<i>C. indicum</i>	150,972	37.48	38	10	4	14	1	0	67	48.8	7	10.4

CDS: protein-coding regions. ^a the percentage ratio of the total length of the CDS to the genome size. ^b the total number of SSRs in CDS. ^c the percentage ratio of the total number of SSRs in CDS to the total number of SSRs in the whole genome.

2.5. Comparison of the Whole cp Genome of Asteraceae

Using *C. carinatum* Schousb as a reference, mVISTA was used to compare the overall sequence of the cp genome of the eight *Asteraceae* species (Figure 3). It was found that the cp genome sequence of *C. carinatum* Schousb had the shortest size, whereas the *K. indica* cp genome had the longest size among the eight *Asteraceae* species. The difference of the sequence length was mainly due to the difference of length of the LSC region and there was no significant differences in the SSC and IRs regions' length. It is important to note that the SSC regions of *A. annua* and *T. mongolicum* are completely different from *C. carinatum* Schousb because of the inversion in most of the *Asteraceae* species, whereas the inversion region in LSC compared to other eudicots is relatively constant [13–25]. In order to illustrate this, we compared the cp genomes between 129 species of *Asteraceae* and some cp genomes of non-*Asteraceae* species using the mVISTA software to confirm the structural changes (Data not shown). Among them, only 16 species of *Asteraceae* (*Artemisia argyi*, *Artemisia capillaries*, *Artemisia frigida*, *Artemisia gmelinii*, *Artemisia montana*, *Aster altaicus*, *Carthamus tinctorius*, *Centaurea diffusa*, *Eclipta prostrate*, *Lactuca sativa*,

Saussurea chabyoungsanica, *Saussurea involucrate*, *Saussurea polylepis*, *Taraxacum mongolicum*, *Taraxacum officinale*, *Taraxacum platycarpum*) did not invert in the SSC region (Supplemental Table S1). However, the LSC region was generally inverted. In addition, the coding area is more conservative than the non-coding area [53,54].

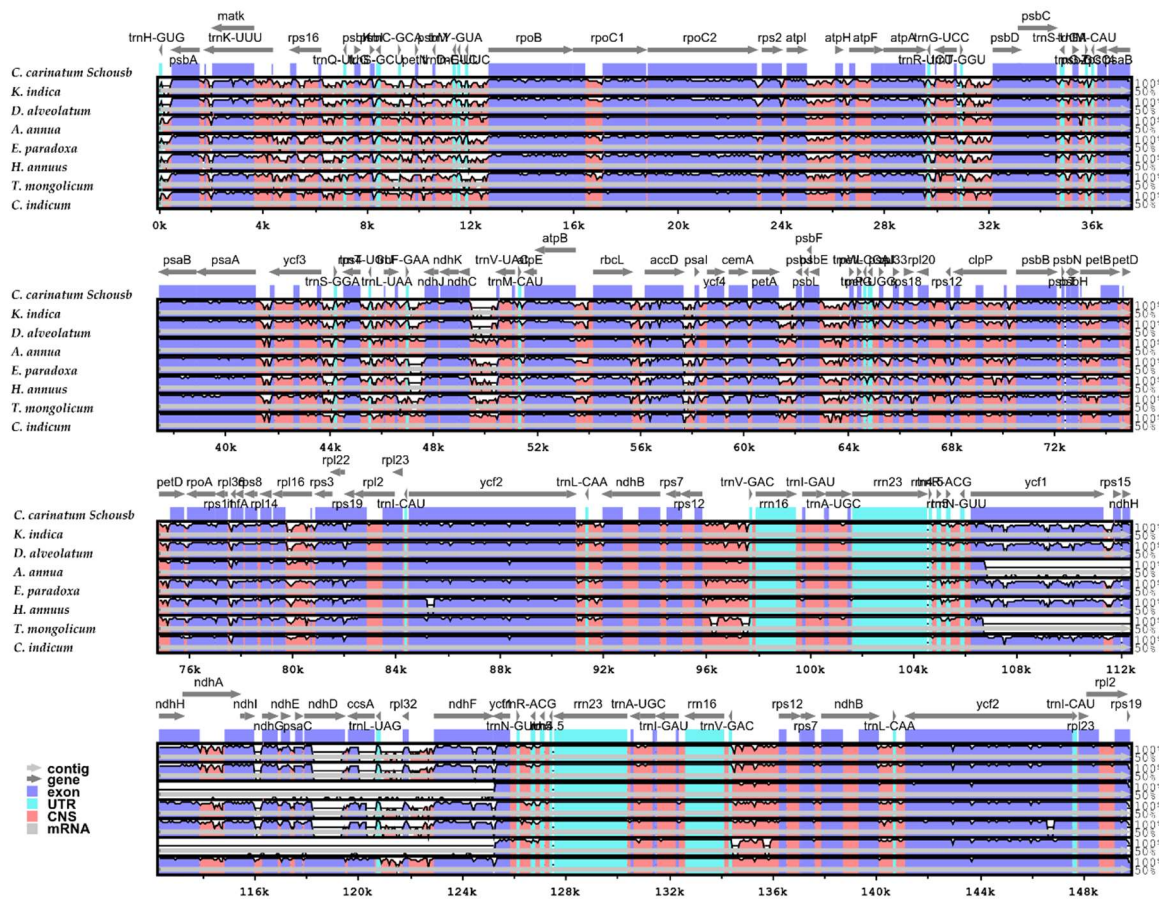


Figure 3. The comparison of eight *Asteraceae* cp genomes by using mVISTA. The grey arrows above the contrast indicate the direction of the gene translation. The y-axis represents the percent identity between 50% and 100%. Protein codes (exon), rRNA, tRNA and conserved non-coding sequence (CNS) are shown in different colors, respectively.

2.6. IR Contraction and Expansion

The IR contraction and expansion of *C. carinatum* Schousb and *K. indica* were analyzed by comparing the LSC/IRB/SSC/IRA boundary among the eight species of *Asteraceae* (Figure 4). For all of them, the junction of the IRB and LSC regions was connected by the *rps19* gene such that pseudogenes *rps19* in the IRA/LSC boundaries were usually found in the duplication of the 3' end of the *rps19* in IRA, in which the pseudogene *rps19* lengths ranged from 60 to 99 bp. The *ycf1* gene was located in both IRB/SSC and IRA/LSC boundaries. Additionally, 477 to 581 bp pseudogenes *ycf1* at the border of IRA/SSC were produced. Among these eight cp genomes, *C. carinatum* Schousb had the largest *ycf1* pseudogene (581 bp in length), whereas the inversion of the SSC region in *A. annua* and *T. mongolicum* caused the pseudogene *ycf1* to appear at the border of IRB/SSC. *T. mongolicum* and *C. indicum* showed the absence of the *ycf1* pseudogene and the *rps19* pseudogene. With the exception of *T. mongolicum*, all *ndhF* genes were located in the SSC region, which has a distance between 35–77 bp from the IRA border, and the *ndhF* gene of *C. carinatum* Schousb has the largest distance to IRA border. Furthermore, the *ndhF* gene of *T. mongolicum* is 10 bp across the IRB region and intersects with the pseudogene *ycf1*. The *trnH-GUG* genes are located in the LSC region between 0–13 bp from the IRA border and

C. carinatum Schousb has the largest one. Oddly, a relatively smaller IR size but longer pseudogene *ycf1* length were found in *C. carinatum* Schousb, which may be due to the occurrence of the contraction in the intergenic regions in *C. carinatum* Schousb [48].

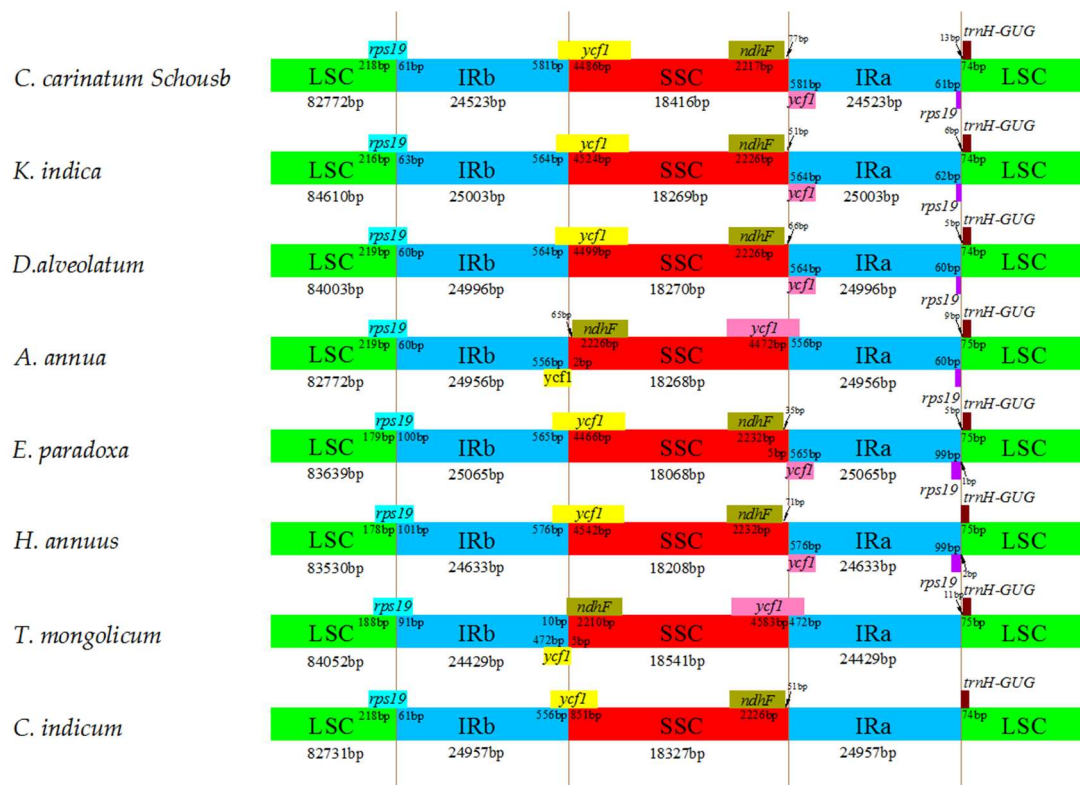


Figure 4. The comparison of the borders of the LSC, SSC, and IR regions among the eight *Asteraceae* cp genomes.

2.7. Phylogenetic Analysis and Barcoding

The cp genome is a useful resource and tool for taxonomy, determining evolutionary relationships within families, and DNA barcoding [42,55–59]. Here, for obtaining a useful barcoding marker and a reasonable phylogenetic status of *C. carinatum* Schousb and *K. indica*, we established the phylogenetic tree of 24 sample species using the Maximum Likelihood (ML) method by the alignment of 18 genes (*atpF*, *clpP*, *matK*, *ndhA*, *ndhB*, *ndhF*, *petB*, *petD*, *psaB*, *psbA*, *rbcL*, *rpl2*, *rpl16*, *rpoB*, *rpoC1*, *rps12*, *rps16*, *rps19*), respectively (Supplemental Figure S1). These 24 species include 18 *Asteraceae* species, four *Orchidaceae* species, one *Chenopodiaceae* species and one *Cruciferae* species. We compared six non-*Asteraceae* species with 18 *Asteraceae* species because the former are morphologically similar to *C. carinatum* Schousb and *K. indica*. From the results of the alignment, the *matK*, *ndhF* and *rbcL* genes are better than the others. There are eight nodes that had bootstrap values >90% when *matK* and *ndhF* were used for gene alignment. However, they both had three nodes with bootstrap values <40% and one node even had a bootstrap value <30%. As for the *rbcL* gene, 10 out of the 19 nodes had bootstrap values >90%. The minimum bootstrap value is 46% and the rest of the nodes had bootstrap values >50%. We show the phylogenetic tree by *rbcL* gene alignment in Figure 5, which shows that *C. carinatum* Schousb is a closely related species with *Chrysanthemum indicum* and that *K. indica* is closely related with the *Conyza bonariensis* species, which is consistent with their classification and morphology. Hence, the *rbcL* gene could be a good candidate gene to be used as a barcoding marker [51,52,54].

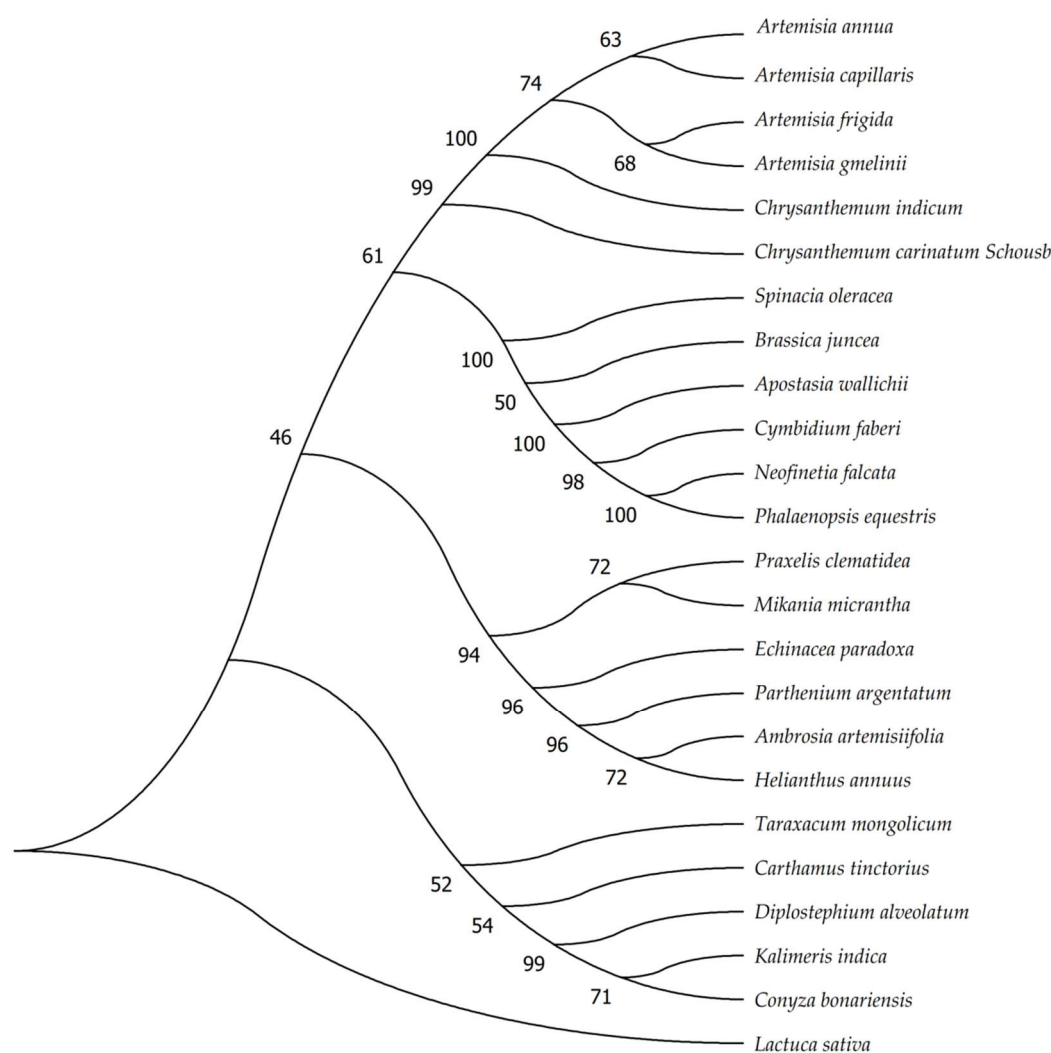


Figure 5. The molecular phylogenetic analysis of the cp protein-coding gene *rcbL* for 24 samples using the Maximum Likelihood method. The tree was constructed by using MEGA7. The stability of each tree node was tested by bootstrap analysis with 1000 replicates.

3. Discussion

We reported on the two genome sequence of *C. carinatum* Schousb and *K. indica*, which provides an important resource to study the evolutionary and inversion mechanism as well as the molecular barcoding of vegetables in the *Asteraceae* family. Although cp genomes of Angiosperms are well-conserved in the genomic structure, the inversion and IR expansion contraction occur frequently [7–11]. These results showed that the inversion of *trnC-GCA* to *trnG-UCC* in the LSC and the whole SSC occurred in these two species, which are congruent with most of the cp genomes of *Asteraceae* family [41,60–62]. Oddly, though some no-inversion occurs in the SSC region, the inversion of *trnC-GCA* to *trnG-UCC* in the LSC does not occur, which maybe could make amplification of *trnC-GCA* or *trnG-UCC* boundary regions a good resource for the molecular taxonomy of *Asteraceae*. Meanwhile, the SSC/IRA and LSC/IRB border extends into the *ycf1* and *rps19* with the subsequent formation of the *ycf1* and *rps19* pseudogenes, respectively [63,64]. The *ycf1* and *rps19* pseudogenes occur in most of the cp genomes, with deletion regions of similar size in species of the same family, but different deletion size between different families (Figure 4) [48].

Here, the analysis of the codon usage frequency and RSCU showed that leucine and isoleucine are the most popular amino acids in the cp genomes in the *C. carinatum* Schousb and *K. indica* as

Angiosperms [23,52,65–67]. Meanwhile, a significantly higher T bases appearance was indicated in the 3th CDS position than in the 2nd position or 1st position. The T bases appear at the end of most favorite synonymous codons (RSCU > 1.0), but the reverse occurs for the G base (Tables 1, 4 and 5). Consistent with most earlier reports about repeats and SSRs, the mononucleotide repeats with A/T repeats are more abundant, which may lead to AT richness in the Angiosperm cp genomes [68–70].

Here, we compared the whole cp genome of six species of *Asteraceae*, which revealed that the inversion is usual in most *Asteraceae* species, with some no-inversion occurring only in the SSC regions' inversions, but not in the inversions of *trnC-GCA* to *trnG-UCC* in the LSC. Among the 129 *Asteraceae* species that have been sequenced, inversion occurs in the LSC region. Whether or not the SSC region is inverted is closely related to the genus. For example, *Diplostephium* has the most sequenced species and all of them have inverted SSC regions. Inconsistent inversions within the genus only occurred in the *Aster* genus and the *Taraxacum* genus (Supplemental Table S1). This phenomenon may be a key feature of the cp genome of *Asteraceae*. It provides an insight into future evolution research.

Here, the *rbcL* genes represents a good diversity among some vegetables. The closer the relationship between species, the higher the sequence similarity of the genes. The phylogenetic tree of the *rbcL* gene among the *Asteraceae* species can obtain a better discrimination result. More importantly, it also could differentiate different family species, which illustrate that the *rbcL* gene is not only highly distinguished within the *Asteraceae* family, but also highly distinguished within most of vegetables. So, *rbcL* would be one of the best choices for DNA barcoding to distinguish between vegetables [71–73].

4. Materials and Methods

4.1. DNA Extraction and Sequencing

The total DNA was extracted from about 100 g fresh leaves of *C. carinatum* Schousb and *K. indica* using the CTAB method [74]. The total DNA quantity was evaluated by the value of the ratio of absorbance measurements at 260 nm and 280 nm (A260/A280) using Nanodrop2000 (Thermo Fisher Scientific, Waltham, MA, USA), whereas visual assessment of the DNA size and integrity was performed using gel electrophoresis. The DNA was sheared to fragments of 300~500 bp. Paired-end libraries were prepared with the TruSeq™ DNA Sample Prep Kit and the TruSeq PE Cluster Kit. The genome was then sequenced using the HiSeq4000 platform (Illumina, Santiago, CA, USA).

4.2. Genome Assembly

The assembly of the cp genome of *C. carinatum* Schousb and *K. indica* were first carried out through the error correction and production of initial contigs by the GS FLX De Novo Assembler Software (Newbler V2.6). PCR amplification and Sanger sequencing were performed to verify the four-f- junction regions between the IRs and the LSC/SSC. The final cp genome of *C. carinatum* Schousb and *K. indica* were submitted to the Genbank with the accession number MG710386 and MG710387, respectively.

4.3. Gene Annotation and Codon Usage Analysis

The cp genome was annotated using BLAST [75] and DOGMA with manual corrections [76]. The tRNAscan-SE was used to identify the tRNA genes [77]. OGDRAW was used to draw the circular genome map [78]. MEGA5 was used to analyze the characteristics of the variations in the synonymous codon [79]. The relative synonymous codon usage values (RSCU), codon usage, and GC content were also determined by MEGA7 [80].

4.4. Repeat Structure and Single Sequence Repeats (SSRs) Analysis

Analysis of tandem repeats with more than 30 bp and a minimum of 90% sequence (forward, palindromic, reverse, and complement) and single sequence repeats (SSRs) was performed by REPuter [81] and MISA [82], respectively, with the same parameters as described in Ni et al. [53].

4.5. Comparative Genome Analysis of the *C. carinatum* Schousb and *K. indica* Genomes

Comparison of the overall cp genome of *C. carinatum* Schousb and *K. indica* with six cp genomes of *Asteraceae* was performed by mVISTA [83,84] using the annotation of *C. carinatum* Schousb as a reference.

4.6. Phylogenetic Analysis

A total of 24 complete cp genome sequences were downloaded from the NCBI Organelle Genome Resources database. MEGA7 [80] was used to construct the evolutionary tree of the cp protein-coding gene *rcbL* of 24 samples by the Maximum Likelihood method.

5. Conclusions

In this study, the complete cp genome of *C. carinatum* Schousb and *K. indica* were reported and analyzed for the first time. Both of them are key traditional Chinese medicines and edible vegetables. Comparing them with other *Asteraceae* species, the cp genome of these two species display two inversions, one is *trnC-GCA* to *trnC-UCC* in the LSC and other is the whole SSC region. We found that most of the inversions of *trnC-GCA* to *trnC-UCC* almost happened in every cp genome of the *Asteraceae* species. Seventy and ninety simple sequence repeats (SSRs) were present in the cp genome of *C. carinatum* Schousb and *K. indica*, respectively. Certainly, these results provide good chances for developing barcoding molecular markers for different families distinguish, which can be obtained through combination by different barcoding markers or through the boundary region marker. The *rcbL* is a good barcoding marker. Meanwhile, these results would be useful for the evolutionary study of *C. carinatum* Schousb and *K. indica*, and might also contribute to the genetics and barcoding of easily-confused leafy vegetables.

Supplementary Materials: Supplementary materials are available online.

Author Contributions: X.L. conceived and designed the experiments; B.Z., H.Y., Y.L., Q.Y., Y.L. (Yuzhuo Lu) and Y.G. performed the experiments; X.L., B.Z., and Y.L. analysed the data; X.L. and B.Z. wrote the paper. All authors read and approved the final manuscript.

Funding: This research was funded by the Natural Science Foundation of Tianjin (17JCZDJC34300) and Technology Pillar Program during the Twelve Five-year Plan Period of China (2015BAD16B02).

Acknowledgments: This work was supported by grants from the Natural Science Foundation of Tianjin (17JCZDJC34300) and the Key Projects in the National Science and Technology Pillar Program during the Twelve Five-year Plan Period of China (2015BAD16B02).

Conflicts of Interest: There is no conflict of interest.

References

1. Cazzonelli, C.I. Carotenoids in nature: Insights from plants and beyond. *Funct. Plant Biol.* **2011**, *38*, 833–847. [[CrossRef](#)]
2. Bobik, K.; Burch-Smith, T.M. Chloroplast signaling within, between and beyond cells. *Front. Plant Sci.* **2015**, *6*, 781. [[CrossRef](#)] [[PubMed](#)]
3. Daniell, H.; Lin, C.; Yu, M.; Chang, W. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol.* **2016**, *17*, 134. [[CrossRef](#)] [[PubMed](#)]
4. Baczkiewicz, A.; Szczecinska, M.; Sawicki, J.; Stebel, A.; Buczkowska, K. DNA barcoding, ecology and geography of the cryptic species of *Aneura pinguis* and their relationships with *Aneura maxima* and *Aneura mirabilis* (Metzgeriales, Marchantiophyta). *PLoS ONE* **2017**, *12*, e0188837. [[CrossRef](#)] [[PubMed](#)]
5. Kang, Y.; Deng, Z.; Zang, R.; Long, W. DNA barcoding analysis and phylogenetic relationships of tree species in tropical cloud forests. *Sci. Rep.* **2017**, *7*, 12564. [[CrossRef](#)] [[PubMed](#)]
6. Song, Y.; Chen, Y.; Lv, J.; Xu, J.; Zhu, S.; Li, M.; Chen, N. Development of Chloroplast Genomic Resources for *Oryza* Species Discrimination. *Front. Plant Sci.* **2017**, *8*, 1854. [[CrossRef](#)] [[PubMed](#)]
7. Stegemann, S.; Keuthe, M.; Greiner, S.; Bock, R. Horizontal transfer of chloroplast genomes between plant species. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 2434–2438. [[CrossRef](#)] [[PubMed](#)]

8. Rogalski, M.; Vieira, L.D.N.; Fraga, H.P.; Guerra, M.P. Plastid genomics in horticultural species: Importance and applications for plant population genetics, evolution, and biotechnology. *Front. Plant Sci.* **2015**, *6*, 586. [[CrossRef](#)] [[PubMed](#)]
9. Ruck, E.C.; Linard, S.R.; Nakov, T.; Theriot, E.C.; Alverson, A.J. Hoarding and horizontal transfer led to an expanded gene and intron repertoire in the plastid genome of the diatom, *Toxarium undulatum* (Bacillariophyta). *Curr. Genet.* **2017**, *63*, 499–507. [[CrossRef](#)] [[PubMed](#)]
10. Kawabe, A.; Nukii, H.; Furihata, H.Y. Exploring the History of Chloroplast Capture in Arabis Using Whole Chloroplast Genome Sequencing. *Int. J. Mol. Sci.* **2018**, *19*, 602. [[CrossRef](#)] [[PubMed](#)]
11. Wang, Z.; Peng, H.; Kilian, N. Molecular Phylogeny of the *Lactuca* Alliance (Cichorieae Subtribe Lactucinae, Asteraceae) with Focus on Their Chinese Centre of Diversity Detects Potential Events of Reticulation and Chloroplast Capture. *PLoS ONE* **2013**, *8*, e82692. [[CrossRef](#)] [[PubMed](#)]
12. Wang, M.; Cui, L.; Feng, K.; Deng, P.; Du, X.; Wan, F.; Song, W.; Nie, X. Comparative Analysis of Asteraceae Chloroplast Genomes: Structural Organization, RNA Editing and Evolution. *Plant. Mol. Biol. Rep.* **2015**, *33*, 1526–1538. [[CrossRef](#)]
13. 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](#)]
14. Jansen, R.K.; Palmer, J.D. Chloroplast DNA from lettuce and Barnadesia (Asteraceae): Structure, gene localization, and characterization of a large inversion. *Curr. Genet.* **1987**, *11*, 553–564. [[CrossRef](#)]
15. Kim, K.J.; Choi, K.S.; Jansen, R.K. Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). *Mol. Biol. Evol.* **2005**, *22*, 1783–1792. [[CrossRef](#)] [[PubMed](#)]
16. Guo, X.; Castillo-Ramirez, S.; Gonzalez, V.; Bustos, P.; Fernandez-Vazquez, J.L.; Santamaria, R.I.; Arellano, J.; Cevallos, M.A.; Davila, G. Rapid evolutionary change of common bean (*Phaseolus vulgaris* L.) plastome, and the genomic diversification of legume chloroplasts. *BMC Genom.* **2007**, *8*, 228. [[CrossRef](#)] [[PubMed](#)]
17. Walker, J.F.; Zanis, M.J.; Emery, N.C. Comparative analysis of complete chloroplast genome sequence and inversion variation in *Lasthenia burkei* (Madieae, Asteraceae). *Am. J. Bot.* **2014**, *101*, 722–729. [[CrossRef](#)] [[PubMed](#)]
18. Choi, K.S.; Park, S. The complete chloroplast genome sequence of *Aster spathulifolius* (Asteraceae); genomic features and relationship with Asteraceae. *Gene* **2015**, *572*, 214–221. [[CrossRef](#)] [[PubMed](#)]
19. Curci, P.L.; De Paola, D.; Danzi, D.; Vendramin, G.G.; Sonnante, G. Complete Chloroplast Genome of the Multifunctional Crop Globe Artichoke and Comparison with Other Asteraceae. *PLoS ONE* **2015**, *10*, e0120589. [[CrossRef](#)] [[PubMed](#)]
20. Schwarz, E.N.; Ruhlman, T.A.; Sabir, J.S.M.; Hajrah, N.H.; Alharbi, N.S.; Al-Malki, A.L.; Bailey, C.D.; Jansen, R.K. Plastid genome sequences of legumes reveal parallel inversions and multiple losses of *rps16* in papilionoids. *J. Syst. Evol.* **2015**, *53*, 458–468. [[CrossRef](#)]
21. Hernandez, L.F.; Rosetti, M.V. Is the abaxial palisade parenchyma in phyllaries of the sunflower (*Helianthus annuus* L.) capitulum a missing trait in modern genotypes? *Phyton Int. J. Exp. Bot.* **2016**, *85*, 291–296.
22. Raman, G.; Park, S. The Complete Chloroplast Genome Sequence of *Ampelopsis*: Gene Organization, Comparative Analysis, and Phylogenetic Relationships to Other Angiosperms. *Front. Plant Sci.* **2016**, *7*, 341. [[CrossRef](#)] [[PubMed](#)]
23. 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](#)]
24. Xie, Q.; Shen, K.; Hao, X.; Phan, N.N.; Bui, T.N.H.; Chen, C.; Zhu, C.; Lin, Y.; Hsiao, C. The complete chloroplast genome of Tianshan Snow Lotus (*Saussurea involucreata*), a famous traditional Chinese medicinal plant of the family Asteraceae. *Mitochondrial DNA Part A* **2017**, *28*, 294–295. [[CrossRef](#)] [[PubMed](#)]
25. Salih, R.H.M.; Majesky, L.; Schwarzacher, T.; Gornall, R.; Heslop-Harrison, P. Complete chloroplast genomes from apomictic *Taraxacum* (Asteraceae): Identity and variation between three microspecies. *PLoS ONE* **2017**, *12*, e168008. [[CrossRef](#)] [[PubMed](#)]
26. Kim, J.; Park, J.Y.; Lee, Y.S.; Woo, S.M.; Park, H.; Lee, S.; Kang, J.H.; Lee, T.J.; Sung, S.H.; Yang, T. The complete chloroplast genomes of two *Taraxacum* species, *T. platycarpum* Dahlst. and *T. mongolicum* Hand.-Mazz. (Asteraceae). *Mitochondrial DNA PART B Res.* **2016**, *1*, 412. [[CrossRef](#)]

27. Smith, D.R.; Keeling, P.J. Mitochondrial and plastid genome architecture: Reoccurring themes, but significant differences at the extremes. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 10177–10184. [[CrossRef](#)] [[PubMed](#)]
28. Chase, M.W.; Reveal, J.L. A phylogenetic classification of the land plants to accompany APG III. *Bot. J. Linn. Soc.* **2009**, *161*, 122–127. [[CrossRef](#)]
29. Bremer, B.; Bremer, K.; Chase, M.W.; Fay, M.F.; Reveal, J.L.; Soltis, D.E.; Soltis, P.S.; Stevens, P.F.; Anderberg, A.A.; Moore, M.J.; et al. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* **2009**, *161*, 105–121.
30. Byng, J.W.; Chase, M.W.; Christenhusz, M.J.M.; Fay, M.F.; Judd, W.S.; Mabberley, D.J.; Sennikov, A.N.; Soltis, D.E.; Soltis, P.S.; Stevens, P.F.; et al. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* **2016**, *181*, 1–20.
31. Xu, W.; Gong, X.; Zhou, X.; Zhao, C.; Chen, H. Chemical constituents and bioactivity of *Kalimeris indica*. *China J. Chin. Mater. Med.* **2010**, *35*, 3172–3174.
32. Wang, G.; Liu, J.; Zhang, C.; Wang, Z.; Cai, B.; Wang, G. Study on Chemical Constituents of *Kalimeris indica*. *J. Chin. Med. Mater.* **2015**, *38*, 81–84.
33. Fan, G.; Kim, S.; Han, B.H.; Han, Y.N. Glyceroglycolipids, a novel class of platelet-activating factor antagonists from *Kalimeris indica*. *Phytochem. Lett.* **2008**, *1*, 207–210. [[CrossRef](#)]
34. Zhong, R.; Xu, G.; Wang, Z.; Wang, A.; Guan, H.; Li, J.; He, X.; Liu, J.; Zhou, M.; Li, Y.; et al. Identification of anti-inflammatory constituents from *Kalimeris indica* with UHPLC-ESI-Q-TOF-MS/MS and GC-MS. *J. Ethnopharmacol.* **2015**, *165*, 39–45. [[CrossRef](#)] [[PubMed](#)]
35. Wang, G.; Yu, Y.; Wang, Z.; Cai, B.; Zhou, Z.; Wang, G.; Liu, J. Two new terpenoids from *Kalimeris indica*. *Nat. Prod. Res.* **2017**, *31*, 2348–2353. [[CrossRef](#)] [[PubMed](#)]
36. Kim, J.; Choi, J.N.; Ku, K.M.; Kang, D.; Kim, J.S.; Park, J.H.Y.; Lee, C.H. A Correlation between Antioxidant Activity and Metabolite Release during the Blanching of *Chrysanthemum coronarium* L. *Biosci. Biotechnol. Biochem.* **2011**, *75*, 674–680. [[CrossRef](#)] [[PubMed](#)]
37. Polatoglu, K.; Karakoc, O.C.; Demirci, B.; Baser, K.H.C. Chemical composition and insecticidal activity of edible garland (*Chrysanthemum coronarium* L.) essential oil against the granary pest *Sitophilus granarius* L. (Coleoptera). *J. Essent. Oil Res.* **2018**, *30*, 120–130. [[CrossRef](#)]
38. Flamini, G.; Cioni, P.L.; Morelli, I. Differences in the fragrances of pollen, leaves, and floral parts of garland (*Chrysanthemum coronarium*) and composition of the essential oils from flowerheads and leaves. *J. Agric. Food Chem.* **2003**, *51*, 2267–2271. [[CrossRef](#)] [[PubMed](#)]
39. Abd-Alla, H.I.; Albalawy, M.A.; Aly, H.F.; Shalaby, N.M.M.; Shaker, K.H. Flavone Composition and Antihypercholesterolemic and Antihyperglycemic Activities of *Chrysanthemum coronarium* L. *Z Naturforsch. C* **2014**, *69*, 199–208. [[CrossRef](#)] [[PubMed](#)]
40. Song, M.; Yang, H.; Jeong, T.; Kim, K.; Baek, N. Heterocyclic compounds from *Chrysanthemum coronarium* L. and their inhibitory activity on hACAT-1, hACAT-2, and LDL-oxidation. *Arch. Pharm. Res.* **2008**, *31*, 573–578. [[CrossRef](#)] [[PubMed](#)]
41. Vargas, O.M.; Ortiz, E.M.; Simpson, B.B. Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (*Asteraceae*: *Astereae*: *Diplostephium*). *New Phytol.* **2017**, *214*, 1736–1750. [[CrossRef](#)] [[PubMed](#)]
42. Fu, Z.; Jiao, B.; Nie, B.; Zhang, G.; Gao, T. A comprehensive generic-level phylogeny of the sunflower family: Implications for the systematics of Chinese *Asteraceae*. *J. Syst. Evol.* **2016**, *54*, 416–437. [[CrossRef](#)]
43. Muse, S.V. Examining rates and patterns of nucleotide substitution in plants. *Plant Mol. Biol.* **2000**, *42*, 25–43. [[CrossRef](#)] [[PubMed](#)]
44. Muse, S.V.; Gaut, B.S. Comparing patterns of nucleotide substitution rates among chloroplast loci using the relative ratio test. *Genetics* **1997**, *146*, 393–399. [[PubMed](#)]
45. Wakeley, J. Substitution-rate variation among sites and the estimation of transition bias. *Mol. Biol. Evol.* **1994**, *11*, 436–442. [[PubMed](#)]
46. Morton, B.R. Codon use and the rate of divergence of land plant chloroplast genes. *Mol. Biol. Evol.* **1994**, *11*, 231–238. [[PubMed](#)]
47. Liu, X.; Yang, H.; Zhao, J.; Zhou, B.; Li, T.; Xiang, B. The complete chloroplast genome sequence of the folk medicinal and vegetable plant purslane (*Portulaca oleracea* L.). *J. Hortic. Sci. Biotechnol.* **2017**, 1–10. [[CrossRef](#)]

48. Xia Liu, Y.L.H.Y. Chloroplast Genome of the Folk Medicine and Vegetable Plant *Talinum paniculatum* (Jacq.) Gaertn.: Gene Organization, Comparative and Phylogenetic Analysis. *Molecules* **2018**, *23*, 857. [[CrossRef](#)] [[PubMed](#)]
49. 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](#)]
50. Zuo, L.; Shang, A.; Zhang, S.; Yu, X.; Ren, Y.; Yang, M.; Wang, J. The first complete chloroplast genome sequences of *Ulmus* species by de novo sequencing: Genome comparative and taxonomic position analysis. *PLoS ONE* **2017**, *12*, e0171264. [[CrossRef](#)] [[PubMed](#)]
51. Jian, H.; Zhang, Y.; Yan, H.; Qiu, X.; Wang, Q.; Li, S.; Zhang, S. The Complete Chloroplast Genome of a Key Ancestor of Modern Roses, *Rosa chinensis* var. *spontanea*, and a Comparison with Congeneric Species. *Molecules* **2018**, *23*, 389. [[CrossRef](#)] [[PubMed](#)]
52. Li, Z.; Saina, J.K.; Gichira, A.W.; Kyalo, C.M.; Wang, Q.; Chen, J. Comparative Genomics of the Balsaminaceae Sister Genera *Hydrocera triflora* and *Impatiens pinfanensis*. *Int. J. Mol. Sci.* **2018**, *19*, 319. [[CrossRef](#)] [[PubMed](#)]
53. Ni, L.; Zhao, Z.; Xu, H.; Chen, S.; Dorje, G. The complete chloroplast genome of *Gentiana straminea* (Gentianaceae), an endemic species to the Sino-Himalayan subregion. *Gene* **2016**, *577*, 281–288. [[CrossRef](#)] [[PubMed](#)]
54. Khakhlova, O.; Bock, R. Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J.* **2006**, *46*, 85–94. [[CrossRef](#)] [[PubMed](#)]
55. Moore, A.J.; Bartoli, A.; Tortosa, R.D.; Baldwin, B.G. Phylogeny, biogeography, and chromosome evolution of the amphitropical genus *Grindelia* (Asteraceae) inferred from nuclear ribosomal and chloroplast sequence data. *Taxon* **2012**, *61*, 211–230.
56. Urbatsch, L.E.; Roberts, R.P.; Karaman, V. Phylogenetic evaluation of *Xylothamia*, *Gundlachia*, and related genera (Asteraceae, Astereae) based on ETS and ITS nrDNA sequence data. *Am. J. Bot.* **2003**, *90*, 634–649. [[CrossRef](#)] [[PubMed](#)]
57. Liu, J.Q.; Gao, T.G.; Chen, Z.D.; Lu, A.M. Molecular phylogeny and biogeography of the Qinghai-Tibet Plateau endemic *Nannoglottis* (Asteraceae). *Mol. Phylogenet. Evol.* **2002**, *23*, 307–325. [[CrossRef](#)]
58. Bayer, R.J.; Cross, E.W. A reassessment of tribal affinities of the enigmatic genera *Printzia* and *Isoetopsis* (Asteraceae), based on three chloroplast sequences. *Aust. J. Bot.* **2002**, *50*, 677–686. [[CrossRef](#)]
59. Eldenas, P.; Kallersjo, M.; Anderberg, A.A. Phylogenetic placement and circumscription of tribes Inuleae s. str. and Plucheeae (Asteraceae): Evidence from sequences of chloroplast gene *ndhF*. *Mol. Phylogenet. Evol.* **1999**, *13*, 50–58. [[CrossRef](#)] [[PubMed](#)]
60. Dempewolf, H.; Kane, N.C.; Ostevik, K.L.; Geleta, M.; Barker, M.S.; Lai, Z.; Stewart, M.L.; Bekele, E.; Engels, J.M.M.; Cronk, Q.C.B.; et al. Establishing genomic tools and resources for *Guizotia abyssinica* (L.f.) Cass.—the development of a library of expressed sequence tags, microsatellite loci, and the sequencing of its chloroplast genome. *Mol. Ecol. Resour.* **2010**, *10*, 1048–1058. [[CrossRef](#)] [[PubMed](#)]
61. Bock, D.G.; Kane, N.C.; Ebert, D.P.; Rieseberg, L.H. Genome skimming reveals the origin of the Jerusalem Artichoke tuber crop species: Neither from Jerusalem nor an artichoke. *New Phytol.* **2014**, *201*, 1021–1030. [[CrossRef](#)] [[PubMed](#)]
62. Zhang, N.; Erickson, D.L.; Ramachandran, P.; Ottesen, A.R.; Timme, R.E.; Funk, V.A.; Luo, Y.; Handy, S.M. An analysis of Echinacea chloroplast genomes: Implications for future botanical identification. *Sci. Rep.* **2017**, *7*, 216. [[CrossRef](#)] [[PubMed](#)]
63. Lee, M.; Park, J.; Lee, H.; Sohn, S.; Lee, J. Complete chloroplast genomic sequence of *Citrus platymamma* determined by combined analysis of Sanger and NGS data. *Hortic. Environ. Biotechnol.* **2015**, *56*, 704–711. [[CrossRef](#)]
64. Su, H.; Hogenhout, S.A.; Al-Sadi, A.M.; Kuo, C. Complete Chloroplast Genome Sequence of Omani Lime (*Citrus aurantiifolia*) and Comparative Analysis within the Rosids. *PLoS ONE* **2014**, *9*, e113049. [[CrossRef](#)] [[PubMed](#)]
65. Yang, Y.; Zhu, J.; Feng, L.; Zhou, T.; Bei, G.; Yang, J.; Zhao, G. Plastid Genome Comparative and Phylogenetic Analyses of the Key Genera in Fagaceae: Highlighting the Effect of Codon Composition Bias in Phylogenetic Inference. *Front. Plant Sci.* **2018**, *9*, 82. [[CrossRef](#)] [[PubMed](#)]

66. 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](#)]
67. Wang, W.; Yu, H.; Wang, J.; Lei, W.; Gao, J.; Qiu, X.; Wang, J. The Complete Chloroplast Genome Sequences of the Medicinal Plant *Forsythia suspensa* (Oleaceae). *Int. J. Mol. Sci.* **2017**, *18*, 2288. [[CrossRef](#)] [[PubMed](#)]
68. Saina, J.K.; Gichira, A.W.; Li, Z.; Hu, G.; Wang, Q.; Liao, K. The complete chloroplast genome sequence of *Dodonaea viscosa*: Comparative and phylogenetic analyses. *Genetica* **2018**, *146*, 101–113. [[CrossRef](#)] [[PubMed](#)]
69. Zhou, J.; Chen, X.; Cui, Y.; Sun, W.; Li, Y.; Wang, Y.; Song, J.; Yao, H. Molecular Structure and Phylogenetic Analyses of Complete Chloroplast Genomes of Two Aristolochia Medicinal Species. *Int. J. Mol. Sci.* **2017**, *18*, 1839. [[CrossRef](#)] [[PubMed](#)]
70. Zhou, T.; Chen, C.; Wei, Y.; Chang, Y.; Bai, G.; Li, Z.; Kanwal, N.; Zhao, G. Comparative Transcriptome and Chloroplast Genome Analyses of Two Related *Dipteronia* Species. *Front. Plant Sci.* **2016**, *7*, 1512. [[CrossRef](#)] [[PubMed](#)]
71. Mosa, K.A.; Soliman, S.; El-Keblawy, A.; Ali, M.A.; Hassan, H.A.; Tamim, A.A.B.; Al-Ali, M.M. Using DNA barcoding to detect adulteration in different herbal plant-based products in the United Arab Emirates: Proof of concept and validation. *Recent Pat. Food Nutr. Agric.* **2018**. [[CrossRef](#)] [[PubMed](#)]
72. Mishra, P.; Shukla, A.K.; Sundaresan, V. Candidate DNA Barcode Tags Combined with High Resolution Melting (Bar-HRM) Curve Analysis for Authentication of *Senna alexandrina* Mill. with Validation in Crude Drugs. *Front. Plant Sci.* **2018**, *9*, 283. [[CrossRef](#)] [[PubMed](#)]
73. Viljoen, E.; Odeny, D.A.; Coetzee, M.P.A.; Berger, D.K.; Rees, D.J.G. Application of Chloroplast Phylogenomics to Resolve Species Relationships Within the Plant Genus *Amaranthus*. *J. Mol. Evol.* **2018**, *86*, 216–239. [[CrossRef](#)] [[PubMed](#)]
74. Allen, G.C.; Flores-Vergara, M.A.; Krasynanski, S.; Kumar, S.; Thompson, W.F. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* **2006**, *1*, 2320–2325. [[CrossRef](#)] [[PubMed](#)]
75. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST plus: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [[CrossRef](#)] [[PubMed](#)]
76. Wyman, S.K.; Jansen, R.K.; Boore, J.L. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **2004**, *20*, 3252–3255. [[CrossRef](#)] [[PubMed](#)]
77. 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*, W686–W689. [[CrossRef](#)] [[PubMed](#)]
78. 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*, 267–274. [[CrossRef](#)] [[PubMed](#)]
79. 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](#)]
80. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)] [[PubMed](#)]
81. Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642. [[CrossRef](#)] [[PubMed](#)]
82. Beier, S.; Thiel, T.; Muench, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [[CrossRef](#)] [[PubMed](#)]
83. Kurtz, S.; Phillippy, A.; Delcher, A.L.; Smoot, M.; Shumway, M.; Antonescu, C.; Salzberg, S.L. Versatile and open software for comparing large genomes. *Genome Biol.* **2004**, *5*.
84. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **2004**, *32*, W273–W279. [[CrossRef](#)] [[PubMed](#)]

Sample Availability: Samples of *C. carinatum* Schousb and *K. indica* are available from the authors.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).