Pak. J. Bot., 57(2): DOI: http://dx.doi.org/10.30848/PJB2025-2(33)

THE COMPLETE CHLOROPLAST GENOME AND PHYLOGENY OF *EUPHORBIA ALTOTIBETICA* (EUPHORBIACEAE), AN ENDEMIC MEDICINAL PLANT SPECIES FROM NORTHWEST CHINA

RONGJU QU[1,2,3], XUANRUI MAO[1,2,3], YONGHUI ZHOU[4], YUPING LIU[1,2,3*], XU SU[1,2,3*], XIAYU HU[1], PING YANG[1], CHANGYUAN ZHENG[1] AND AFTAB AHMAD[5*]

[1]*School of Life Sciences, Qinghai Normal University, Xining, China*
[2]*Key Laboratory of Biodiversity Formation Mechanism and Comprehensive Utilization of the Qinghai-Tibet Plateau in Qinghai Province, Qinghai Normal University, Xining, China*
[3]*Academy of Plateau Science and Sustainability, Qinghai Normal University, Xining, China*
[4]*Ganzhou Research Institute of Vegetables and Flowers, Ganzhou, China*
[5]*School of Ecology and Environment, Northwestern Polytechnical University, Xian, China*
[*]*Corresponding author's email: lyp8527970@126.com, xusu8527970@126.com, aftabahmadkhan003@gmail.com*

Abstract

Euphorbia altotibetica is an endemic medicinal herb species to Northwest China, belonging to genus *Euphorbia* of Euphorbiaceae. We sequenced and assembled the complete chloroplast genome of *Euphorbia altotibetica* followed by annotation using high-throughput sequencing technology, and annotation pipelines. Our result shows that the complete chloroplast genome of *E. altotibetica* is 161,141 bp in length with a GC content of 35.5%, with a typical tetrad ring structure. The chloroplast assembly consists of a large single copy region, a small single copy region, and two inverted repeat regions, with size 89,398 bp, 17,802bp, and 26,133 bp respectively. A total of 130 genes were annotated, including eight rRNA genes, 38 tRNA genes, and 84 protein genes. We also detected 90 simple sequence repeats (SSRs) loci in the chloroplast genome, which are mainly composed of mononucleotide repeats. Codon bias analysis showed that isoleucine (Ile) is the most abundant amino acid with frequency (8.78%), and four codons (*ycf1*, *ycf2*, *rpl20*, and *rpl22*) have the relative synonymous codon usage (RSCU) value above one. In the IRa regions of *E. altotibetica*, there is deletion in gene *rps19*. Phylogenetic analysis based on the chloroplast genome of *E. altotibetica* along with 68 other species of genus *Euphorbia* showed that the *E. altotibetica* and *E. peplus* were closest relatives, and nested together with a support rate of 100%. In this study, we obtained and characterized the complete chloroplast genome of *E. altotibetica* followed by its phylogenetic analysis for the first time. Our study provides theoretical basis for future research to understand the genetic diversity, phylogeny, population genetic structure, and speciation mechanism of genus Euphorbiaceae.

Key words: *Euphorbia altotibetica*, Chloroplast genome, IR region, Phylogenetic analysis, Codon usage.

Introduction

E. altotibetica is a perennial herb belonging to the genus *Euphorbia* of Euphorbiaceae, which is mainly distributed in the hillside, tussocks, and lakeside regions of Qinghai, Gansu, Ningxia, and Tibet of Northwest China, with the altitude of 2,800-3,900 meters (Li *et al*., 2008). *E. altotibetica* is a unique Tibetan medicinal plant in China and is mainly used in the treatment of skin tinea and swelling (Pan *et al*., 2003). Pharmacological studies showed the main chemical composition of this herb consist of a variety of Diterpenoids, Triterpenoids, Sterol, Flavonoids, Coumarin, Anthraquinone, Lignans with anti-cancer, anti-bacterial and anti-inflammatory activity (Shi *et al*., 2008). Previous studies on *E. altotibetica* mainly focused on its chemical composition (Pan *et al*., 2003; Zhang *et al*., 2013), and there are no reports available on the characterization of chloroplast genomes and phylogenetic relationship of *E. altotibetica.*

Chloroplast is an important semi-autonomous organelle in plants that perform photosynthesis and energy conversion (Corriveau & Coleman, 1988) and plays a very important role in plant stress resistance (Gray, 1989). The typical angiosperm chloroplast genome size is usually about 120-180 kb and encoding 110-130 genes (Palmer, 1985; Xu *et al*., 2020). Chloroplast genomes are mostly inherited maternally, and only a few are biparental or paternally inherited. The chloroplast genome exhibits a highly conserved and quadripartite structure with slow evolution rate, including a large single copy region (LSC), a small single copy area (SSC), and two inverted repeat (IR) sequences of the same size (Jansen *et al*., 2005). Previous studies indicated that the chloroplast genome has greater function on plant taxonomy (Parks *et al*., 2009) and adaptive evolution (Lemieux *et al*., 2016), especially for inter-specific identification (Yu *et al*., 2017; Du *et al*., 2020) and related species phylogeny (Straub *et al*., 2012; Zhou *et al*., 2017; Liu *et al*., 2018).

In this study we used Illumina high-throughput sequencing platform to sequence the whole chloroplast genome of *E. altotibetica*, followed by genome annotation and structural characterization, providing new genomic resources for this species. We also carried out phylogenetic analysis to evaluate the sequence divergence in chloroplast regions of *E. altotibetica*, when comparing it to other known species of the genus *Euphorbia*. Our results provide important information regarding the complete chloroplast genome, and the phylogenetic status of *E. altotibetica* for future studies.

Material and Methods

Sample collection: We collected fresh leaves of *E. altotibetica* from Yangkang Township, Tianjun County, Haixi Mongolian and Tibetan Autonomous Prefecture, Qinghai Province (37°42'8.784"N, 98°33'58.464"E, altitude 3,673.45 m). The leaves were dried immediately

after picking, using denatured silica gel. The voucher specimens were stored in the Herbarium of Northwest Institute of Plateau Biology (HNWP), Chinese Academy of Sciences, Qinghai Provence, China.

DNA extraction and sequencing: Leaf genomic DNA was extracted by the modified CTAB method (Tai & Tanksley, 1990). DNA quality was evaluated using horizontal electrophoresis with 1% agarose gel. The qualified DNA was randomly sheared by the Covaris Ultrasonic Breaker, and the whole library preparation was completed with repaired end, added ploy-A, sequenced adapters to 3′ ends, followed by purification and PCR amplification. The amplified DNA was sequenced using the Illumina NovaSeq 6000 platform after the library quality inspection.

Chloroplast genome assembly and annotation: *De novo* genome assembly from clean data was accomplished utilizing NOVOPlasty (Dierckxsens *et al*., 2017; Ding *et al*., 2020), using default settings, and genomes of *E. micractina* (OL622067), *E. pekinensis* (MZ707776) and *E. kansuensis* (MZ962400) as the reference. The genome annotation was performed using the program PGA (Qu *et al*., 2019) and GeSeq (Tillich *et al*., 2017) combined with manual correction. The correctness of the assembly was confirmed and verified manually. The complete chloroplast genome sequence of *E. altotibetica* obtained was deposited in the GenBank under accession number OR120371. Finally, chloroplast genome features of *E. altotibetica* were visualized using the online resource OGDRAW (Lohse *et al*., 2013).

Codon usage and repeat sequence analyses: We detected simple sequence repeats (SSRs) in the chloroplast genome sequence of *E. altotibetica* using the online program MISA (Beier *et al*., 2017). Within MISA, we set the minimum number of repeated units to 10, 6, 5, 5, 5, and 5 for mono-, di-, tri-, tetra-, penta- and hexanucleotides, respectively. Moreover, we used CodonW (Sharp & Li, 1987; Shields & Sharp, 1987) software to perform statistical analysis of RSCU in *E. altotibetica* chloroplast genome.

IR boundary regions and genome comparative analysis: We compared the chloroplast genome of *E. altotibetica* with those of nine species of *Euphorbia* to identify hotspots of genomic variations (Table S1). We especially visualized differences among the LSC/IRb/ SSC/IRa junctions of ten chloroplast genomes based on their annotations using IRscope (Amiryousefi *et al*., 2018). We further compared the chloroplast genome of *E. altotibetica* with nine additional species of *Euphorbia* using the Shuffle-LAGAN alignment method within mVISTA (Frazer *et al*., 2004). We used *E. peplus* (MZ678242.1) as a reference.

Nucleotide diversity and Ka/Ks ratio: We compared the chloroplast genome of *E. altotibetica* and its close relatives *E. fauriei* (OP477345.1), *E. pekinensis* (MZ707776.1), *E. micractina* (OL622067.1), *E. jolkinii* (LC661698.1), *E. helioscopia* (MN199031.1), *E. kansuensis* (MZ962400.1), *E. esula* (KY000001.1), *E. kansui* (MH392274.1), *E. peplus* (MZ678242.1) and *E. lathyris* (MT241376.1) to assess their nucleotide diversity (Table S1). To evaluate the complete nucleotide diversity (Pi) among chloroplast genomes of these ten species, the chloroplast genome sequences were aligned using MAFFT (Katoh *et al*., 2002) alignment tool to estimate the synonymous (*Ks*) and non-synonymous (*Ka*) substitution rates, and manually adjusted with Bioedit. We then carried out a sliding window analysis using DnaSP (DNA Sequences Polymorphism version V6.0) to obtain nucleotide diversity (Pi) and the *Ka/Ks* for each gene. We set window length of 600 bp and a step size of 200 bp, and selected protein-coding genes to detect synonymous (*Ks*) and non-synonymous (*Ka*) substitution rates (Kong *et al*., 2021).

Phylogenetic analysis: For the phylogenetic analysis of *Euphorbia*, we downloaded 78 chloroplast genomes from 68 species of genus *Euphorbia*, as well as two species i.e., *Triadica sebifera* (MT424756) and *Balakata baccata* (MW266130) as outgroups. All the genome sequences were obtained from GenBank (Table S1). The nucleotide sequences were aligned using MAFFT (Katoh *et al*., 2002) with default parameters and filtered it using Gblocks to remove ambiguously aligned regions. For the final alignment, we determined the best-fit substitution model (TVM+F+I+G4) via Model Finder, and performed phylogenetic analysis using maximum-likelihood (ML) method (Guindon *et al*., 2010). The ML method comprised reconstruction in IQ-TREE (Nguyen *et al*., 2015) with 5,000 ultrafast bootstrap replications, while the Bayesian inference (BI) method was performed using MrBayes. For the BI, we ran two independent chains of the Markov Chain Monte Carlo (MCMC) for 10^6 generations with sampling every 1,000 generations. At the end of the analysis, the average split frequency between chains was 0.01, indicating convergence, and we discarded the first 25% of all generations as burn-in, and the remaining trees were combined in a maximum clade credibility (MCC) tree.

Results

Chloroplast genome characteristics and annotation: The chloroplast genome of *E. altotibetica* has a typical quadripartite circular structure that is 161,141 bp in length and consisted of a large single copy (LSC, 90,586 bp) and a small single copy (SSC, 17,429 bp) region separated by a pair of inverted repeats (IRa and IRb, each with 26,563 bp) regions. The total GC content is 35.5%, and the highest GC content found in the two IR regions is 42.3%, while the GC content of LSC and SSC is 32.5% and 30.2%, respectively (Fig. 1, Table 1).

There are 130 encoded genes in total, including 84 protein coding genes (PCGs), 38 transfer RNAs (tRNAs), and eight ribosomal RNAs (rRNAs). Among 130 genes, 81 genes are located in the LSC, 13 in the SSC, and 19 are duplicated in the IR regions, which comprised five PCGs (*rpl2*, *rpl23*, *ndhB*, *rps7*, *rps12*), seven tRNAs (*trnA-UGC*, *trnI-GAU*, *trnI-CAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, *trnV-UGC*), four rRNAs (*rrn4.5S*, *rrn5S*, *rrn16S*, *rrn23S*), and three genes of unknown function (*ycf1*, *ycf2*, *ycf15*). Meanwhile, 12 genes also contain one intron of the total 130 genes. Of these, six are PCGs (*atpF*, *ndhA*, *ndhB*, *rpl16*, *rpl12*, *rpoC1*) and six are tRNAs

(*trnA-UGC*, *trnG-UCC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, *trnV-UAC*). Besides, only three genes (*clpP*, *rps12*, and *ycf3)* have two introns (Table 2).

Repeat sequences analysis: We detected 90 SSR loci representing two different categories of SSRs, including mononucleotides and dinucleotides, in the chloroplast genome of *E. altotibetica*. The number of mononucleotide repeats is 84 and that of dinucleotides repeats is six, whose length ranges between 13 and 18 bp (Table 3). Among these, 62 SSRs are in the LSC, 16 are in the SSC, six are in the IRa, and six are in the IRb. Most of SSRs (59) are located in the intergenic spacer (IGS) regions, and only a few (36) are present in the coding regions (CDS) (Table 4).

Table 1. The detail characteristics of the complete chloroplast genome of *E. altotibetica*.

Category	Items	Characteristics
Construction of chloroplast genome	LSC region (bp)	90,586
	IRA region (bp)	26,563
	SSC region (bp)	17,429
	IRB region (bp)	26,563
	Size of chloroplast genome (bp)	161,141
Gene content	Total genes	130
	Protein-coding genes	84
	tRNA	38
	rRNA	8
	Duplicate genes	20
	Genes on LSC region	85
	Genes on IRA region	19
	Genes on SSC region	13
	Genes on IRB region	19
GC content	GC content of LSC region (%)	32.5
	GC content of IRA region (%)	42.3
	GC content of SSC region (%)	30.2
	GC content of IRB region (%)	42.3
	Overall GC content (%)	35.5

Table 2. Gene annotation of the chloroplast genome of *E. altotibetica.*

Gene category	Gene group	Gene name
Genes for photosynthesis	Subunits of photosystem I	*psaA, psaB, psaC, psaI, psaJ*
	Subunits of photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ*
	Subunit of cytochrome b/f complex	*petA, petD, petG, petL, petN*
	Subunits of ATP synthase	*atpA, atpB, atpE, atpF*[1], *atpH, atpI*
	Subunits of NADH dehydrogenase	*ndhA*[1], *ndhB*[1*], *ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
	Large subunit of rubisco	*rbcL*
Self replication	DNA dependent RNA polymerase	*rpoA, rpoB, rpoC1*[1], *rpoC2*
	Small subunit of ribosome	*rps2, rps3, rps4, rps7*[*], *rps8, rps11, rps12*[2*], *rps14, rps15, rps18, rps19*
	Large subunit of ribosome	*rpl2*[1*], *rpl14, rpl16*[1], *rpl20, rpl22, rpl23*[*], *rpl33, rpl36*
	Transfer RNA gene	*trnA-UGC*[1*], *trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU, trnG-UCC*[1], *trnG-GCC, trnH-GUG, trnI-GAU*[1*], *trnI-CAU*[*], *trnK-UUU*[1], *trnL-CAA*[*], *trnL-UAA*[1], *trnL-UAG, trnM-CAU, trnN-GUU*[*], *trnP-UGG, trnQ-UUG, trnR-UCU, trnR-ACG*[*], *trnS-GCU, trnS-UGA, trnS-GGA, trnT-GGU, trnT-UGU, trnV-UAC*[1], *trnV-GAC*[*], *trnW-CCA, trnY-GUA*
	Ribosomal RNA gene	*rrn4.5S*[*], *rrn5S*[*], *rrn16S*[*], *rrn23S*[*]
Other genes	Maturase	*matK*
	Envelop membrane protein	*cemA*
	C-type eytochrome synthesis gene	*ccsA*
	Submit of acetyl-CoA-carboxylase	*accD*
	ATP-dependent protease subunit P	*clpP*[2]
Genes of unknown function	Conserved open reading frame	*ycf1*[*], *ycf2*[*], *ycf3*[2], *ycf4, ycf15*[*]

Note: [1] Genes with one intron. [2] Genes with two introns. [*] Duplicated genes located in the IR regions.

Table S1. Species names and accession numbers of the complete chloroplast genome sequences.

No.	Species names	Gene Bank	No.	Species names	Gene Bank
1.	*Euphorbia peplus*	MZ678242.1	36.	*Euphorbia invenusta*	MT395005.1
2.	*Euphorbia micractina*	OL622067.1	37.	*Euphorbia invenusta_var. Augusta*	MT394997.1
3.	*Euphorbia kansui*	MH392274.1	38.	*Euphorbia neostolonifera*	MT395023.1
4.	*Euphorbia esula*	KY000001.1	39.	*Euphorbia rhizophora*	MT395020.1
5.	*Euphorbia kansuensis*	MZ962400.1	40.	*Euphorbia schubei*	MT395017.1
6.	*Euphorbia helioscopia*	MN199031.1	41.	*Euphorbia lugardae*	MT395019.1
7.	*Euphorbia pekinensis*	MZ707776.1	42.	*Euphorbia succulenta*	MT395015.1
8.	*Euphorbia fauriei*	OP477345.1	43.	*Euphorbia ritchiei*	MT394998.1
9.	*Euphorbia lathyris*	MT241376.1	44.	*Euphorbia renneyi*	MT394999.1
10.	*Euphorbia jolkinii*	LC661698.1	45.	*Euphorbia neovirgata*	MT395018.1
11.	*Triadica sebifera*	MT424756.1	46.	*Euphorbia neogillettii*	MT395021.1
12.	*Balakata baccata*	MW266130.1	47.	*Euphorbia lindenii*	MT395026.1
13.	*Euphorbia neoglabrata*	MT394996.1	48.	*Euphorbia echinulata*	MT395011.1
14.	*Euphorbia neocymosa*	MT395006.1	49.	*Euphorbia bisglobosa*	MT395041.1
15.	*Euphorbia umbellata*	MT395046.1	50.	*Euphorbia orobanchoides*	MT395043.1
16.	*Euphorbia mbuinzauensis*	MT395000.1	51.	*Euphorbia ampliphylla*	MT395036.1
17.	*Euphorbia bicompacta*	MT395030.1	52.	*Euphorbia poissonii*	MT395035.1
18.	*Euphorbia bicompacta_var.*	MT395029.1	53.	*Euphorbia drupifera*	MW496383.1
19.	*Euphorbia pereskiifolia*	MT395004.1	54.	*Euphorbia milii*	MN713924.1
20.	*Euphorbia kirkii*	MW300679.1	55.	*Euphorbia hedyotoides*	MT395028.1
21.	*Euphorbia syncameronii*	MT395039.1	56.	*Euphorbia alluaudii*	MT395034.1
22.	*Euphorbia pseudolaevis*	MT395047.1	57.	*Euphorbia tirucalli*	MH890571.1
23.	*Euphorbia mafingensis*	MW300676.1	58.	*Euphorbia pteroneura*	MW496386.1
24.	*Euphorbia cupricola*	MW300677.1	59.	*Euphorbia enteropgora*	MT395033.1
25.	*Euphorbia discoidea*	MT395042.1	60.	*Euphorbia thymifolia*	MW496379.1
26.	*Euphorbia neogossweileri*	MT395027.1	61.	*Euphorbia humifusa*	OM791345.1
27.	*Euphorbia neoglaucescens*	MT395038.1	62.	*Euphorbia hirta*	MW822040.1
28.	*Euphorbia pseudomollis*	MW300680.1	63.	*Euphorbia prostrata*	ON631059.1
29.	*Euphorbia aff. Neococcinea Luke s.n.*	MT395031.1	64.	*Euphorbia maculata*	MT830858.1
30.	*Euphorbia neoarborescens*	MT395032.1	65.	*Euphorbia schlechtendalii*	MW496378.1
31.	*Euphorbia magnifica*	MT395016.1	66.	*Euphorbia espinosa*	MW496384.1
32.	*Euphorbia sp. NW-2021b*	MW300678.1	67.	*Euphorbia larica*	MN646683.1
33.	*Euphorbia biselegans*	MT395022.1	68.	*Euphorbia scheffleri*	MT395025.1
34.	*Euphorbia neorubella*	MT395003.1	69.	*Euphorbia smithii*	MN646684.1
35.	*Euphorbia guentheri*	MT395002.1			

Table 3. Number of SSRs identified in the chloroplast genome of *E. altotibetica.*

Repeat unit	Type	Number	Largest repeat
1	A	41	15
	C	1	13
	T	42	18
2	AT	4	7
	TA	2	8
Total	**5**	**90**	-

Analysis of codon usage: The codon usage bias and RSCU were analyzed based on 52 CDS sequences of the chloroplast genome of *E. altotibetica*. Our results indicate that protein-coding genes have a total of 21,462 codons, representing 61 amino acids and three stop codons. Among these codons, isoleucine is the most abundant (8.78%), followed by glycine (6.83%) and phenylalanine (6.06%), whereas cysteine is the least abundant, representing only 1.11% (Fig. 2, Table 5).

We also detected 31 degenerate codons with RSCU values above one, indicating their usage bias in the chloroplast genome of *E. altotibetica*. In particular, the UUA codon encoding leucine had the highest usage bias with a value of 2.11. In addition, almost all RSCU values were greater than one in the A/U-ending codons, while all RSCU values were less than 1 in the C/G-ending codons, except for UUG. Met (AUC) and Trp (UGG) were encoded by only one codon, with no codon preference (Table 5).

IR expansion and contraction analysis of *Euphorbia*: In this study, the contraction and expansion of IR boundaries were visualized using ten chloroplast genomes of *Euphorbia* (Table S1). IR regions have four boundaries with LSC-IRb, namely IRb-SSC, SSC-IRa, IRa-LSC, and LSC-IRb. We compared the difference in the binding sites of IR/LSC and IR/SSC of the ten *Euphorbia* species. Our analysis revealed that these ten species have roughly similar genetic composition and structure (Fig. 3).

Table 4. SSR information of the chloroplast genome in *E. altotibetica*.

No.	SSR type	SSR	Size	Start	End	Location
1.	p1	(T)14	14	494	507	IGS
2.	p1	(A)12	12	4,877	4,888	IGS
3.	p1	(T)10	10	5,226	5,235	IGS
4.	p1	(A)10	10	6,466	6,475	IGS
5.	p1	(A)11	11	7,916	7,926	IGS
6.	p1	(A)10	10	8,819	8,828	IGS
7.	p1	(A)13	13	9,471	9,483	Intron (*trnS-GCU*)
8.	p1	(T)10	10	9,887	9,896	IGS
9.	p1	(A)12	12	11,549	11,560	Intron (*trnR-UCU*)
10.	p2	(AT)6	12	11,664	11,675	IGS
11.	p1	(T)14	14	11,782	11,795	CDS (*atpA*)
12.	p1	(T)11	11	14,100	14,110	CDS (*atpF*)
13.	p1	(A)11	11	14,492	14,502	CDS (*atpF*)
14.	p1	(T)12	12	14,776	14,787	IGS
15.	p1	(T)11	11	15,797	15,807	IGS
16.	p1	(T)10	10	16,212	16,221	CDS (*atpI*)
17.	p1	(A)10	10	18,406	18,415	CDS (*rpoC2*)
18.	p1	(A)10	10	19,493	19,502	CDS (*rpoC2*)
19.	p1	(T)14	14	20,343	20,356	CDS (*rpoC2*)
20.	p1	(A)10	10	24,749	24,758	CDS (*rpoC1*)
21.	p1	(T)12	12	28,047	28,058	CDS (*rpoB*)
22.	p1	(A)11	11	29,158	29,168	IGS
23.	p1	(A)12	12	29,777	29,788	IGS
24.	p1	(C)13	13	30,778	30,790	IGS
25.	p1	(T)10	10	31,680	31,689	IGS
26.	p1	(A)10	10	32,028	32,037	IGS
27.	p2	(AT)6	12	32,433	32,444	IGS
28.	p1	(A)15	15	32,681	32,695	IGS
29.	c	(A)10ttttcaatg(A)12	31	33,133	33,163	IGS
30.	p1	(T)11	11	34,392	34,402	IGS
31.	p1	(A)10	10	35,451	35,460	IGS
32.	p1	(T)14	14	38,449	38,462	Intron (*trnS-UGA*)
33.	p1	(T)10	10	38,703	38,712	IGS
34.	p2	(AT)7	14	41,221	41,234	IGS
35.	p1	(A)10	10	48,203	48,212	CDS (*ycf3*)
36.	p1	(A)12	12	50,498	50,509	IGS
37.	p1	(T)12	12	52,340	52,351	IGS
38.	p1	(T)10	10	52,755	52,764	IGS
39.	c	(T)15caatttattacatattttctatattaaatagtttaaaa atttattaaatttctattatattactattattatattac(TA)6	104	54,778	54,881	IGS
40.	p1	(T)10	10	58,823	58,832	CDS (*atpB*)
41.	c	(A)10ttg(T)11	24	61,472	61,495	IGS
42.	p1	(A)11	11	64,198	64,208	IGS
43.	p1	(T)11	11	64,730	64,740	IGS
44.	p1	(A)10	10	66,040	66,049	CDS (*cemA*)
45.	p2	(TA)8	16	66,886	66,901	IGS
46.	p1	(T)11	11	68,075	68,085	IGS
47.	p1	(T)18	18	68,200	68,217	IGS
48.	p1	(A)11	11	68,459	68,469	IGS

Table 4. (Cont'd.).

No.	SSR type	SSR	Size	Start	End	Location
49.	p1	(T)10	10	68,650	68,659	IGS
50.	p1	(A)10	10	69,260	69,269	CDS (*psbF*)
51.	p1	(A)13	13	71,550	71,562	IGS
52.	p1	(A)11	11	71,687	71,697	IGS
53.	c	(A)10tattcaatatatattcatcaat(A)15	47	73,829	73,875	IGS
54.	p1	(A)12	12	74,250	74,261	IGS
55.	c	(T)13agtttgactactttacttgatcatattatcatgatat tttatcataccaatttctattctaccttcccggaa(T)11	96	74,408	74,503	IGS
56.	p1	(A)11	11	75,121	75,131	IGS
57.	p1	(A)15	15	76,404	76,418	Intron (*clpP*)
58.	p1	(A)10	10	76,571	76,580	Intron (*clpP*)
59.	p1	(T)10	10	77,282	77,291	Intron (*clpP*)
60.	p1	(T)10	10	77,584	77,593	Intron (*clpP*)
61.	p1	(A)10	10	82,537	82,546	IGS
62.	p1	(A)11	11	86,021	86,031	IGS
63.	p1	(T)10	10	87,386	87,395	IGS
64.	p2	(TA)7	14	88,171	88,184	CDS (*rpl16*)
65.	p1	(T)12	12	88,287	88,298	CDS (*rpl16*)
66.	p1	(T)13	13	88,702	88,714	CDS (*rpl16*)
67.	p1	(T)10	10	90,066	90,075	IGS
68.	p1	(T)14	14	90,848	90,861	IGS
69.	p1	(T)10	10	100,559	100,568	IGS
70.	p1	(T)13	13	105,497	105,509	IGS
71.	p1	(T)13	13	109,534	109,546	Intron (*trnI-GAU*)
72.	p1	(A)12	12	114,551	114,562	IGS
73.	p1	(A)14	14	116,884	116,897	CDS (*ycf1*)
74.	p1	(A)10	10	117,901	117,910	CDS (*ndhF*)
75.	p1	(A)13	13	120,049	120,061	IGS
76.	p1	(A)10	10	120,171	120,180	IGS
77.	p1	(A)12	12	120,320	120,331	IGS
78.	p1	(T)12	12	120,547	120,558	IGS
79.	p1	(T)11	11	121,246	121,256	CDS (*ccsA*)
80.	p1	(T)12	12	121,686	121,697	IGS
81.	p1	(T)11	11	123,284	123,294	IGS
82.	p1	(A)11	11	130,017	130,027	IGS
83.	p2	(AT)6	12	130,211	130,222	CDS (*ycf1*)
84.	p1	(T)10	10	131,820	131,829	CDS (*ycf1*)
85.	p1	(T)12	12	132,477	132,488	CDS (*ycf1*)
86.	p1	(T)10	10	132,657	132,666	CDS (*ycf1*)
87.	p1	(T)11	11	133,098	133,108	CDS (*ycf1*)
88.	p1	(A)10	10	134,353	134,362	CDS (*ycf1*)
89.	p1	(T)16	16	134,529	134,544	CDS (*ycf1*)
90.	p1	(T)14	14	134,831	134,844	CDS (*ycf1*)
91.	p1	(T)12	12	137,166	137,177	IGS
92.	p1	(A)13	13	142,182	142,194	Intron (*trnI-GAU*)
93.	p1	(A)13	13	146,219	146,231	IGS
94.	p1	(A)10	10	151,160	151,169	IGS
95.	p1	(A)14	14	160,867	160,880	IGS

Note: P represents a single SSR type, The number in P1/P2 indicates the number of bases constituting the motif, respectively; C represents the compound SSR type; IGS represents the intergenic region.

Table 5. RSCU analysis of protein-coding region in *E. altotibetica*.

Amino acid	Codon	Number	RSCU	Ratio/%
Phe	UUU	885	1.36	6.06%
	UUC	416	0.64	
Leu	UUA	797	2.11	5.68%
	UUG	423	1.12	
Ser	UCU	456	1.70	5.49%
	UCC	244	0.91	
	UCA	334	1.25	
	UCG	144	0.54	
Tyr	UAU	658	1.63	3.76%
	UAC	148	0.37	
Cys	UGU	182	1.52	1.11%
	UGC	57	0.48	
Trp	UGG	381	1.00	1.78%
Leu	CUU	478	1.27	4.87%
	CUC	127	0.34	
	CUA	294	0.78	
	CUG	146	0.39	
Pro	CCU	348	1.59	4.08%
	CCC	161	0.74	
	CCA	256	1.17	
	CCG	111	0.51	
His	CAU	399	1.54	2.41%
	CAC	119	0.46	
Gln	CAA	599	1.57	3.55%
	CAG	162	0.43	
Arg	CGU	264	1.29	3.41%
	CGC	94	0.46	
	CGA	292	1.42	
	CGG	82	0.40	
Ile	AUU	948	1.51	8.78%
	AUC	327	0.52	
	AUA	610	0.97	
Met	AUG	473	1.00	2.20%
Thr	ACU	457	1.69	5.03%
	ACC	180	0.67	
	ACA	341	1.26	
	ACG	101	0.37	
Asn	AAU	840	1.56	5.02%
	AAC	238	0.44	
Lys	AAA	924	1.57	5.50%
	AAG	257	0.44	
Ser	AGU	331	1.23	2.01%
	AGC	100	0.37	
Arg	AGA	383	1.87	2.33%
	AGG	116	0.57	
Val	GUU	411	1.47	5.22%
	GUC	143	0.51	
	GUA	418	1.49	
	GUG	148	0.53	
Ala	GCU	529	1.84	5.36%
	GCC	176	0.61	
	GCA	321	1.12	
	GCG	124	0.43	
Asp	GAU	697	1.61	4.04%
	GAC	171	0.39	
Glu	GAA	905	1.54	5.47%
	GAG	270	0.46	
Gly	GGU	504	1.38	6.83%
	GGC	149	0.41	
	GGA	582	1.59	
	GGG	231	0.63	

The boundary of JLB (LSC/IRb) was situated in the coding region between *rpl22* and *rps19* genes, and *rps19* and *rpl2* genes, while JLA (IRa/LSC) boundary is present between *rpl2* and *trnH* genes of *Euphorbia*. Gene deletion of *rps19* exists in the IRa region of *E. altotibetica* and *E. kansuensis*. *Ycf1* gene crossed JSB (IRb/SSC) and JSA (SSC/IRa) boundary regions of all chloroplast genomes of *Euphorbia*, but it was absent in the IRb region of *E. fauriei*. Besides, we observed several expansions and contractions of IR/LSC and IR/ SSC boundary regions. For example, *rps19* gene of *E. altotibetica*, *E. peplus*, and *E. kansuensis* crossed the LSC/IRb boundary region; however it is contained within the LSC or IRb regions in the other seven species of *Euphorbia*. In addition, *ndhF* gene is deleted within SSC region of *E. lathyris*, but it is smaller and restricted within SSC region in other nine *Euphorbia* species (Fig. 3).

For the gene position relative to IRs, *ycf1* gene crossed the SSC/IRa boundary in all species and are about five times larger than that of SSC/IRb region (5693 bp to 5789 bp vs. 995 bp to 1430 bp). The position of *trnH* gene is variable among all taxa analyzed, and mostly located in the LSC region. But it slightly crossed the boundary with the IRa in *E. helioscopia*, *E. micractina*, *E. pekinensis* and *E. fauriei* (Fig. 3).

Comparative analysis of chloroplast genomes in *Euphorbia*: Comparisons of nine chloroplast genomes of *Euphorbia* using a multiple sequence alignment, our result revealed they are highly conserved structures. Gene coding regions are more conserved than noncoding regions, and IRs are more conserved than LSC and SSC regions. Besides, we observed that the intergenic spacer regions between several pairs of genes varied greatly, especially in the single copy (SC) and non-coding regions, such as *trnH-psbA*, *trnK-trnQ*, *trnG-trnR*, *psbM-trnD*, *trnT-psbD*, *psbZ-trnG*, *trnF-ndhJ*, *ndhC-trnV*, *accD-psaI*, *psbE-petL*, *rpl33-rps18*, and *ndhF-trnL*. In addition, *atpF*, *rpl16*, and *ycf1* genes showed high levels of variation within their protein coding regions (Fig. 4, Table S1).

Analyses of nucleotide diversity and Ka/Ks ratio: The analysis of polymorphic sites among *Euphorbia* species indicated that the average value of nucleotide diversity (Pi) is 0.01972, ranging from 0.09479 to 0.0003. Specifically, we detected a total of ten regions with high levels of nucleotide diversity (Pi > 0.06), including *psbI*, *trnG-trnR*, *trnY-trnE*, *trnT-psbD*, *psaA-ycf3*, *trnF-ndhJ*, *accD*, *trnP-psaJ*, *psbT-psbN* and *rpoA-rps11*, all of which existed in the LSC region (Fig. 5). Meanwhile, compared with *E. altotibetica*, we also calculated the *Ka/Ks* substitution rates of protein-coding genes for ten *Euphorbia* species, respectively (Table S1). Most of genes had *Ka/Ks* values less than 1 and the change trend of ratios is similar, except that *ycf 1*, *ycf 2*, *rpl 20* and *rpl 22* showed a higher ratio than 1. Besides, genes with values zero or uncalculatable were excluded from the analysis (Fig. 6).

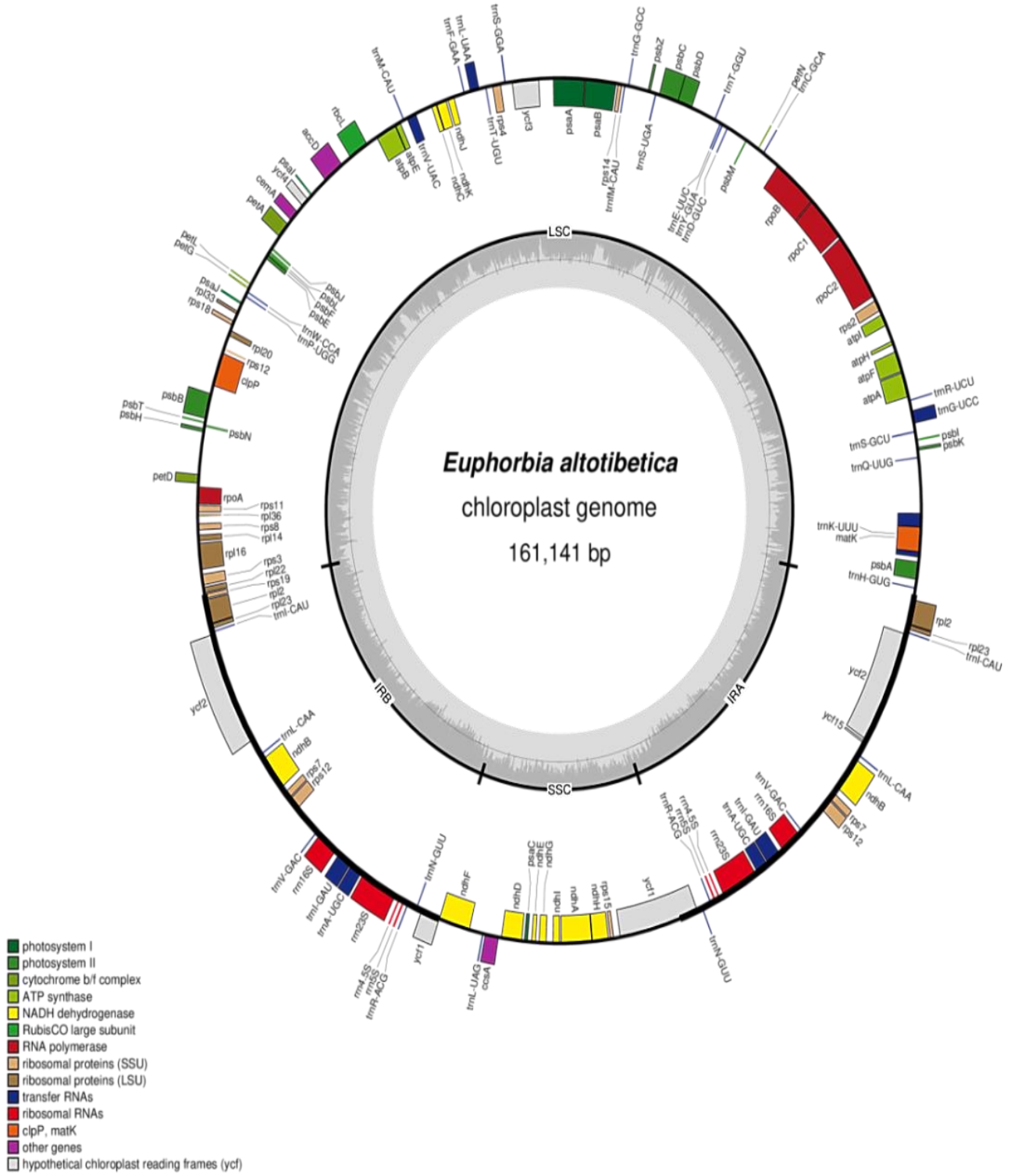

Fig. 1. Map of the chloroplast genome of *E. altotibetica*.
Note: Genes on the outer and inner portions of the circle are transcribed in clockwise and counterclockwise directions, respectively. Genes belonging to different functional groups are color-coded, and the inner circle shows their locations within the LSC, SSC, IRa, and IRb regions of the chloroplast genome. The dark-gray bars in the inner circle correspond to GC content.

Phylogenetic analysis: We reconstructed the phylogenetic trees based on the chloroplast genomes of 68 *Euphorbia* species with ML and BI approaches, using *Triadica sebifera* and *Balakata baccata* as outgroups (Table S1). The results showed that 68 *Euphorbia* species constituted a monophyletic group. Besides monophyletic group, it also formed two major clades (Clade 1 and Clade 2) with a high bootstrap support (100%). Clade 1 consists of *E. fauriei*, *E. pekinensis*, *E. micractina*, *E. jolkinii*, *E. helioscopia*, *E. kansuensis*, *E. esula*, *E. kansui*, *E. peplus* and *E. altitibetica*, all of which in this clade belong to *Euphorbia* Subgen. Esula Pers. However, the other attribute of clade 2, the *E. altotibetica* nested closely with the *E. peplus* indicating them to be the closest relatives (Fig. 7).

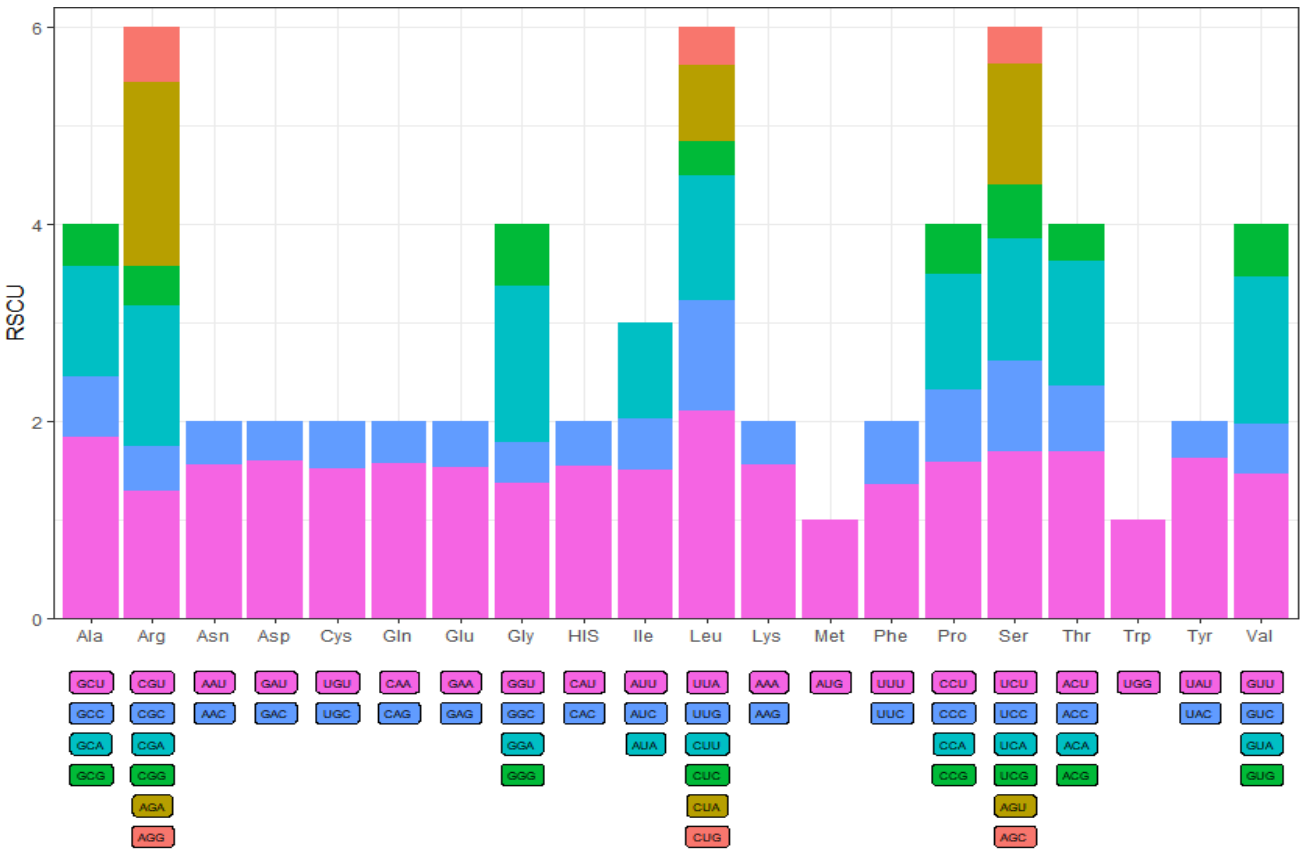

Fig. 2. RSCU analysis of each amino acid in *E. altotibetica*.
Note: The bars (color-coded) depict the relative synonymous codon usage values for each codon.

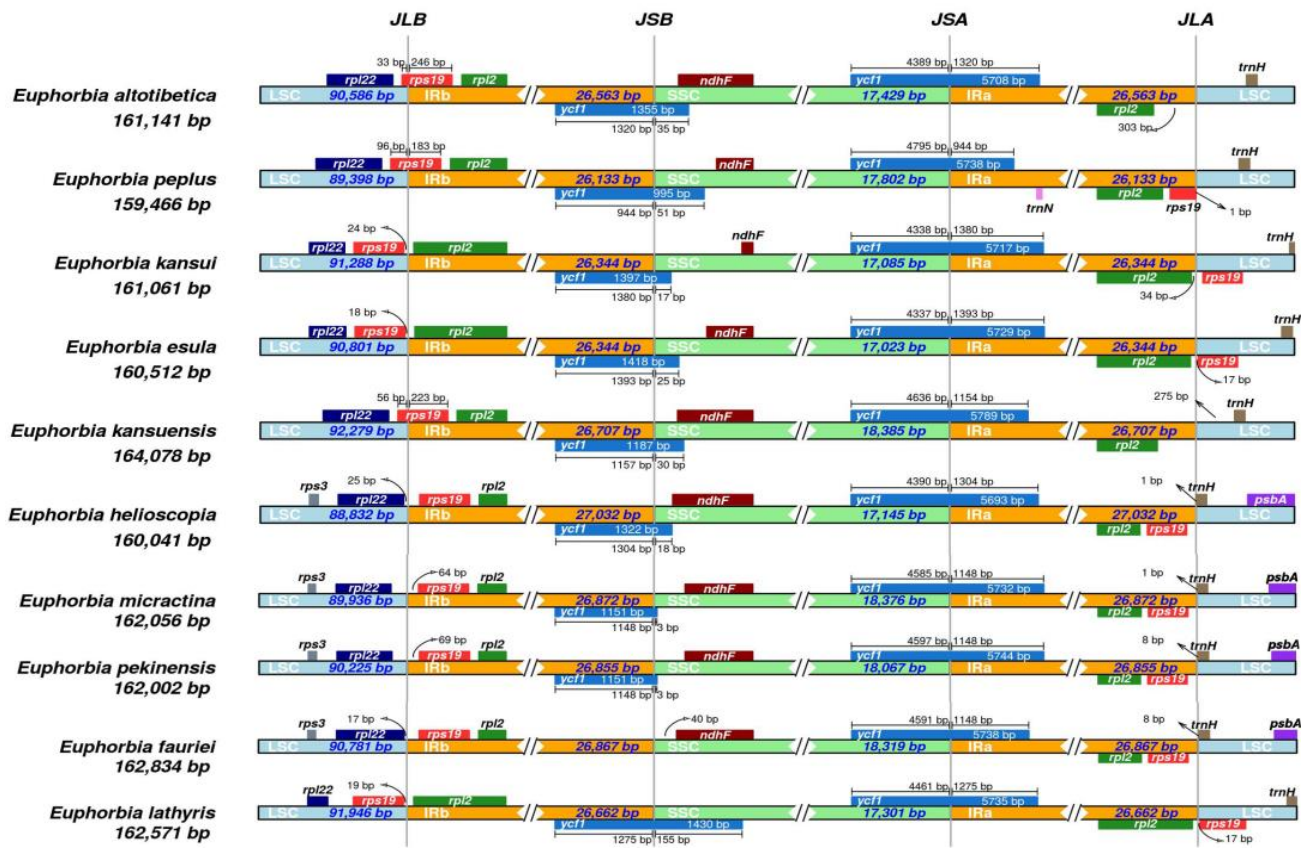

Fig. 3. Boundary analysis from IR regions of chloroplast genomes in ten *Euphorbia* species.
Note: JLB: junction line between LSC and IRB; JSB: junction line between IRB and SSC; JSA: junction line between SSC and IRA; JLA: junction line between IRA and LSC.

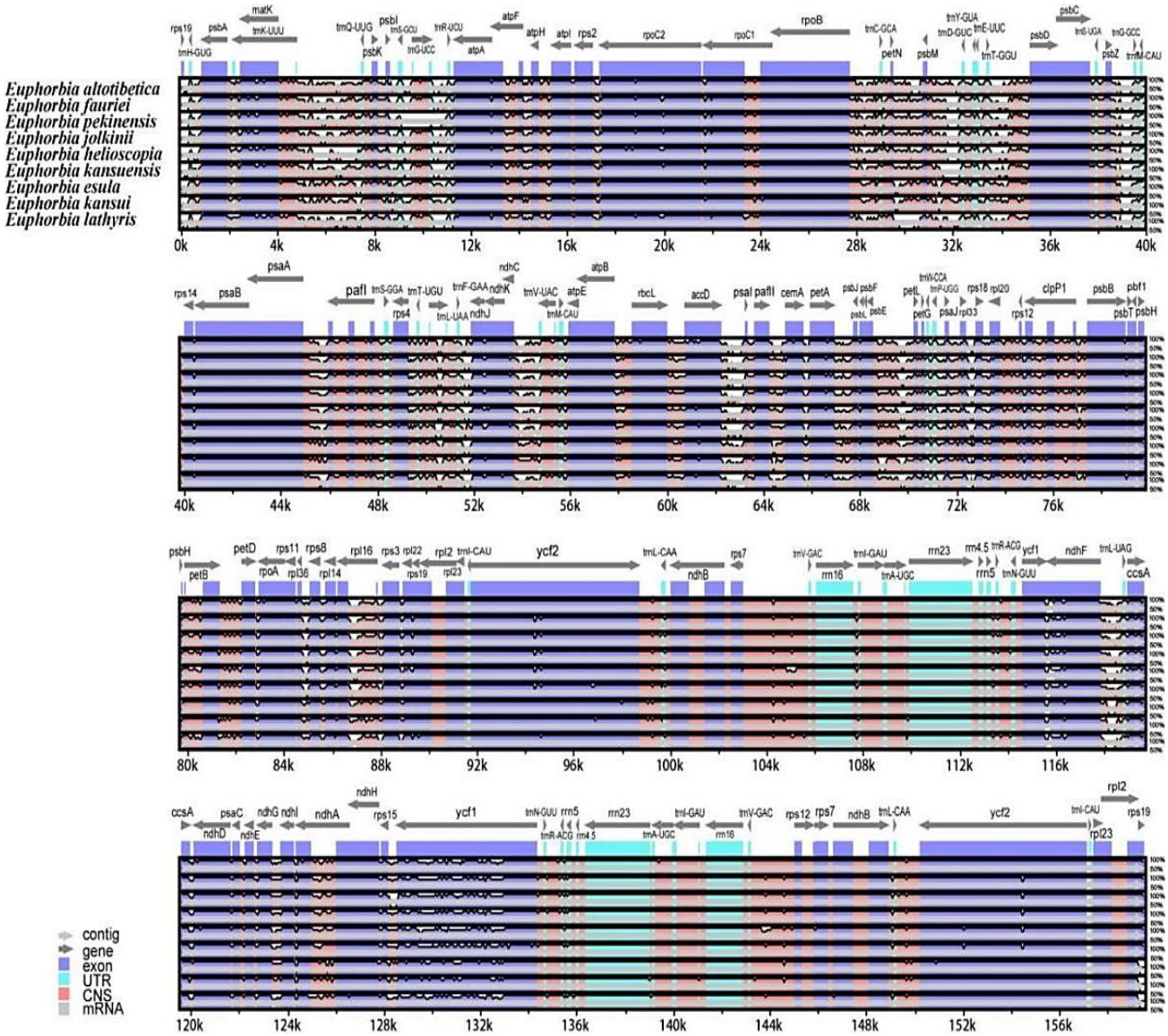

Fig. 4. Multiple sequence alignments and visualization of nine chloroplast genomes from *Euphorbia were* performed with mVISTA using *E. peplus* as a reference.

Note: The top arrows show transcription direction; exons of protein-coding genes are marked in purple, cyan indicates tRNAs and rRNAs, and non-coding as red. The *X*-axis represents the positions in the chloroplast genome, while the *Y*-axis indicates the percent of identity, ranging from 50% to 100%.

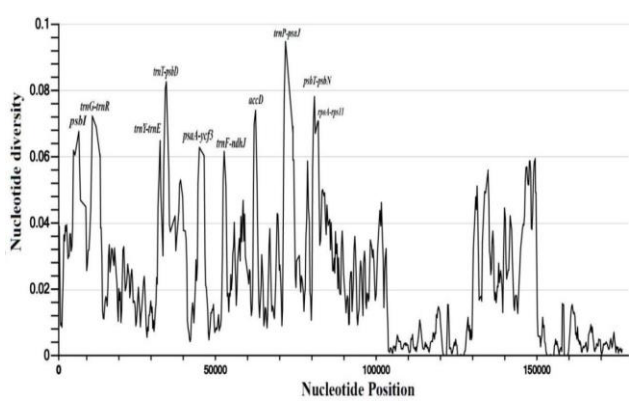

Fig. 5. Nucleotide diversity (Pi) analysis with a sliding window size of 600 bp and a step size of 200 bp.

Note: The *X*-axis shows nucleotide positions of the midpoint of a window, while the *Y*-axis represents nucleotide diversity values.

Discussion

In the present study, we presented the assembly and annotation of the chloroplast genome for *E. altotibetica* for the first time and compared it with other *Euphorbia* species. The chloroplast genome analyses of these species broaden our knowledge regarding the phylogenetic relationships of species within the genus *Euphorbia*.

Chloroplast is the most important organelle and photosynthesis site of higher plants, that contains rich genetic information in the form of second-largest genome (Song *et al*., 2017). Recently, the database of chloroplast genomes has been improved gradually, especially for medicinal plants (Jiang *et al*., 2020). Previous studies indicated that chloroplast genome sequences have been widely used to taxonomic revision, population genetic structure, genetic diversity, historical population dynamics, and phylogenetic relationships (Xiong *et al*.,

2009; Chen *et al*., 2010; Twyford & Ness, 2017). A lot of data revealed that the structure and gene content of chloroplast genomes for angiosperms were highly conservative. Almost all chloroplast genomes have a circular and quadripartite structure (Bendich, 2004), including a large single copy (LSC, 80~90 kb) and a small single copy (SSC, 16~27 kb) region separated by a pair of inverted repeats (IRa and IRb, each with 20~28 kb) regions, whose length is probably between 120 and 180 kb (Raubeson *et al*., 2007). GC content ranged from 30% to 40% (Zhang *et al*., 2012). Besides, coding genes often contain protein coding genes, tRNA coding genes and rRNA coding genes (Daniell *et al*., 2016).

Our results revealed that the chloroplast genome of *E. altotibetica* has a circular and quadripartite structure with the length of 161,141 bp, including a large single copy (LSC, 90,586 bp) and a small single copy (SSC, 17,429 bp) region separated by a pair of inverted repeats (IRa and IRb, each with 26,563 bp) regions, that is similar to those of other *Euphorbia* species and other angiosperms reported previously (Zhao *et al*., 2018; Li *et al*., 2019; Zhao *et al*., 2020; Ma *et al*., 2020; Zhang *et al*., 2021; Yu *et al*., 2022; Wang *et al*., 2022). The total GC content is 35.5% and the highest GC value is found in IR regions (42.3%) in our analysis, that may be caused by the high GC content of tRNA genes in reverse repeat IR regions (Jiang *et al*., 2021). The total number of coding genes of the chloroplast genome in *E. altotibetica* is 130, including 84 PCGs, 38 tRNAs and eight rRNAs. Among these, 12 genes contain one intron, while three of them ones have two introns. This result is similar to the closely related species of *E. lathyris* (encoding 128 genes, comprising 85 protein-coding, 35 tRNA genes and eight rRNA genes) (Ma *et al*., 2020) and *E. pekinensis* (encoding 129 functional genes, comprising 85 protein-coding genes, 36 tRNA genes, and eight rRNA genes) (Wang *et al*., 2022).

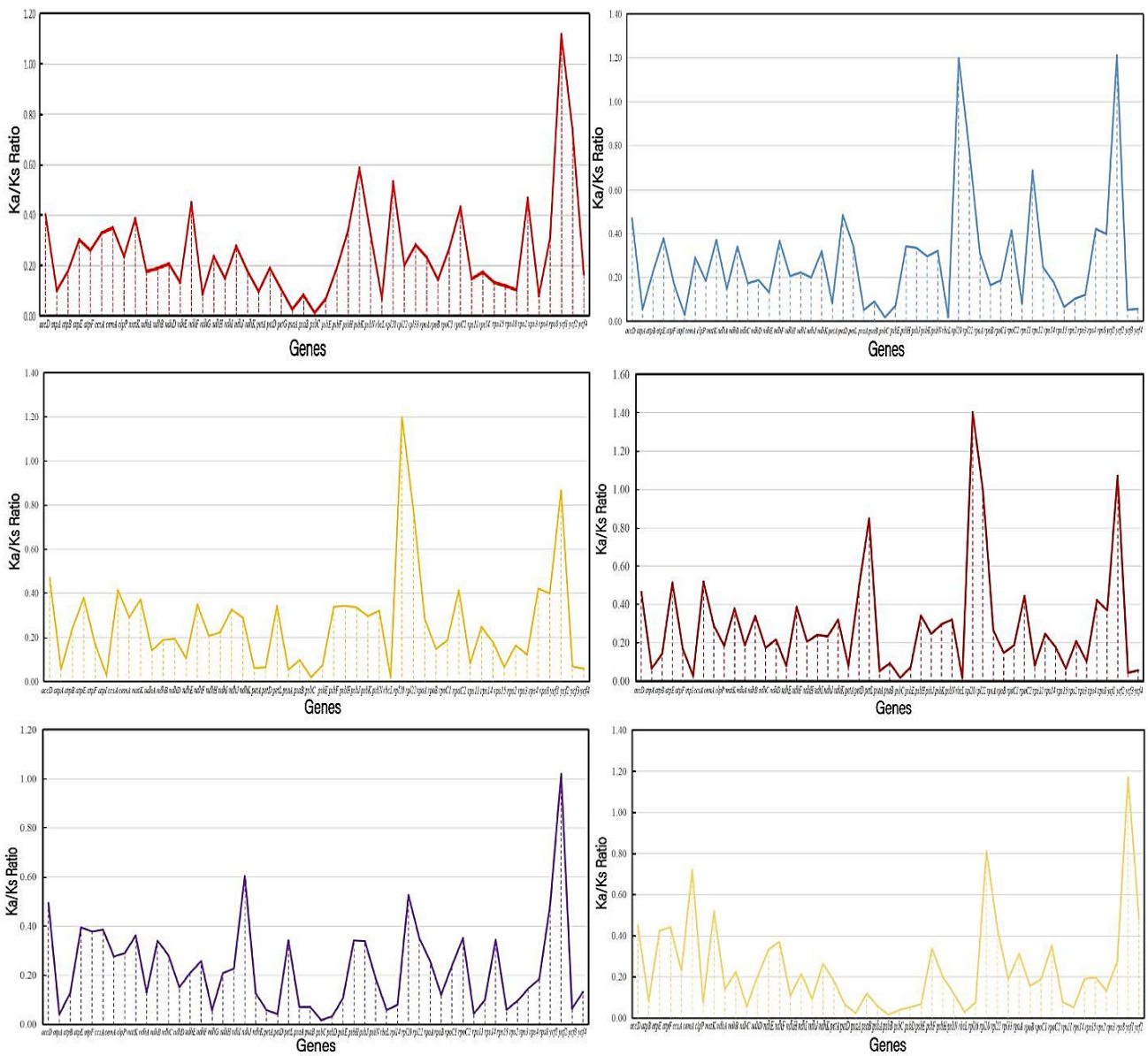

Fig. 6. The *Ka/Ks* ratio of protein-coding genes.
Note: Genes, where *Ka/Ks* values were zero and uncalculatable are not shown. Crimson show *E. peplus* (MZ678242.1) as the reference; Blue show *E. kansui* (MH392274.1) as the reference; Gold show *E. kansuensis* (MZ962400.1) as the reference; DarkRed show *E. helioscopia* (MN199031.1) as the reference; Purple show *E. jolkinii* (LC661698.1) as the reference; Yellow show *E. pekinensis* (MZ707776.1) as the reference.

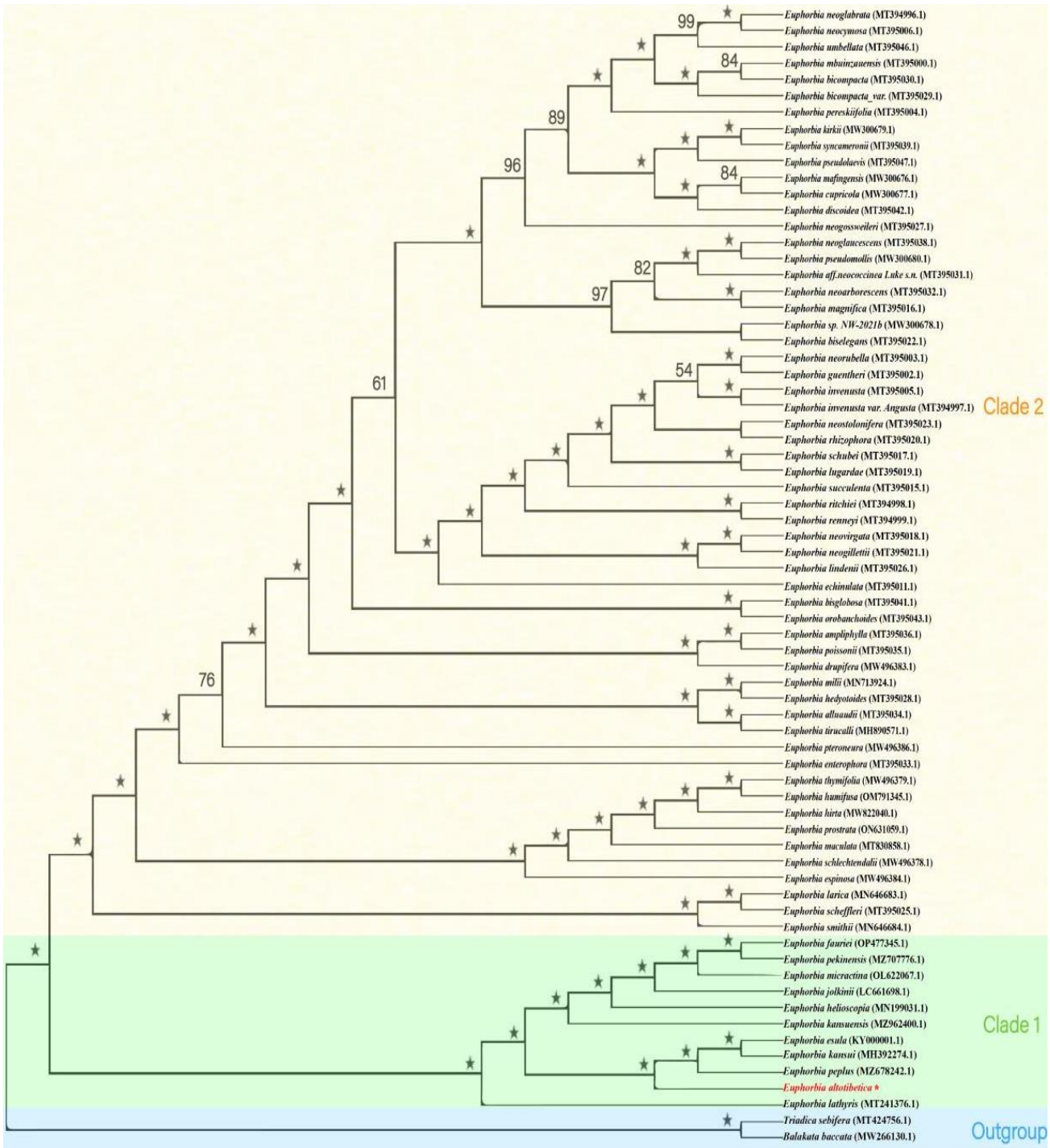

Fig. 7. Molecular phylogenetic analysis of *E. altotibetica*.
Note: Numbers above branches represent support values for the ML and BI analyses, respectively. Stars ("★") represent fully supported nodes (i.e., 100% bootstrap support and 1.00 pp) in the two analyses.

Simple sequence repeats (SSRs) sequences are tandem repeats of short motifs (1-6 bp) (Yuan *et al*., 2021), that are an important part of genomes in higher eukaryotes. They have been widely used as molecular markers for population genetics and evolutionary studies, because of they provide high rates of polymorphism at the species level and exhibit a codominant inheritance pattern (Powell *et al*., 1995; Pugh *et al*., 2004). The copy number variants of chloroplast SSRs are a significant molecular marker because it had a larger taxonomic distance than that of nuclear genes and mitochondrial SSRs (Deguilloux *et al*., 2004; Redwan *et al*., 2015). Thus, it plays a vital role in species delimitation, systematics and evolution, population genetic of polymorphism and genetic map construction (Du *et al*., 2012; Yang *et al*., 2016; Dong *et al*., 2016; Asaf *et al*., 2017). In the present study, we found that the dominant repeat motifs of SSRs within the chloroplast genome of *E. altotibetica* are 84 mononucleotide and six di-nucleotide. The AT/TA is the main type of di-nucleotide repeats, that is also reported by the previous studies in angiosperm (Du *et al*., 2020). Our results further provide that SSRs are mainly composed of short A and T, instead of C or G (Kuang *et al*., 2011; Peng *et al*., 2022).

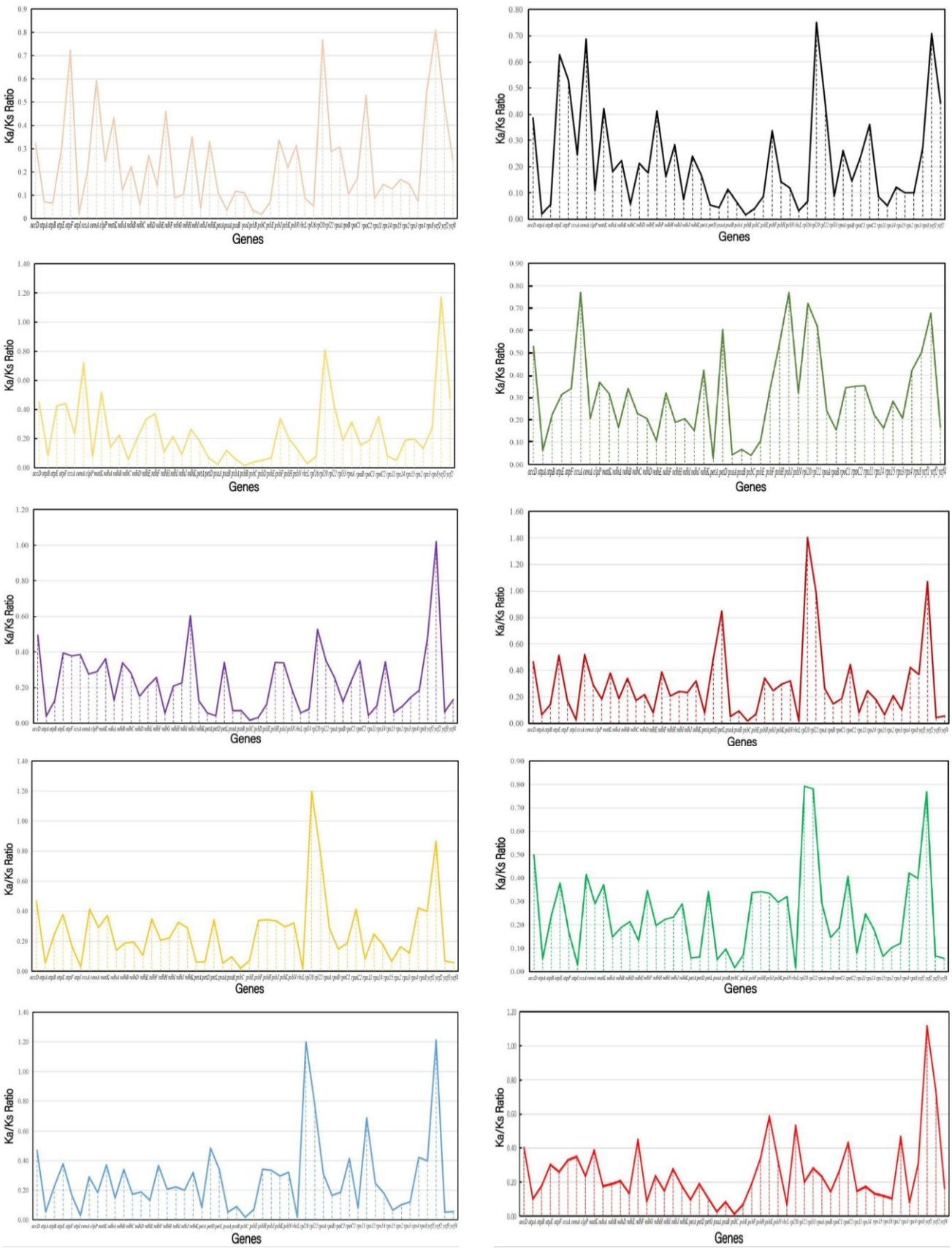

Fig. S1. The *Ka/Ks* ratio of protein-coding genes.
Note: Genes, where *Ka/Ks* values were zero and uncalculatable are not shown. Pink show *E. lathyris* (MT241376.1) as the reference; Black show *E. fauriei* (OP477345.1) as the reference; Gold show *E. kansuensis* (MZ962400.1) as the reference; YellowGreen show *E. micractina* (OL622067.1) as the reference; Purple show *E. jolkinii* (LC661698.1) as the reference; DarkRed show *E. helioscopia* (MN199031.1) as the reference; Yellow show *E. pekinensis* (MZ707776.1) as the reference; BlueGreen show *E. esula* (KY000001.1) as the reference; Blue show *E. kansui* (MH392274.1) as the reference; Crimson show *E. peplus* (MZ678242.1) as the reference.

Previous studies indicated that methionine (Met) and tryptophan (Trp) were encoded by unique codon, while the other 18 amino acids were coded by two to six codons (Campos *et al*., 2013). For many plants, the usage frequency of specific codons was higher than that of other synonymous codons, which was called codon usage bias (Campos *et al*., 2013). Codon usage bias is an important evolutionary feature for species that could reflect the origin and evolution mode of genes and species, and could affect gene function and expression (Staden & McLachlan, 1982; Li *et al*., 2017). Analysis of codon usage bias of the chloroplast genome for *E. altotibetica* demonstrated that isoleucine (Ile) has the highest proportion. There are 29 synonymous codons having the values of RSCU above one, of which 28 codons ended with A/U and only one codon with G, which is consistent with the codon usage bias of most angiosperms reported previously (Li *et al*., 2019; Zhao *et al*., 2020). In addition, previous study has also found that the codons ending with A or U are the preferred codons in plant chloroplast genomes (Liu & Xue, 2021). Species rich in G and C bases tend to have the best codons rich in G and C, while species rich in A and U bases prefer the best codons rich in A and U (Nie *et al*., 2014). Our results support the previous studies indicating the codon usage patterns in chloroplast genomes of *E. altotibetica* is same as six *Euphorbiaceae* species (Wang *et al*., 2020). Therefore, we considered that such high codon usage bias might be due to higher AU content in the *E. altotibetica* chloroplast genomes.

During the course of evolution, chloroplast genomes had expansions and contractions in the boundary regions of LSC/IRb, IRb/SSC, SSC/IRa and IRa/LSC (Kim & Lee, 2004; Hansen *et al*., 2007; Xue *et al*., 2019), which led to the change in length and structure of chloroplast genomes, and even pseudogenization (Huang *et al*., 2014). Here, we found that the ten chloroplast genomes of *Euphorbia* were highly similar. IRb/SSC boundaries of *E. altoibetica*, *E. kansuensis* and *E. helioscopia* were located in the overlay region of *ycf1* and *ndhF* genes, which had different degrees of expansions and contractions. Meanwhile, *E. fauriei* genes do not overlap at the IRb/SSC boundary and existed *ycf1* gene loss. *Rps19* gene lost at the IRa/LSC region in *E. altotibetica*, is similar as in *E. kansuensis*, except that *E. lathyris* lost *ndhF* gene, while others contained *ndhF* gene in the IRb/SSC boundary. However, we found *rps19* gene shrinkage at the LSC/IRb and IRa/LSC boundaries of *E. kansui*, *E. esula* and *E. lahyris*, but expansion into the IR regions of other *Euphorbia* species. Therefore, we inferred that IR regions of all *Euphorbia* is highly conserved, and variations in LSC/IRb and IRa/LSC boundaries was the main cause of expansions and contractions in the IR regions.

Ka/Ks ratio could represent selective pressure in special protein-coding genes. *Ka/Ks* value greater than 1 indicates the genes were under the influence of positive selection, and lower than 1 is considered that they are experiencing negative selection or purifying selection, while *Ka/Ks* value equal to zero indicates they are under neutral selection (Wang *et al*., 2010). Comparative analysis of *Ka/Ks* ratio of *E. altoibetica* with ten other species of *Euphorbia*, we found only four genes (*ycf1*, *ycf2*, *rpl20* and *rpl22*) have Ka/Ks value greater than 1. Y*cf1* and *ycf2* genes were indispensable for chloroplasts (Esposito *et al*., 2001). Recent studies have shown that *ycf1* gene encoded Tic214 is an essential component of Tic (translocon of the inner membrane of chloroplasts) in *Arabidopsis* (Kikuchi *et al*., 2013) and Wolfe presumed *ycf2* gene could encode ATP synthase (Wolfe, 1994). *Rpl20* and *rpl22* genes separately encoded the ribosomal proteins of L20 and L22, whose adaptive evolution involved the process of proteins synthesis in the chloroplast of angiosperms (Wang *et al*., 2012). Therefore, positive selection of these genes could help *E. altoibetica* to maintain normal growth and metabolism in arid environments, thus improving plant stress adaptation.

We reconstructed the phylogenetic relationship based on the chloroplast genome of 68 *Euphorbia* species. The result showed that all *Euphorbia* species constituted a monophyletic group with a high bootstrap support, and they are classified into two major clades (Clade 1 and Clade 2). All species from clade 1 belonged to *E.* Subgen. *Esula* Pers, *E. altotibetica,* and *E. peplus* had the closest relationship. Meanwhile, *E. jolkinii*, *E. pekinensis*, *E. micractina,* and *E. fauriei* are classified into a subclade. Previous studies considered that the former three species of *Euphorbia* belonged to the Sect. *Tulocarpa* (Raf.) Prokh (Li *et al*., 2008), thus we inferred *E. fauriei* should integrate into Sect. *Tulocarpa* (Raf.) Prokh. All of species in clade 2 had clear definition with high support values, but their sect. tribe, and subgenus nomenclature are not defined yet. Therefore, our study provided an important reference to identify the phylogenetic relationship of *Euphorbia* species.

Conclusion

In the present study, we sequenced and described the chloroplast genome of *E. altotibetica* for the first time. We found that the chloroplast genome of *E. altotibetica* is similar to other *Euphorbia* species with little variations. Our phylogenetic analyses provided strong support for the relationship among *Euphorbia* species and clarified its phylogenetic position.

Acknowledgements

This research was funded by the Second Tibetan Plateau Scientific Expedition and Research (STEP) Program (2019QZKK0502), the 2023 First Fund for Central Government to Guide Local Science and Technology Development in Qinghai Province (2023ZY019), and the National Natural Science Foundation of China (32160297, 31960052).

References:

Amiryousefi, A., J. Hyvönen and P. Poczai. 2018. IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinform.,* 34: 3030-3031.

Asaf, S., M. Waqas, A.L. Khan, M.A. Khan, S.M. Kang, Q.M. Imran, R. Shahzad, S. Bilal, B.W. Yun and I.J. Lee. 2017. The complete chloroplast genome of Wild Rice (*Oryza minuta*) and Its comparison to related species. *Front. Plant Sci*., 8: 304.

Beier, S., T. Thiel, T. Münch, U. Scholz and M. Mascher. 2017. MISA-web: a web server for microsatellite prediction. *Bioinform.,* 33: 2583-2585.

Bendich, A.J. 2004. Circular chloroplast chromosomes: the grand illusion. *Plant Cell*, 16: 1661-1666.

Campos, J.L., K. Zeng, D.J. Parker, B. Charlesworth and P.R. Haddrill. 2013. Codon usage bias and effective population sizes on the X chromosome versus the autosomes in *Drosophila melanogaster*. *Mol. Biol. Evol*., 30: 811-823.

Chen, S., H. Yao, J.P. Han, L. Chang, J.Y. Song, L.C. Shi, Y.J. Zhu, X.Y. Ma, T. Gao, X.H. Pang, K. Luo, Y. Li, X.W. Li, X.C. Jia, Y.L. Lin and C. Leon. 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS One*, 5: e8613.

Corriveau, J.L. and A.W. Coleman. 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *Amer. J. Bot*., 75: 1443-1458.

Daniell, H., C.S. Lin, M. Yu and W.J. Chang. 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Gen. Biol*., 17: 134.

Deguilloux, M.F., M.H. Pemonge and R.J. Petit. 2004. Use of chloroplast microsatellites to differentiate oak populations. *Ann. Forest. Sci*., 61: 825-830.

Dierckxsens, N., P. Mardulyn and G Smits. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucl. Acids Res*., 45: e18.

Ding, B.C., Y. Sun, F.X. Rong, C. Guo, Y.Q. Wang, Y. Zhang, P.P. Gao and J. Ding. 2020. The complete mitochondrial genome of *Holothuria spinifera* (Théel, 1866). *Mitochondrial DNA Part B*, 5: 1679-1680.

Dong, W., C. Xu, D. Li, X.B. Jin, R. Li, Q. Lu and Z. Suo. 2016. Comparative analysis of the complete chloroplast genome sequences in psammophytic *Haloxylon* species (Amaranthaceae). *Peer J*., 4: e2699.

Du, Q.Z., B.W. Wang, Z.Z. Wei, D.Q. Zhang and B. Li. 2012. Genetic diversity and population structure of Chinese white poplar (*Populus tomentosa*) revealed by SSR markers. *J. Hered*., 103: 853-862.

Du, X.Y., T. Zeng, Q. Feng, L.J. Hu, X. Luo, Q.B. Weng, J.F. He and B. Zhu. 2020. The complete chloroplast genome sequence of yellow mustard (*Sinapis alba* L.) and its phylogenetic relationship to other Brassicaceae species. *Gene*, 731: 144340.

Esposito, D., D.C. Higgs, R.G. Drager, D.B. Stern and J. Girard-Bascou. 2001. A nucleus-encoded suppressor defines a new factor which can promote *petD* mRNA stability in the chloroplast of *Chlamydomonas reinhardtii*. *Curr. Genet*., 39: 40-48.

Frazer, K.A., L. Pachter, A. Poliakov, E.M. Rubin and I. Dubchak. 2004. VISTA: computational tools for comparative genomics. *Nucl. Acids Res*., 32: W273-W279.

Gray, J.C. 1989. Genetic manipulation of the chloroplast genome. *Plant Biotechnol.*, 12: 317-335.

Guindon, S., J.F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk and O. Gascuel. 2010. New algorithms and methods to estimate maximum-Likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.*, 59: 307-321.

Hansen, D.R., S.G. Dastidar, Z.Q. Cai, C. Penaflor, J.V. Kuehl, J.L. Boore and R.K. Jansen. 2007. Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). *Mol. Phylogen. Evol.*, 45: 547-563.

Huang, H., C. Shi, Y. Liu, S.Y. Mao and L.Z. Gao. 2014. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evol. Biol.*, 14: 151.

Jansen, R.K., L.A. Raubeson, J.L. Boore, C.W. dePamphilis, T.W. Chumley, R.C. Haberle, S.K. Wyman, A.J. Alverson, R. Peery, S.J. Herman, H.M. Fourcade, J.V. Kuehl, J.R. McNeal, J.L. Mack and L.Y. Cui. 2005. Methods for obtaining and analyzing whole chloroplast genome sequences. *Meth. Enzymol.*, 395: 348-384.

Jiang, W.J., M.Y. Guo and X.H. Pang. 2020. Application of chloroplast genome in identification and phylogenetic analysis of medicinal plants. *World Chin. Med*., 15: 702-708+716.

Jiang, Y., Q.S. Yang, J. Wang, C.J. Yang, L.F. Huang, Y. Yang and B.Z. Duan. 2021. Complete chloroplast genome of *Paris mairei*: characterization and phylogeny. *Chin. Trad. Herb. Drugs*, 52: 4014-4022.

Katoh, K., K. Misawa, K. Kuma and T. Miyata. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.*, 30: 3059-3066.

Kikuchi, S., J. Bedard, M. Hirano, Y. Hirabayashi, M. Oishi, M. Imai, M. Takase, T. Ide and M. Nakai. 2013. Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science*, 339: 571-574.

Kim, K.J. and H.L. Lee. 2004. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng Nees*) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res.*, 11: 247-261.

Kong, B.L.H., H.S. Park, T.W. Lau, Z. Lin, T.J. Yang and P.C. Shaw. 2021. Comparative analysis and phylogenetic investigation of Hong Kong Ilex chloroplast genomes. *Sci. Rep.*, 11: 5153.

Kuang, D.Y., H. Wu, Y.L. Wang and L.M. Gao. 2011. Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics. *Genome*, 54: 663-673.

Lemieux, C., C. Otis and M. Turmel. 2016. Comparative chloroplast genome analyses of Streptophyte green algae uncover major structural alterations in the Klebsormidiophyceae, Coleochaetophyceae and Zygnematophyceae. *Front. Plant Sci.*, 7: 697.

Li, B., F. Lin, P. Huang, W. Guo and Y. Zheng. 2017. Complete chloroplast genome sequence of *Decaisnea insignis*: Genome organization, genomic resources and comparative analysis. *Sci. Rep.*, 7: 10073.

Li, B.T., H.X. Qiu, J.S. Ma, H. Zhu, G.G. Michael, E. Hans-Joachim, D. Stefan, H. Petra, J.G. Lynn, V. Maria and D.M. Gordon. 2008. Euphorbiaceae. pp. 163-314. In: (Eds.): Wu, Z.Y. & P.H. Raven, *Flora of China*, Vol. 11. Science Press, Beijing, and Missouri Botanical Garden Press, St. Louis. Available online at: http://www.efloras.org/florataxon.aspx? flora_id=2&taxon_id= 242321386 (accessed on 26 April 2023).

Li, P., X.Q. Liang and X.N. Zhang. 2019. Characterization of the complete chloroplast genome of *Euphorbia helioscopia* Linn. (Euphorbiaceae), a traditional Chinese medicine. *Mitochondrial DNA Part B* 4.

Liu, Q. and Q.Z. Xue. 2021. Comparative studies on codon usage pattern of chloroplasts and their host nuclear genes in four plant species. *J. Genet*., 84: 55-62.

Liu, S.S., Y.H. Hu, F. Maghuly, I.M. Porth and J.F. Mao. 2018. The complete chloroplast genome sequence annotation for *Malania oleifera*, a critically endangered and important bioresource tree. *Conserv. Gen. Resour.*, 11: 271-274.

Lohse, M., O. Drechsel, S. Kahlau and R. Rock. 2013. Organellar Genome DRAW, A suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucl. Acids Res.*, 41: W575-W581.

Ma, W.Q., C.G. Lü, D.Q. Jiang, C.Z. Kang and D. Zhao. 2020. The complete chloroplast genome sequence of *Euphorbia lathyris* L. (Euphorbiaceae). *Mitochondrial DNA B Resour.*, 5: 3678-3680.

Nguyen, L.T., H.A. Schmidt, A. von Haeseler and B.Q. Minh. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol*., 32: 268-274.

Nie, X.J., P.C. Deng, K.W. Feng, P.X. Liu, X.H. Du, M.Y. Frank and W.N. Song. 2014. Comparative analysis of codon usage patterns in chloroplast genomes of the Asteraceae family. *Plant Mol. Biol. Rep*., 32: 828-840.

Palmer, J.D. 1985. Comparative organization of chloroplast genomes. *Ann. Rev. Genet*., 19: 325-354.

Pan, L., X.F. Zhang, Y. Deng, W. Huan, W.D. Gang and L.X. Dong. 2003. Chemical constituents from the whole plant of Euphorbia altotibetic. *Helv. Chim. Acta*., 86: 2525-2532.

Parks, M., R. Cronn and A. Liston. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.*, 7: 84.

Peng, J.Y., X.S. Zhang, D.G. Zhang, Y. Wang, T. Deng, X.H. Huang, T.H. Kuang and Q. Zhou. 2022. Newly reported chloroplast genome of *Sinosenecio albonervius* Y. Liu & Q. E. Yang and comparative analyses with other *Sinosenecio* species. *BMC Genom.*, 23: 639.

Powell, W., M. Morgante, R. Mcdevitt, G.G. Vendramin and J.A. Rafalski. 1995. Polymorphic simple sequence repeat regions in chloroplast genomes: Applications to the population genetics of pines. *Proc. Natl. Acad. Sci. U.S.A.*, 92: 7759-7763.

Pugh, T., O. Fouet, A. M. Risterucci, P. Brottier, M. Abouladze, C. Deletrez, B. Courtois, D. Clement, P. Larmande, J.A.K. N'Goran and C. Lanaud. 2004. A new cacao linkage map based on codominant markers: development and integration of 201 new microsatellite markers. *Theor. Appl. Genet.*, 108: 1151-1161.

Qu, X.J., M. Moore, D.Z. Li and T.S. Yi. 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods*, 15: 1-12.

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

Redwan, R.M., A. Saidin and S.V. Kumar. 2015. Complete chloroplast genome sequence of MD-2 pineapple and its comparative analysis among nine other plants from the subclass Commelinidae. *BMC Plant Biol.*, 15: 196.

Sharp, P.M. and W.H. Li. 1987. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucl. Acids Res.*, 15: 1281-1295.

Shi, Q.W., X.H. Su and H. Kiyota. 2008. Chemical and pharmacological research of the plants in genus *Euphorbia*. *Chem. Rev.*, 108: 4295-4327.

Shields, D.C. and P.M. Sharp. 1987. Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational biases. *Nucl. Acids Res.*, 15: 8023-8040.

Song, J., Y.H. Long, L.M. Lin and F. Yin. 2017. Analysis on structure and phylogeny of chloroplast genomes in *Araliaceae* species. *Chin. Trad. Herb. Drugs*, 48: 5070-5075.

Staden, R. and A.D. McLachlan. 1982. Codon preference and its use in identifying protein coding regions in long DNA sequences. *Nucl. Acids Res.*, 10: 141-156.

Straub, S.C.K., M. Parks, K. Weitemier, M. Fishbein, R.C.Cronn and A. Liston. 2012. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *Amer. J. Bot.*, 99: 349-364.

Tai, T.H. and S.D. Tanksley. 1990. A rapid and inexpensive method for isolation of total DNA from dehydrated plant tissue. *Plant Mol. Biol. Rep.*, 8: 297-303.

Tillich, M., P. Lehwark, T. Pellizzer, E.S. Ulbricht-Jones, A. Fischer, R. Bock and S. Greiner. 2017. GeSeq-versatile and accurate annotation of organelle genomes. *Nucl. Acids Res.*, 45: 6-11.

Twyford, A.D. and R.W. Ness. 2017. Strategies for complete plastid genome sequencing. *Mol. Ecol. Resour.*, 17: 858-868.

Wang, B., L. Gao, Y. Su and T. Wang. 2012. Adaptive evolutionary analysis of chloroplast genes in euphyllophytes based on complete chloroplast genome sequences. *Acta Scientiarum Naturalium Universitatis Sunyatseni*, 51: 108-113+146.

Wang, D.P., Y.B. Zhang, Z. Zhang, J. Zhu and J. Yu. 2010. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genom. Proteom. Bioinf.*, 8: 77-80.

Wang, Y.L., X. Jian and S. Wang. 2022. Characterization of the complete chloroplast genome of *Euphorbia pekinensis* Rupr. (Euphorbiaceae). *Mitochondrial DNA Part B*, 7: 1550-1552.

Wang, Z.J., B.B. Xu, B. Li, Q.Q. Zhou, G.Y. Wang, X.Z. Jiang, C.C. Wang and Z.D. Xu. 2020. Comparative analysis of codon usage patterns in chloroplast genomes of six Euphorbiaceae species. *Peer. J.*, 8: e8251.

Wolfe, K.H. 1994. Similarity between putative ATP-binding sites in land plant plastid ORF2280 proteins and the FtsH/CDC48 family of ATPases. *Curr. Genet.*, 25: 379-383.

Xiong, A.S., R.H. Peng, J. Zhuang, F. Gao, B. Zhu, X.Y. Fu, Y. Xue, X.F. Jin, Y.S. Tian, W. Zhao and Q.H. Yao. 2009. Gene duplication, transfer, and evolution in the chloroplast genome. *Biotechnol. Adv.*, 27: 340-347.

Xu, W.B., B.S. Xia and X.W. Li. 2020. The complete chloroplast genome sequences of five pinnate-leaved Primula species and phylogenetic analyses. *Sci. Rep*., 10: 20782.

Xue, S., T. Shi, W.J. Luo, X.P. Ni, S. Iqbal, Z.J. Ni, X. Huang, D. Yao, Z.J. Shen and Z.H. Gao. 2019. Comparative analysis of the complete chloroplast genome among *Prunus mume*, *P. armeniaca*, and *P. salicina*. *Hort. Res.*, 6: 89.

Yang, Y., T. Zhou, D. Duan, J. Yang, L. Feng and G. Zhao. 2016. Comparative analysis of the complete chloroplast genomes of five *Quercus* species. *Front. Plant Sci.*, 7: 959.

Yu, T., F. Pu, Q. Guan and M. Fan. 2022. Chloroplast genome of *Euphorbia peplus* L. characteristic and phylogenetic analysis. *Mol. Plant Breed.*, 20: 1828-1837.

Yu, X.Q., B. Drew, J.B. Yang, L.M. Gao and D.Z. Li. 2017. Comparative chloroplast genomes of eleven *Schima* (Theaceae) species: Insights into DNA barcoding and phylogeny. *PLoS One*, 12: e0178026.

Yuan, J.B., X.J. Zhang, M. Wang, Y.M. Sun, C.Z. Liu, S.H. Li, Y. Yu, Y. Gao, F. Liu, X.X. Zhang, J. Kong, J.Y. Fan, C.S. Zhang, L. Feng, J.H. Xiang and F.H. Li. 2021. Simple sequence repeats drive genome plasticity and promote adaptive evolution in penaeid shrimp. *Comm. Biol.*, 4: 186.

Zhang, B.B., X.L. Han, Q. Jiang, Z.X. Liao and H.S. Wang. 2013. Cytotoxic cholestane-type and ergostane-type steroids from the aerial parts of Euphorbia altotibetic. *Steroids*, 78: 38-43.

Zhang, T.W., Y.J. Fang, X.M. Wang, X. Deng, X.W. Zhang, S.N. Hu and J. Yu. 2012. The complete chloroplast and mitochondrial genome sequences of *Boea hygrometrica*: insights into the evolution of plant organellar genomes. *PLoS One*, 7: e30531.

Zhang, Y.C., Z.C. Lu, D. Zhang and J.J. Li. 2021. The complete chloroplast genome of *Euphorbia hirta* (Euphorbiaceae), a commonly used medicinal plant in China. *Mitochondrial DNA Part B*, 6.

Zhao, X.Y., Q. Wang, F.Y. Wang and Y. Li. 2020. Characterization of the complete chloroplast genome of *Euphorbia lathyris* (Euphorbiaceae), a traditional Chinese medicine. *Mitochondrial DNA Part B*, 5.

Zhao, Z.Y., X. Wang, Y. Yu, S.H. Yuan, D. Jiang, Y.J. Zhang, T. Zhang, W.H. Zhong, Q.J. Yuan and L.Q. Huang. 2018. Complete chloroplast genome sequences of *Dioscorea*: Characterization, genomic resources, and phylogenetic analyses. *Peer. J.*, 6: e6032.

Zhou, J.G., X.L. Chen, Y.X. Cui and W. Sun. 2017. Molecular structure and phylogenetic analyses of complete chloroplast genomes of two *Aristolochia* medicinal species. *Int. J. Mol. Sci.*, 18: 1839.

(Received for publication 16 August 2023)