

## 葵花大蓟叶绿体基因组结构特征及系统发育分析

郑长远<sup>1,2</sup>, 苏旭<sup>1,2,3</sup>, 刘玉萍<sup>1,2\*</sup>, 刘涛<sup>1,2</sup>, 王亚男<sup>1,2</sup>, 苏丹丹<sup>1,2</sup>,  
张雨<sup>1,2</sup>, 秦娜<sup>1,2</sup>

(1. 青海师范大学生命科学学院, 青海 西宁 810008; 2. 青海师范大学, 青海省青藏高原药用动植物资源  
重点实验室, 青海 西宁 810008; 3. 青海师范大学高原科学与可持续发展研究院, 青海 西宁 810016)

**摘要:** 葵花大蓟 (*Cirsium souliei*) 是菊科 (Compositae)、蓟属 (*Cirsium*) 的一种多年生药用草本植物, 具有重要的药用和生态价值。基于高通量测序技术, 本研究通过对葵花大蓟的叶绿体基因组测序、组装、注释和特征分析, 结果发现葵花大蓟叶绿体基因组全长 152 470 bp, 为典型的环状四分体结构, GC 含量为 37.7%; 叶绿体基因组共编码 134 个基因, 其中蛋白编码基因 89 个, tRNA 基因 37 个, rRNA 基因 8 个; 叶绿体基因组共检测到 188 个简单重复序列 (simple sequence repeats, SSR) 位点, 以单核苷酸重复为主; 同时, 密码子偏好性分析显示, 亮氨酸 (Leu) 是使用频率最高的氨基酸 (10.51%), 同义密码子相对使用频率 (relative synonymous codon usage, RSCU) 值大于 1 的密码子有 30 个且以 A/U 结尾为主。此外, 基于叶绿体基因组构建系统发育树, 结果显示管状花亚科 (Carduoideae) 46 种植物中, 葵花大蓟与蓟属的翼蓟 (*C. vulgare*) 的亲缘关系最近, 支持率高达 100%, 旨在为葵花大蓟遗传多样性、群体遗传结构、系统进化及物种形成机制提供理论依据。

**关键词:** 叶绿体基因组; 系统发育; 简单重复序列; IR 区边界分析; 葵花大蓟

中图分类号: R931 文献标识码: A 文章编号: 0513-4870(2022)07-2224-10

## Characteristics of complete chloroplast genome and phylogenetic analysis of *Cirsium souliei* (Asteraceae)

ZHENG Chang-yuan<sup>1,2</sup>, SU Xu<sup>1,2,3</sup>, LIU Yu-ping<sup>1,2\*</sup>, LIU Tao<sup>1,2</sup>, WANG Ya-nan<sup>1,2</sup>, SU Dan-dan<sup>1,2</sup>,  
ZHANG Yu<sup>1,2</sup>, QIN Na<sup>1,2</sup>

(1. School of Life Sciences, Qinghai Normal University, Xining 810008, China; 2. Key Laboratory of Medicinal Animal and Plant Resources of the Qinghai-Tibet Plateau in Qinghai Province, Qinghai Normal University, Xining 810008, China; 3. Academy of Plateau Science and Sustainability, Qinghai Normal University, Xining 810016, China)

**Abstract:** *Cirsium souliei* (Asteraceae) is a perennial medicinal herb of *Cirsium* with important medicinal and ecological values. Here, we sequenced the complete chloroplast (cp) genome of *C. souliei* based on high-throughput sequencing technology, then assembled and annotated it, and analysed the structure and characteristics of the cp genome. The result indicated that the cp genome of *C. souliei* was a typical quadripartite circular structure of 152 470 bp in length, and GC content was 37.7%. The cp genome of *C. souliei* encoded 134 genes, including 89 protein-coding genes, 37 tRNA genes and 8 rRNA genes. Meanwhile, we detected 188 simple sequence repeats (SSR) loci in the cp genome, which were mainly composed of mononucleotide repeats. Codon bias analysis showed that leucine (Leu) was the highest amino acids with frequency (10.51%), and there were 30 codons with the value of relative synonymous codon usage (RSCU) above one, of which mostly ended with A/U. Additionally,

收稿日期: 2022-03-01; 修回日期: 2022-03-29.

基金项目: 国家自然科学基金资助项目 (41761009, 32160297, 31800310).

\*通讯作者 E-mail: lyp8527970@126.com

DOI: 10.16438/j.0513-4870.2022-0270

the result from phylogenetic analysis based on 46 cp genomes of Carduoideae showed that *C. souliei* and *C. vulgare* were sister species, and had the closest relationship with 100% bootstrap within *Cirsium*. This study provides theoretical basis for future studying genetic diversity, population genetic structure, systematics and evolution, and speciation mechanism.

**Key words:** chloroplast genome; phylogeny; simple sequence repeat; boundary analysis of IR region; *Cirsium souliei*

叶绿体是植物一种重要的半自主性细胞器, 拥有独立的基因组<sup>[1]</sup>。研究表明, 高等植物叶绿体基因组长度一般介于120~180 kb, 通常为典型的环状四分体结构, 包括1个大单拷贝区 (large single copy region, LSC)、1个小单拷贝区 (small single copy region, SSC) 和2个反向重复区 (inverted repeats region a/b, IRa/IRb)<sup>[2,3]</sup>。叶绿体基因组编码110~130个基因, 具有合成蛋白质、淀粉、色素等多种功能<sup>[4,5]</sup>。被子植物中, 叶绿体基因组 (cpDNA) 多为母系遗传, 仅少数为双亲或父系遗传, 因具有高度保守、结构稳定、进化速率慢、分子质量小等优点, 被广泛应用于物种界定、系统进化、遗传多样性、群体遗传结构及物种形成等领域, 取得理想的研究成果<sup>[6,7]</sup>。譬如, Jansen等<sup>[8]</sup>基于葡萄 (*Vitis vinifera*) 和其他26种被子植物叶绿体基因组序列构建系统发育树, 从全新的角度阐明了葡萄在被子植物中的系统发育关系; 梁凤萍等<sup>[9]</sup>基于27种菊科植物叶绿体基因组序列进行系统性研究, 为菊科植物的分类界定和系统进化提供了更准确的证据。

葵花大蓟 (*Cirsium souliei*) 是菊科 (Compositae)、管状花亚科 (Carduoideae)、蓟属 (*Cirsium*) 的一种多年生药用草本植物, 别名“聚头蓟”, 主要分布于青海、甘肃、四川和西藏等地, 生于海拔1 930~4 800 m的山坡、路旁、林缘、荒地、河滩地、田间、水旁潮湿处<sup>[10,11]</sup>。葵花大蓟全草入药, 具有凉血止血、散瘀消肿等功效, 主治吐血、鼻出血、尿血、子宫出血、崩漏、黄疸、疮痍等症状<sup>[10,12]</sup>。目前, 国内外学者对葵花大蓟的耐逆特性和化学成分进行了研究, 发现葵花大蓟具有较好的温度耐受性, 并从中分离鉴定出多种化合物<sup>[13-16]</sup>。然而, 有关葵花大蓟叶绿体基因组结构特征、系统位置及亲缘关系的研究至今未见报道。据此, 本研究采用高通量测序技术对葵花大蓟的叶绿体基因组进行测序和数据分析, 阐明叶绿体基因组的结构特征; 在此基础上, 通过对葵花大蓟及其45种近缘种系统发育树的构建, 明晰其系统位置及亲缘关系, 以期对葵花大蓟遗传多样性、群体遗传结构、系统进化及物种形成机制提供理论依据。

## 材料与方法

**实验材料** 新鲜无病虫害的葵花大蓟叶片采自青海省海北藏族自治州祁连县默勒镇 (N37.72898°, E100.6297°, alt. 3 471.12 m), 采集后迅速放于变色硅胶中干燥保存。凭证标本存放于中国科学院西北高原生物研究所青藏高原生物标本馆 (HNWP)。

**基因组DNA提取和测序** 采用改良的CTAB法提取基因组总DNA<sup>[17]</sup>, 然后利用1.0%琼脂糖凝胶电泳检测DNA的完整性, Nanodrop检测DNA的纯度 (OD 260/280比值) 以及Qubit 2.0荧光定量仪对样品DNA浓度定量; DNA样品检测合格后, 使用Covaris超声破碎仪随机破碎合格的DNA样品, 接着对测序片段进行末端修复、3'端加A尾、加测序接头和纯化; 通过PCR扩增构建DNA文库, 质检合格后利用Illumina NovaSeq6000高通量测序平台 (北京诺禾致源科技有限公司) 进行测序。

**叶绿体基因组组装和注释** 原始序列 (raw reads) 经Fastap软件<sup>[18]</sup>过滤、去除低质量序列, 获得待分析的高质量序列 (clean reads)。以蓟 (*Cirsium japonicum*; Genbank 登录号: NC\_053767.1) 叶绿体基因组序列为参考, 运用NOVOPlasty软件<sup>[19]</sup>对葵花大蓟叶绿体基因组序列进行组装, 设置为默认参数; 然后, 以葵花大蓟近缘种林蓟 (*Cirsium schantarense*; Genbank 登录号: MN871982) 和丝路蓟 (*Cirsium arvense*; Genbank 登录号: NC\_036965) 为参考物种, 利用在线工具GeSeq (<https://chlorobox.mpimp-golm.mpg.de/geseq.html>)<sup>[20]</sup>和Plastid Genome Annotator (PGA) 软件<sup>[21]</sup>分别对葵花大蓟的叶绿体基因组进行注释, 并将两种注释结果对比矫正, 去除错误及冗余注释; 接着, 采用在线工具Chloroplot (<https://irscope.shinyapps.io/Chloroplot/>)<sup>[22]</sup>绘制葵花大蓟叶绿体基因组图谱; 最后, 将注释好的葵花大蓟叶绿体基因组序列提交至NCBI数据库 (序列登录号ON720162)。

**密码子偏好性和SSR位点分析** 利用Codon W1.4.2软件<sup>[23]</sup>检测葵花大蓟叶绿体基因组中大于200 bp蛋白编码基因的密码子使用频率和相对密码子使用频率 (RSCU); 使用在线程序MISA (<https://webblast>).

ipk-gatersleben.de/misa/index.php?action=1)<sup>[24]</sup>对葵花大蓟叶绿体基因组序列进行SSR位点分析,其中单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸和六核苷酸的重复数分别设置为8、5、3、3、3和3。

**叶绿体基因组比较分析** 运用在线软件IRscope (<https://irscope.shinyapps.io/irapp/>)<sup>[25]</sup>对葵花大蓟及蓟属8种植物的叶绿体基因组进行IR区边界比较分析;以丝毛飞廉 (*Carduus crispus*; Genbank 登录号: NC\_053726.1) 为参考基因组,利用在线软件mVista (<http://genome.lbl.gov/vista/mvista/submit.html>)<sup>[26]</sup>中的Shuffle-LAGAN模型对上述9种植物的叶绿体基因组作比较分析。

**系统发育分析** 采用MAFFT v7软件<sup>[27]</sup>,对从GenBank数据库中下载的菊科管状花亚科45种植物及葵花大蓟的叶绿体基因组序列进行比对,以菊科舌状花亚科(Cichorioideae)中苣荬菜(*Sonchus arvensis*)为外类群,运用IQ-TREE<sup>[28]</sup>软件构建46种管状花亚科植物的最大似然(maximum likelihood, ML)树,bootstrap值设置为5 000,最佳模型设置为TVM+F+R2。

## 结果与分析

### 1 叶绿体基因组结构、分类、功能及特征

葵花大蓟叶绿体基因组全长152 470 bp,由一个长度为83 485 bp的LSC、一个长度为18 623 bp的SSC和一对长度为25 181 bp的反向重复区(inverted repeat regions, IRs)构成的环状四分体结构(图1)。葵花大蓟叶绿体基因组总GC含量为37.7%,其中SSC区、LSC区和IRs区的GC含量分别为35.8%、31.3%和43.1%。葵花大蓟叶绿体基因组共编码134个基因,包括蛋白编码基因89个、tRNA基因37个、rRNA基因8个,其中17个基因存在两个拷贝(表1)。此外,葵花大蓟叶绿体基因组中共发现13个基因含有内含子,其中7个基因含有1个内含子,6个基因含有2个内含子(表1)。

### 2 密码子偏好性分析

本研究从葵花大蓟叶绿体基因组中,获得71条大于200 bp的蛋白编码序列(coding DNA sequence, CDS),用于密码子偏好性分析。研究结果表明,葵花大蓟叶绿体基因组71条CDS序列共编码25 473个密码子。其

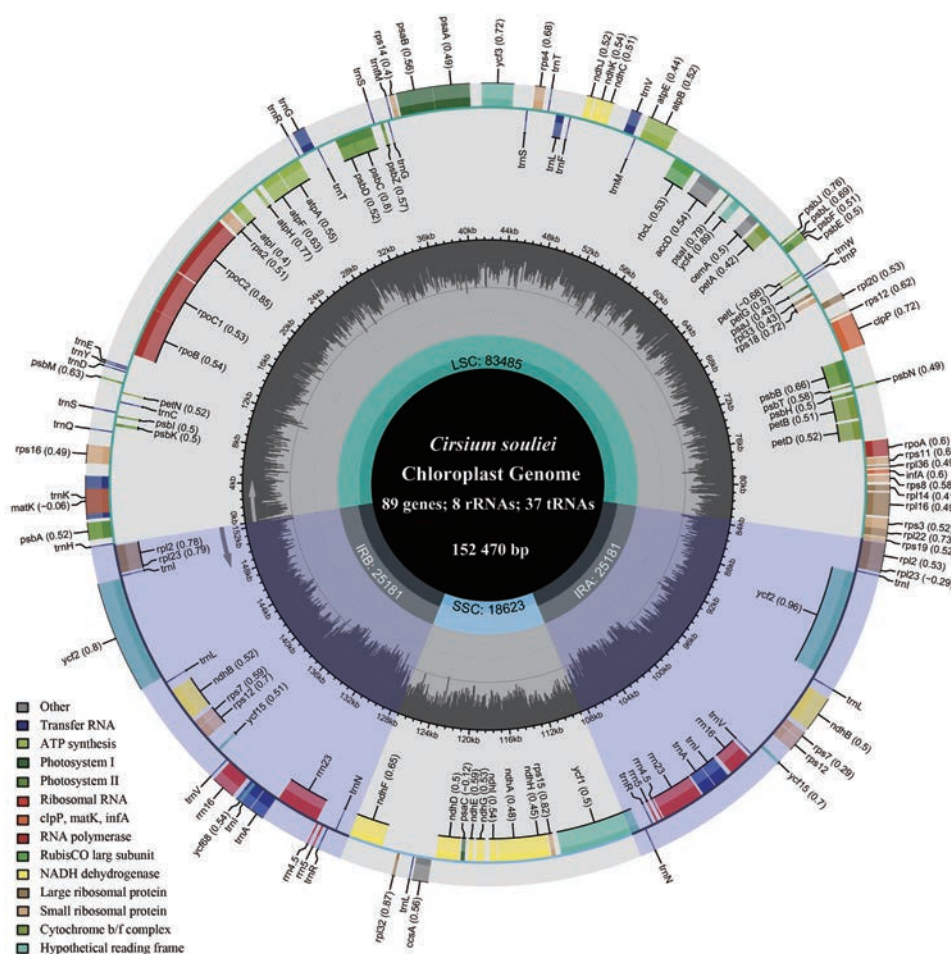


Figure 1 Circularized map of the chloroplast genome of *C. soubiei*

**Table 1** Gene annotation of the chloroplast genome of *C. souliei*. <sup>a, b</sup>Present an intron and two introns in protein-coding genes, respectively; <sup>c</sup>Presents duplicated genes

Gene category	Gene group	Gene name	
Genes for photosynthesis	Subunits of photosystem I	<i>psaA, psbA, psbC, psal, psaj</i>	
	Subunits of photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>	
	Subunit of cytochrome b/f complex	<i>petA, petB, petD, petG, petN, petL</i>	
	Subunits of ATP synthase	<i>atpA, atpB, atpE, atpF<sup>a</sup>, atpH, atpI</i>	
	Subunits of NADH dehydrogenase	<i>ndhA, ndhB<sup>bc</sup>, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>	
	Large subunit of Rubisco	<i>rbcL</i>	
	Self-replication	DNA dependent RNA polymerase	<i>rpoA, rpoB, rpoC1<sup>a</sup>, rpoC2</i>
		Small subunit of ribosome	<i>rps2, rps3, rps4, rps7<sup>c</sup>, rps8, rps11, rps12<sup>c</sup>, rps14, rps15, rps16, rps18, rps19</i>
		Large subunit of ribosome	<i>rpl2<sup>bc</sup>, rpl14, rpl16<sup>a</sup>, rpl20, rpl22, rpl23<sup>c</sup>, rpl32, rpl33, rpl36</i>
		Transfer RNA gene	<i>trnA-UGC<sup>bc</sup>, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnG-UCC<sup>a</sup>, trnH-GUG, trnI-CAU<sup>c</sup>, trnI-GAU<sup>bc</sup>, trnK-UUU<sup>a</sup>, trnL-CAA<sup>c</sup>, trnL-UAA<sup>a</sup>, trnL-UAG, trnM-CAU, trnM-CAU, trnN-GUU<sup>c</sup>, trnP-UGG, trnQ-UUG, trnR-ACG<sup>c</sup>, trnR-UCU, trnS-GGA, trnS-GCU, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC<sup>c</sup>, trnV-UAC<sup>c</sup>, trnW-CCA, trnY-GUA</i>
Ribosomal RNA gene		<i>rrn4.5<sup>c</sup>, rrn5<sup>c</sup>, rrn16<sup>c</sup>, rrn23<sup>c</sup></i>	
Translation initiation factor		<i>infA</i>	
Other genes		Maturase	<i>matK</i>
		Envelop membrane protein	<i>cemA</i>
		c-Type eytochrome synthesis gene	<i>ccsA</i>
		Submit of acetyl-CoA-carboxylase	<i>accD</i>
	ATP-dependent protease subunit P	<i>clpP<sup>b</sup></i>	
	Genes of unknown function	Conserved open reading frame	<i>ycf1, ycf2<sup>c</sup>, ycf3<sup>b</sup>, ycf4, ycf15<sup>c</sup>, ycf68</i>

**Table 2** Relative synonymous codon usage of each amino acid in *C. souliei*. <sup>\*</sup>Represents termination codon

Amino acid	Codon	Number	RSCU	Ratio	Amino acid	Codon	Number	RSCU	Ratio		
Phe	UUU	935	1.29	5.68%	Ala	GCU	606	1.77	5.37%		
	UUC	511	0.71			GCC	213	0.62			
Leu	UUA	824	1.85	10.51%	TER <sup>*</sup>	GCA	399	1.17	0.28%		
	UUG	569	1.28			UAA	38	1.58			
	CUU	571	1.28			UAG	18	0.75			
	CUC	173	0.39			UGA	16	0.67			
	CUA	363	0.81			His	CAU	465		1.54	2.37%
	CUG	177	0.40				CAC	138		0.46	
Ile	AUU	1 049	1.46	8.48%	Gln	CAA	681	1.51	3.54%		
	AUC	434	0.60			CAG	220	0.49			
	AUA	677	0.94			Asn	AAU	972		1.56	4.88%
Met	AUG	596	1.00	2.34%	Lys		AAA	1 016	1.48	5.39%	
	Val	GUU	486	1.44			5.30%	AAG	358		
		GUC	170	0.50		Asp	GAU	832	1.59		4.11%
GUA		513	1.52	GAC	214		0.41				
GUG	182	0.54	Glu	GAA	977		1.46	5.24%			
Ser	UCU	573		1.76	7.68%	GAG	359		0.54		
	UCC	301		0.92	Cys	UGU	206		1.47	1.10%	
	UCA	396	1.21	UGC		74	0.53				
	UCG	170	0.52	Trp		UGG	450	1.00	1.77%		
	AGU	393	1.21		Arg	CGU	335	1.29		6.12%	
	AGC	123	0.38			CGC	102	0.39			
Pro	CCU	401	1.51	4.17%		CGA	339	1.31			
	CCC	189	0.71	Thr	CGG	123	0.47				
	CCA	312	1.18		AGA	489	1.88				
	CCG	159	0.60		AGG	170	0.65				
Tyr	ACU	510	1.62		4.95%	Gly	GGU	565	1.30	6.85%	
	ACC	235	0.74	GGC	199		0.46				
	ACA	390	1.24	GGA	679		1.56				
	ACG	127	0.40	GGG	302		0.69				

中, 编码亮氨酸 (Leu) 的密码子占比最多, 为 10.51%; 编码半胱氨酸 (Cys) 的密码子占比最少, 为 1.10% (表 2)。同时, 还检测到 30 个相对同义密码子使用度 (RSCU) 大于 1 的密码子, 除 UUG 外其余密码子都以 A/U 结尾, 且编码精氨酸 (Arg) 的密码子 AGA 的 RSCU 值最高, 为 1.88 (表 2 和图 2)。

### 3 SSR 位点分析

葵花大蓟叶绿体基因组共检测到 188 个 SSR 位点, 其中单核苷酸重复有 108 个, 包括 A、C 和 T 3 种重

复类型; 二核苷酸重复有 4 个, 仅有 AT 和 TA 两种类型; 三核苷酸重复有 67 个, 包括 AAC、AAG、AAT、ACC、ACG、AGA、AGT、ATA 等 29 种类型; 四核苷酸重复有 9 个, 包括 AAAT、AATC、ATAA、GATT、GTTT、TATT 和 TTTC 7 种类型 (图 3A)。此外, 葵花大蓟叶绿体基因组中大部分 SSR 位点位于蛋白编码序列 (CDS), 占比约为 47%, 其次在基因间隔区 (intergenic spacer, IGS) 的 SSR 位点约为 45%, 而位于内含子 (intron) 中的 SSR 位点占比最小, 仅为 8% (图 3B)。

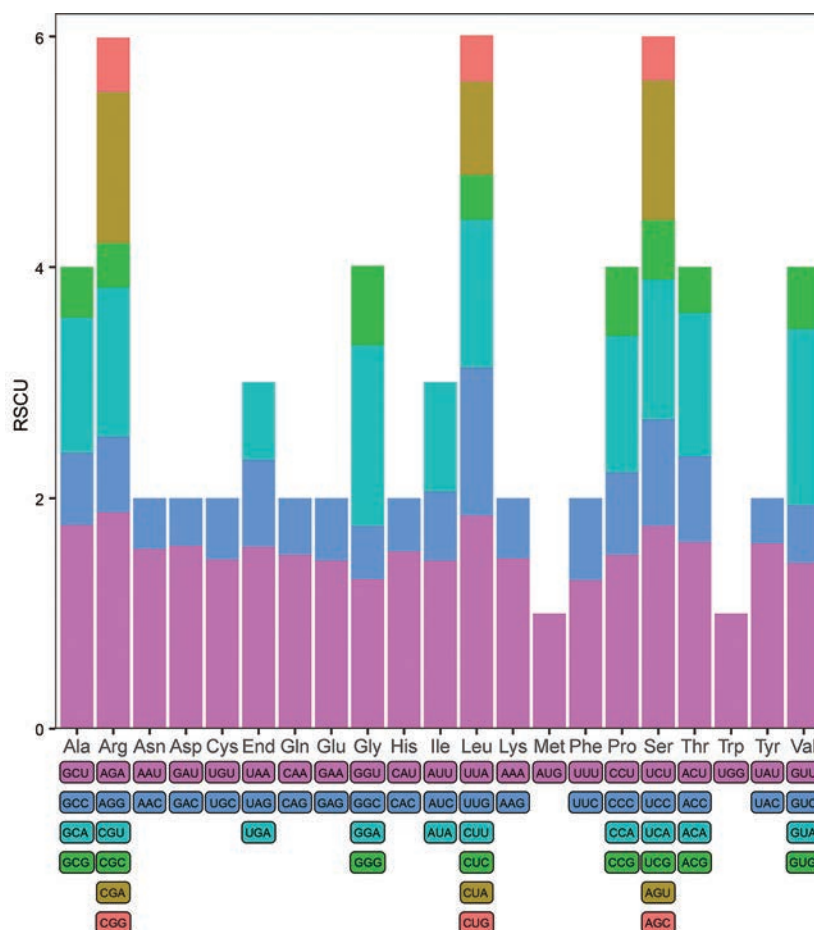


Figure 2 RSCU analysis of each amino acid in *C. souliei*

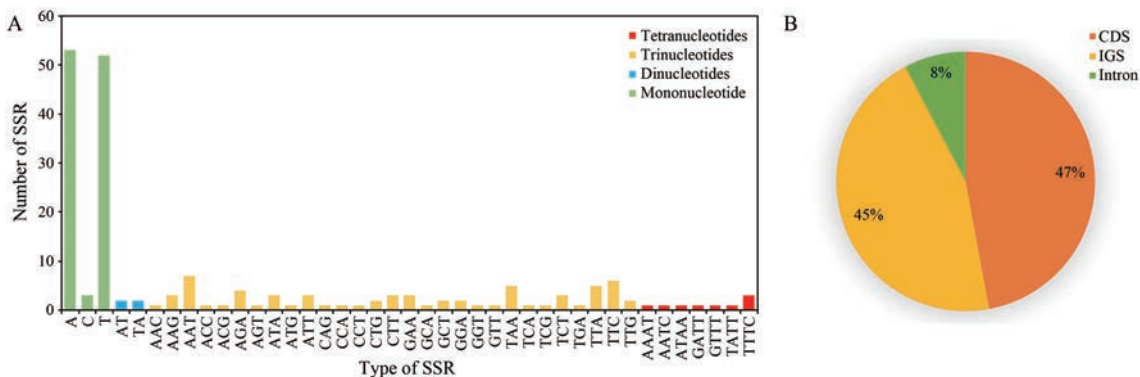


Figure 3 SSR types and distributions in the chloroplast genome of *C. souliei*. A: Types and numbers of SSR in the chloroplast genome of *C. souliei*; B: Distribution of SSRs in the coding DNA sequence (CDS), intergenic spacer (IGS) and intron regions

#### 4 叶绿体基因组比较分析

叶绿体基因组IR区扩张和收缩分析表明, 蓟属所有参试物种在LSC/IR和SSC/IR边界区具有相似的基因种类、排列和长度, 但在这些边界区也存在明显的扩张和收缩(图4)。就LSC/IRb边界区而言, 葵花大蓟、丝路蓟、绵毛蓟(*C. eriophorum*)、刺儿菜(*C. setosum*)、蓟和*C. nipponicum*中*rps19*基因均向IRb区扩张了60 bp; IRb/SSC边界区, 除葵花大蓟和蓟的*ycf1*基因缺失外, 其余蓟属物种的*ycf1*基因均向IRb区扩张了6~10 bp; 同样, SSC/IRa边界区, 蓟属所有参试物种的*ycf1*基因均向IRa区扩张了566~571 bp(图4)。此外, 叶绿体基因组全序列比对分析显示, 蓟属所有参试物种的叶绿体基因组高度保守, 未检测到大量的基因缺失和变异(图5), 但也有微量的基因变异存在, 如在*petN-psbM*、*trnR-UCU-trnM-CAU*、*petD-rpoA*、*ndhB-ycf15*、*trnL-UAG-ndhF*、*ycf1*等区域; 并且, 参试物种叶绿体基因组中LSC区和SSC区的遗传变异显著高于IR区(图5)。

#### 5 系统发育分析

以苣荬菜为外类群, 基于管状花亚科46种植物叶绿体基因组构建的最大似然(ML)树结果显示, 管状花亚科46种植物聚为Clade 1和Clade 2两大分支, 其中

Clade 1由飞廉亚族(*Carduinae*)和矢车菊亚族(*Centau-reinae*)的物种构成, Clade 2由刺苞亚族(*Carlininae*)和旋覆花亚族(*Inulinae*)的物种组成(图6)。在Clade 1中, 葵花大蓟与翼蓟为姊妹类群, 构成一个单系分支, 支持率(bootstrap)高达100%, 表明两者亲缘关系较近; 并且, 其与飞廉属(*Carduus*)丝毛飞廉和*C. tenuiflorus*构成的单系分支互为姊妹类群, 支持率为100%(图6)。此外, 牛口刺(*C. shansiense*)和丝路蓟构成的单系分支与刺儿菜互为姊妹类群, 支持率为100%; 蓟和*C. rhinoceros*为姊妹种, 构成一个单系分支, 支持率也为100%; 然而, 蓟属的绵毛蓟与*C. nipponicum*构成的单系分支与*Saussurea tanguensis*也为姊妹类群, 但這些物种与蓟属其他物种亲缘关系较远(图6)。由此看来蓟属参试物种并非是一个单系类群。

#### 讨论

研究表明, 陆生植物叶绿体基因组大小常介于120~180 kb之间, 基因结构、组成和排列顺序较为保守<sup>[3]</sup>。葵花大蓟叶绿体基因组长度为152 470 bp、GC含量为37.7%、共编码134个基因, 这与先前报道的蓟属其他植物的叶绿体基因组结构特征基本类似<sup>[29-31]</sup>,

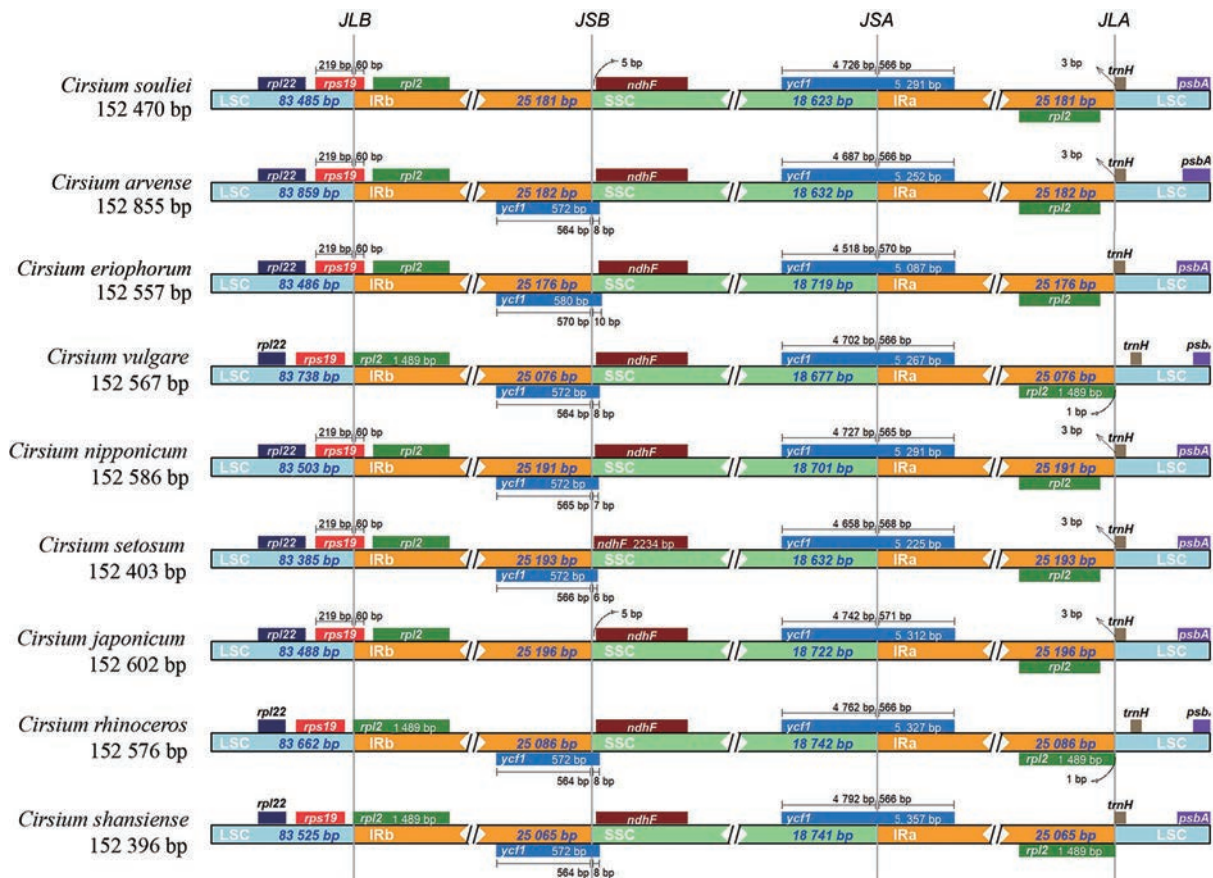
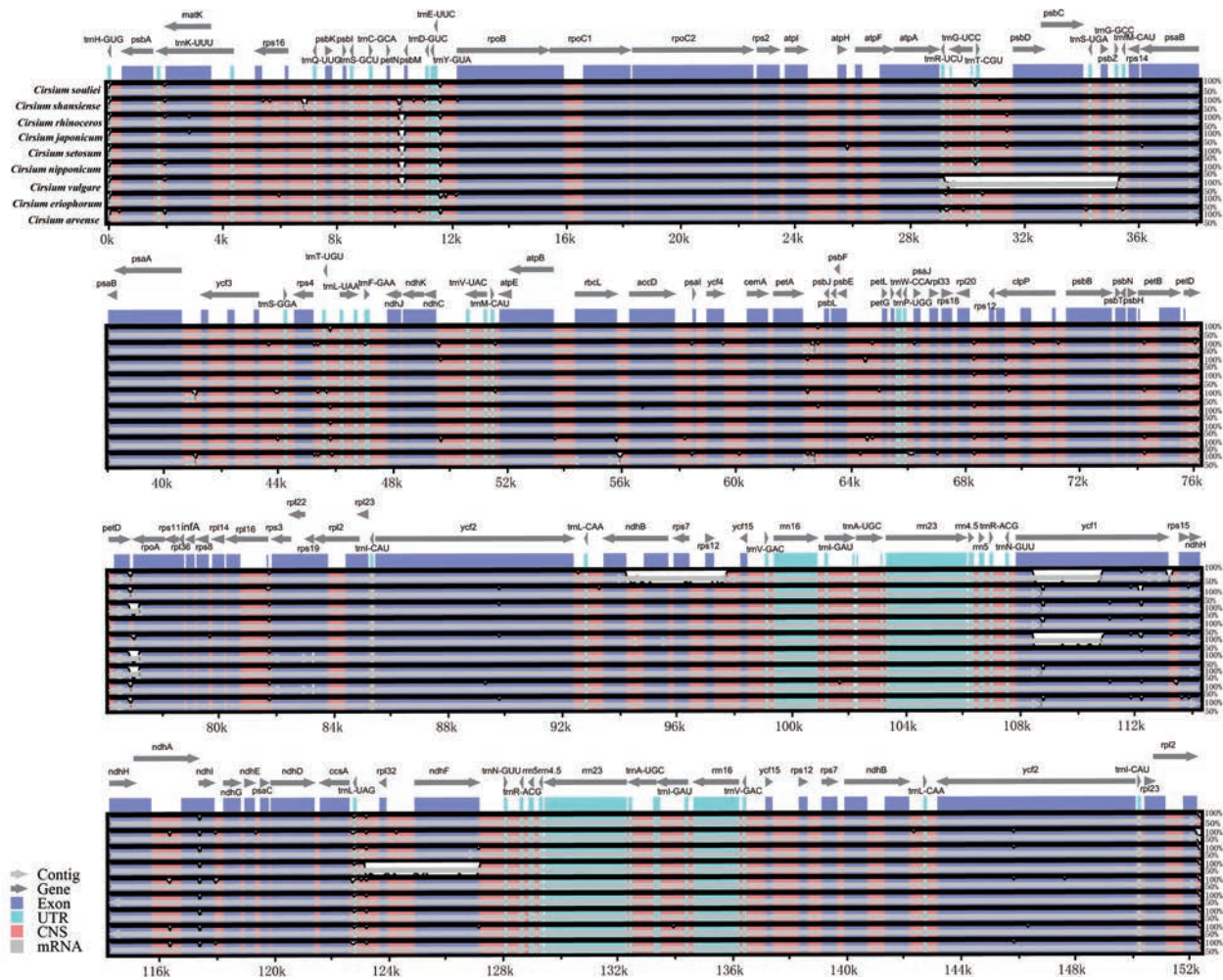


Figure 4 Boundary analysis of IR regions in chloroplast genomes from nine *Cirsium* species



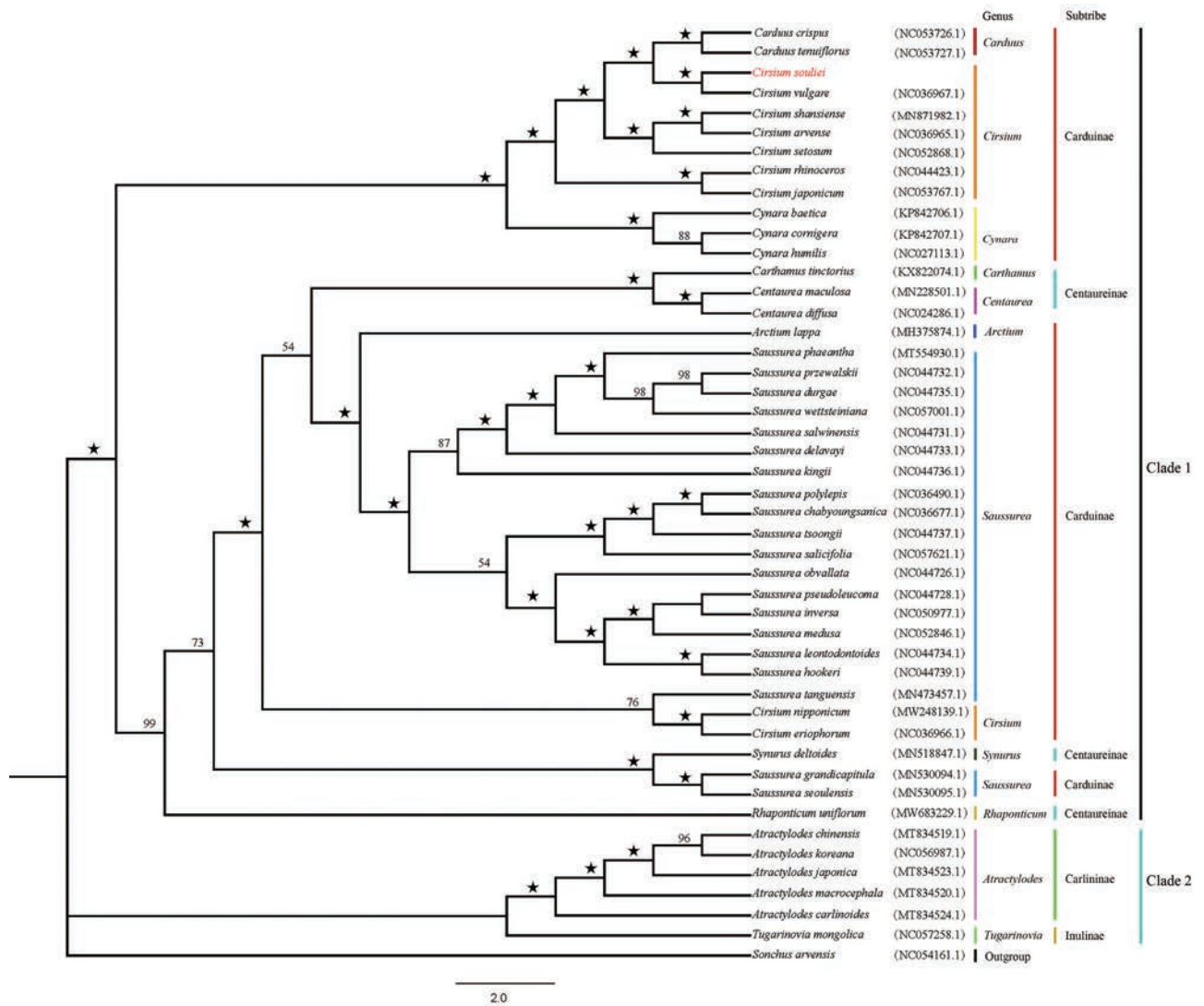
**Figure 5** Full sequence alignment of chloroplast genomes of nine *Cirsium* species. The top arrows show transcription direction; purple color indicates the exons of protein-coding genes (exon); cyan color indicates tRNAs and rRNAs; pink-orange color shows conserved non-coding sequence (CNS); The X-axis represents the positions in the cp genome while Y-axis represents percentage identity within 50%–100%

因此葵花大蓟的叶绿体基因组结构相对保守, 这进一步验证了先前研究结论的可靠性和合理性<sup>[3]</sup>。

IR、LSC和SSC区扩张和收缩是植物进化中的常见现象, 也是造成叶绿体基因组长度差异与结构变异的主要原因<sup>[32,33]</sup>。譬如, 苏玥等<sup>[34]</sup>通过对乳苣 (*Mulgedium tataricum*) 及其他5个菊科莴苣族 (*Lactuceae*) 植物叶绿体基因组LSC/IR和SSC/IR边界区扩张和收缩的研究, 发现莴苣族6个物种的叶绿体基因组序列整体高度相似, 但 *rps19*、*rpl2*、*ycf1*、*trnN*、*ndhF*、*trnH* 和 *rpl23* 等基因在LSC/IR和SSC/IR边界区的位置和大小存在稍许差异。本研究结果也同样表明, 葵花大蓟及蓟属其他8种植物叶绿体基因组的LSC/IR和SSC/IR边界基因排列和类型高度相似, 仅有少量差异, 如葵花大蓟和蓟的 *ycf1* 基因仅存在IRb/SSC边界区的丢失和SSC/IRa边界区向IRa区的扩张; 同时, 葵花大蓟、丝路蓟、绵毛蓟、刺儿菜、蓟和 *C. nipponicum* 的 *rps19* 基因也存

在向IRb区扩张的现象。此外, 叶绿体基因组序列比对结果显示, 蓟属所有参试物种叶绿体基因组中LSC和SSC区的遗传变异显著高于IR区, 尤其存在高变异区域, 如 *petN-psbM*、*trnR-UCU-trnfM-CAU*、*petD-rpoA*、*ndhB-ycf15*、*trnL-UAG-ndhF* 和 *ycf1*, 其往往会发生快速的核苷酸替换, 因而有望从这些区域中开发用于蓟属种间鉴定和系统发育分析的分子标记<sup>[35]</sup>。

叶绿体简单序列重复 (SSR) 拷贝数变异是一种重要的分子标记, 因其具有比核基因与线粒体SSR更大的分类距离, 被广泛应用于植物群体遗传多样性和系统发育研究<sup>[36-38]</sup>。通过对葵花大蓟叶绿体基因组的分析, 本研究共检测到188个SSR位点, 其主要位于CDS区; 并且, 单核苷酸A/T碱基在SSR位点中出现频率最高, 这与先前报道的菊科其他植物的叶绿体基因组类似, 如乳苣、茵陈蒿 (*Artemisia capillaris*) 和银叶蒿 (*Artemisia argyrophylla*) 等<sup>[34,39]</sup>。然而, 葵花大蓟SSR



**Figure 6** Phylogenetic tree of 46 species in Carduoideae based on the chloroplast genomes using maximum likelihood (ML). "★" Represent fully supported nodes

位点的三核苷酸和四核苷酸重复类型与已报道的菊科其他植物却有较大差异,如葵花大蓟检测到的三核苷酸和四核苷酸类型分别有29种和7种,远高于菊科其他物种<sup>[39-42]</sup>,说明葵花大蓟与菊科其他植物叶绿体基因组SSR重复类型差异较大,必将为葵花大蓟SSR引物开发和系统发育研究提供较好的基础。

研究表明,自然界存在的20种氨基酸中,除甲硫氨酸(Met)和色氨酸(Trp)由唯一密码子编码外,其他氨基酸均对应2~6个同义密码子,由于同义密码子在机体内的使用频率存在差异,从而形成植物密码子的使用偏性,即密码子偏好性<sup>[43]</sup>。密码子偏好性是生物体进化的重要特征,其对基因功能、物种进化等研究具有重要意义<sup>[44]</sup>。本研究密码子偏好性分析表明,亮氨酸(Leu)是葵花大蓟叶绿体基因组中占比最高的氨基酸,并且相对同义密码子使用度(RSCU)大于1的30个密

码子通常以A/U结尾,这与先前发表的菊科植物基本相似<sup>[41,45]</sup>,认为这可能是由于叶绿体基因组富含大量A、U碱基,从而导致不同植物的密码子偏好性不同,但植物间亲缘关系越近则往往具有相似的密码子使用模式,表明植物密码子偏好性在进化关系上较为保守<sup>[46,47]</sup>。

探究菊科植物的系统发育关系已成为当前系统与进化植物学研究的热点领域<sup>[48,49]</sup>。纵观国内外研究进展,发现菊科植物系统发育的研究主要集中于科内属间与一些属内<sup>[9,34,50,51]</sup>,而关于蓟属种内系统发育分析的报道较少<sup>[29-31]</sup>,尤其有关葵花大蓟系统发育关系的研究目前尚属空白。本研究基于菊科管状花亚科46种植物的叶绿体基因组数据,通过对其系统发育的分析发现葵花大蓟与翼蓟为姊妹类群,靴带支持率高达100%,表明两者具有最近的亲缘关系。

**作者贡献:** 郑长远负责实验数据分析和论文初稿撰写; 苏旭和刘玉萍负责实验设计和论文修改; 刘涛、王亚男、苏丹丹、张雨和秦娜参与野外采样和数据分析。全体作者都阅读并同意最终文本。

**利益冲突:** 无相关利益冲突。

## References

- [1] Corriveau JL, Coleman AW. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species [J]. *Am J Bot*, 1988, 75: 1443-1458.
- [2] Raman G, Park KT, Kim JH, et al. Characteristics of the completed chloroplast genome sequence of *Xanthium spinosum*: comparative analyses, identification of mutational hotspots and phylogenetic implications [J]. *BMC genomics*, 2020, 22: 855.
- [3] Xu WB, Xia BS, Li XW. The complete chloroplast genome sequences of five pinnate-leaved *Primula* species and phylogenetic analyses [J]. *Sci Rep*, 2020, 10: 20782.
- [4] Palmer JD. Comparative organization of chloroplast genomes [J]. *Annu Rev Genet*, 1985, 19: 325-354.
- [5] Rodríguez-Ezpeleta N, Brinkmann H, Burey SC, et al. Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes [J]. *Curr Biol*, 2005, 15: 1325-1330.
- [6] Zhang YJ, Li DZ. Advances in phylogenomics based on complete chloroplast genomes [J]. *Plant Diver and Resour (植物分类与资源学报)*, 2011, 33: 365-375.
- [7] Du XY, Zeng T, Feng Q, et al. The complete chloroplast genome sequence of yellow mustard (*Sinapis alba* L.) and its phylogenetic relationship to other Brassicaceae species [J]. *Gene*, 2020, 731: 144340.
- [8] Jansen RK, Kaitanis C, Sasaki C, et al. Phylogenetic analyses of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosids [J]. *BMC Evol Biol*, 2006, 6: 32.
- [9] Liang FP, Wen XN, Gao HY, et al. Analysis of chloroplast genomes features of Asteraceae species [J]. *Genom Appl Biol (基因组学与应用生物学)*, 2018, 37: 5437-5447.
- [10] Chen YL, Shi Z. *Flora Reipublicae Popularis Sinicae: Vol 78 (中国植物志第 78 卷)* [M]. Beijing: Science Press, 1987: 86.
- [11] Liu SW. *Flora Qinghaiica (青海植物志)* [M]. Xining: Qinghai People's Publishing House, 1996: 439.
- [12] Northwest Institute of Plateau Biolog, CAS. *Economic Flora of Qinghai (青海经济植物志)* [M]. Xining: Qinghai People's Publishing House, 1987: 611.
- [13] Yang D, Luo Q, Cao ML, et al. Temperature stress on physiological indices of Compositae weeds and their adaptive significance [J]. *J Jishou Univ (吉首大学学报)*, 2006, 27: 75-79.
- [14] He ZW, Lv ZP, Wu WS, et al. Lipophilic chemical constituents from *Cirsium souliei* [J]. *Acta Bot Bor-Occid Sin (西北植物学报)*, 2007, 27: 1884-1887.
- [15] Lv MY, Ge H, Liu SJ, et al. Studies on the chemical constituents of *Cirsium souliei* (Franch.) Mattf. [J]. *Northwest Pharm J (西北药学报)*, 2013, 28: 118-121.
- [16] He XF, He ZW, Jin XJ, et al. Caryolane-type sesquiterpenes from *Cirsium souliei* [J]. *Phytochem Lett*, 2014, 10: 80-85.
- [17] Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue [J]. *Phytochem Bull*, 1987, 19: 11-15.
- [18] Chen SF, Zhou YQ, Chen YR, et al. Fastp: an ultra-fast all-in-one FASTQ preprocessor [J]. *Bioinformatics*, 2018, 34: i884-i890.
- [19] Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data [J]. *Nucleic Acids Res*, 2017, 45: e18.
- [20] Tillich M, Lehwark P, Pellizzer T, et al. GeSeq - versatile and accurate annotation of organelle genomes [J]. *Nucleic Acids Res*, 2017, 45: W6-W11.
- [21] Qu XJ, Moore MJ, Li DZ, et al. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes [J]. *Plant Methods*, 2019, 15: 50.
- [22] Zheng SY, Poczai P, Hyvönen J, et al. Chlorplot: an online program for the versatile plotting of organelle genomes [J]. *Front Genet*, 2020, 11: 576124.
- [23] Sharp PM, Li WH. The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications [J]. *Nucleic Acids Res*, 1987, 15: 1281-1295.
- [24] Beier S, Thiel T, Münc T, et al. Misa-web: a web server for microsatellite prediction [J]. *Bioinformatics*, 2017, 33: 2583-2585.
- [25] Amiryousefi A, Hyvönen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes [J]. *Bioinformatics*, 2018, 34: 3030-3031.
- [26] Frazer KA, Pachter L, Poliakov A, et al. VISTA: computational tools for comparative genomics [J]. *Nucleic Acids Res*, 2004, 32: W273-W279.
- [27] Katoh K, Misawa K, Kuma K, et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform [J]. *Nucleic Acids Res*, 2002, 30: 3059-3066.
- [28] Nguyen LT, Schmidt HA, von Haeseler A, et al. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies [J]. *Mol Biol Evol*, 2015, 32: 268-274.
- [29] Nam SJ, Kim JM, Kim YS, et al. The complete chloroplast genome of Korean endemic species, *Cirsium rhinoceros* (H. Lév. & Vaniot) Nakai (Asteraceae) [J]. *Mitochondrial DNA B*, 2019, 4: 2351-2352.
- [30] Xie BX, Jiang CY, Zhu ZY, et al. The complete chloroplast genome of *Cirsium setosum* and its phylogenetic analysis [J]. *Mitochondrial DNA B*, 2019, 4: 3756-3757.
- [31] Xu JS, Dang H, Li HZ. The complete chloroplast genome of *Cirsium shansiense* and its phylogenetic analysis [J]. *Mitochondrial DNA B*, 2020, 5: 1134-1135.

- [32] Kim KJ, Lee HL. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants [J]. DNA Res, 2004, 11: 247-261.
- [33] Xue S, Shi T, Luo WJ, et al. Comparative analysis of the complete chloroplast genome among *Prunus mume*, *P. armeniaca*, and *P. salicina* [J]. Hort Res, 2019, 6: 89.
- [34] Su Y, Liu JJ, Wan B, et al. Chloroplast genome structure characteristic and phylogenetic analysis of *Mulgedium tataricum* [J]. J Agric Sci Technol (中国农业科技导报), 2021, 23: 33-42.
- [35] Ma ML, Meng HL, Zhang W, et al. Complete chloroplast genome and phylogenetic analysis of *Amomum villosum* [J]. Acta Bot Bor-Occid Sin (西北植物学报), 2020, 40: 978-986.
- [36] Powell W, Morgante M, Mcdevitt R, et al. Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines [J]. Proc Natl Acad Sci U S A, 1995, 92: 7759-7763.
- [37] Pugh T, Fouet O, Risterucci AM, et al. A new cacao linkage map based on codominant markers: development and integration of 201 new microsatellite markers [J]. Theor Appl Genet, 2004, 108: 1151-1161.
- [38] Wang HK, Lou XM, Zhang Z. Application in germplasm resource research using chloroplast simple sequence repeat [J]. Mol Plant Breed (分子植物育种), 2006, 4: 92-98.
- [39] Kim GB, Lim CE, Kim JS, et al. Comparative chloroplast genome analysis of *Artemisia* (Asteraceae) in East Asia: insights into evolutionary divergence and phylogenomic implications [J]. BMC Genomics, 2020, 21: 415.
- [40] Iram S, Abdullah, Furrkh M, et al. Chloroplast genome sequences of *Artemisia maritima* and *Artemisia absinthium*: comparative analyses, mutational hotspots in genus *Artemisia* and phylogeny in family Asteraceae [J]. Genomics, 2020, 112: 1454-1463.
- [41] Tyagi S, Jung JA, Kim JS, et al. Comparative analysis of the complete chloroplast genome of mainland *Aster spathulifolius* and other *Aster* species [J]. Plants, 2020, 9: 568.
- [42] Loeuille B, Thode V, Siniscalchi C, et al. Extremely low nucleotide diversity among thirty-six new chloroplast genome sequences from *Aldama* (Heliantheae, Asteraceae) and comparative chloroplast genomics analyses with closely related genera [J]. PeerJ, 2021, 9: e10886.
- [43] Campos JL, Zeng K, Parker DJ, et al. Codon usage bias and effective population sizes on the X chromosome versus the autosomes in *Drosophila melanogaster* [J]. Mol Biol Evol, 2013, 30: 811-823.
- [44] Wang J, Wang TY, Wang LY, et al. Assembling and analysis of whole chloroplast genome sequence of *Elaeagnus angustifolia* and its codon usage bias [J]. Acta Bot Bor-Occid Sin (西北植物学报), 2019, 39: 1559-1572.
- [45] Gichira AW, Avoga S, Li ZZ, et al. Comparative genomics of 11 complete chloroplast genomes of Senecioneae (Asteraceae) species: DNA barcodes and phylogenetics [J]. Bot Stud, 2020, 60: 17.
- [46] Zhang MW, Zhang TK, Liu CY, et al. Biosynthesis pathway evolution and codon usage bias analysis of lignin in *Punica granatum* [J]. Mol Plant Breed (分子植物育种), 2018, 16: 2131-2138.
- [47] Yuan XL, Li YQ, Zhang JF, et al. Analysis of codon usage bias in the chloroplast genome of *Dalbergia odorifera* [J]. Guihaia (广西植物), 2021, 41: 622-630.
- [48] Zhu SX. Systematics of *Chaetoseris* and *Stenoseris* (Compositae-Lactuceae) (毛鳞菊和细莴苣属的系统学研究) [D]. Beijing: Graduate School, the Chinese Academy of Science (Institute of Botany, the Chinese Academy of Sciences), 2004.
- [49] Gu JY. A Phylogenetic Study of *Anaphalis* Based on Chloroplast Genomes (基于叶绿体基因组的香青属分子系统学研究) [D]. Zhengzhou: Zhengzhou University, 2019.
- [50] Wu X, Jiang M, Chen HM, et al. Comparative analysis of three complete chloroplast genomes of *Inula* genus with phylogenetic analysis of 49 plants from Carduoideae [J]. Acta Pharm Sin (药学报), 2020, 55: 1042-1049.
- [51] Zhu B, Hu LJ, Tian ZJ, et al. Phylogenetic relationship analysis of 50 species in Asteraceae family based on the chloroplast genome sequence [J]. J Guizhou Norm Univ (贵州师范大学学报), 2021, 39: 15-25.