Frontiers in Bioscience Landmark

Front. Biosci. (Landmark Ed) **2023**; 28(6): 110
https://doi.org/10.31083/j.fbl2806110

Original Research

Chloroplast Genome Structure and Phylogenetic Analysis of 13 Lamiaceae Plants in Tibet

Yiming Niu[1], Qian Qin[1], Yanjing Dong[1], Xiaoyun Wang[1], Shouwen Zhang[1,*], Zejing Mu[1,*]

[1]The Research Center for Traditional Chinese Medicine Resources and Ethnic Minority Medicine, Jiangxi University of Chinese Medicine, 330004 Nanchang, Jiangxi, China

*Correspondence: wtzsw@163.com (Shouwen Zhang); 20131060@jxutcm.edu.cn (Zejing Mu)

Academic Editor: Jen-Tsung Chen

Submitted: 7 December 2022 Revised: 12 January 2023 Accepted: 1 February 2023 Published: 8 June 2023

Abstract

Background: The chloroplast (cp) genome has unique and highly conserved characteristics and is therefore widely used in species identification and classification, as well as to improve the in–depth understanding of plant evolution. **Methods**: In this study, the cp genomes of 13 Lamiaceae plants in the Tibet Autonomous Region of China were sequenced, assembled and annotated using bioinformatics methods. Phylogenetic trees were constructed to reveal the phylogenetic relationship of related species in the Lamiaceae. **Results**: The results showed that all 13 cp genomes had a typical four–segment structure, including one large single–copy (LSC) region, one pair of inverted repeat (IR) regions and one small single–copy (SSC) region. The sequence lengths of the 13 cp genomes were between 149,081 bp and 152,312 bp, and the average GC content was 37.6%. These genomes contained 131–133 annotated genes, including 86–88 protein–coding genes, 37–38 tRNA genes, and 8 rRNA genes. A total of 542 SSR loci were detected using MISA software. The repeat types were mostly single–nucleotide repeats, accounting for 61% of simple repeats. A total of 26,328–26,887 codons were detected in 13 cp genomes. According to the RSCU value analysis, the codons mostly ended with A/T. Analysis of IR boundaries showed that the other species were relatively conserved, except for *Nepeta laevigata* (D. Don) Hand.–Mazz., which differed in gene type and location on both sides of the boundary. By analysing nucleotide diversity, two highly mutated regions located in the LSC and SSC regions were identified in the 13 cp genomes. **Conclusions**: Using the cp genome of *Lycium ruthenicum* Murray as the outgroup, 97 cp genomes of the Lamiaceae were used to construct an Maximum Likehood (ML) phylogenetic tree, in which these species were divided into eight major clades, corresponding to eight subfamilies based on morphological classification. The phylogenetic results based on monophyletic relationships were consistent with the morphological classification status at the tribe level.

Keywords: Lamiaceae; chloroplast genome; phylogeny

1. Introduction

Lamiaceae is a large family with worldwide distribution. There are 10 subfamilies, approximately 220 genera and 3500 species in this family in the world, and more than 800 species in 99 genera in China [1]. Compared to mainland China, in the Tibetan plateau region, the sunlight exposure time is longer, and the altitude is higher. Because of this special geographical environment, Tibet has bred some unique medicinal plants, commonly known as Tibetan medicine [2,3]. In China, 24 species in 16 genera of the Lamiaceae are used in Tibetan medicine [4], e.g., *Elsholtzia fruticosa*, *N. laevigata*, and *Elsholtzia densa*. These Tibetan medicines are widely used in the treatment of influenza, liver and stomach diseases, dysentery, and pharyngitis [3]. Some Lamiaceae plants have very good economic value, e.g., *Galeopsis bifida* is a famous oil crop, and *Salvia sikkimensis* is a landscape plant.

To date, studies on species in the Lamiaceae have mainly focused on pharmacological effects [5], morphological investigations [6], chemical composition analysis [7,8], and species classification [9]. However, Chinese scholars have published the cp genomes of Tibetan medicine *Elsholtzia* Willd. and *Lamiophlomis* Kud., and confirmed that plant morphological classification can be used as a basis to support molecular classification and species identification of medicinal materials [10,11]. Through clustering analysis of cp genomes of some species in the Lamiaceae, studies have confirmed that medicinal materials with similar chemical compositions and pharmacological effects have close genetic relationships [12]. In order to explore the relationships of different subfamilies of the Lamiaceae, investigators have established the phylogenetic trees with the cp genomes of some species in the Lamiaceae. They found that the Nepetoideae subfamily formed a clade independently of the other six subfamilies, and that the Lamioideae subfamily and Scutellarioideae subfamily were closely related [13,14].

Cp is the main intracellular location for plant photosynthesis and one of the organelles with independent genetic material. Compared with the mitochondrial or nuclear genome, the cp genome has a higher structure, gene number and gene composition [15]. In recent years, an increasing number of researchers have used cp genomes to analyse plant genetic diversity [16–18] and to explore the relationships among different families, genera and species [19–26].

Table 1. Collection information for 13 species of the Lamiaceae.

Special	Collect location	Latitude and longitude	Altitude	Sample number	Accession number
Ajuga nubigena Diels.	Gyirong County, Rikaze City, Tibet	28°23′53.9″E 85°18′35″N	2868.1m	H3050404	OP186457
Elsholtzia densa Benth.	Chagyab County, Changdu City, Tibet	98°6′0.51″E 30°12′31.98″N	4365.0m	H3050021	OP186458
Elsholtzia eriostachya Benth.	Nyalam County, Rikaze City, Tibet	28°19’6.2″E 86°2’19.5″N	4172.6m	H3050264	OP186459
Elsholtzia fruticosa (D. Don) Rehd.	Gyirong County, Rikaze City, Tibet	28°23’52.4″E 85’19’39.1″N	2813.1m	H3050369	OP186460
Galeopsis bifida Boenn.	Bomê County, Nyingchi City, Tibet	93°14′52.0″E 29°53′5.54″N	3630.0m	H3050082	OP186461
Marmoritis complanatum (Dunn) A. L. Budantzev.	Qushui County, Lhasa City, Tibet	90°38′9.46″E 29°18′33.49″N	3596.0m	H3050160	OP186462
Nepeta dentata C. Y. Wu et Hsuan.	Karuo District, Changdu City, Tibet	31°24′36.8″E 97°28′1.8″N	3677.4m	H3050533	OP186463
Nepeta hemsleyana Oliver ex Prain.	Bomê County, Nyingchi City, Tibet	93°14′52.0″E 29°53′5.54″N	3630.0m	H3050095	OP186464
Nepeta laevigata (D. Don) Hand.–Mazz.	Mira Mountain Pass, Nyingchi City, Tibet	94°38’54.14”E 29°36’55.22”N	4544.0m	H3050131	OP186465
Nepeta thomsonii Benth. ex Hook. f.	Chagyab County, Changdu City, Tibet	98°6′0.51″E 30°12′31.98″N	4365.0m	H3050023	OP186466
Phlomis betonicoides (Diels) Kamelin & Makhm.	Ranwu Canyon, Basu County, Changdu District, Tibet	96°46′4.64″E 29°30′55.02″N	4013.0m	H3050063	OP186467
Salvia sikkimensis Stib.	Yadong County, Rikaze City, Tibet	89°0′1.34″E 27°33′20.36″N	3693.0m	H3050217	OP186468
Thymus linearis Benth.	Nyalam County, Rikaze City, Tibet	28°19′6.2″E 86°2′19.5″N	4172.6m	H3050268	OP186469

Many species in the *Nepeta*, *Elsholtzia*, and *Salvia* genera of the Lamiaceae [1] are distributed in the Qinghai–Tibet Plateau in China, providing a solid basis for the study of these species. Therefore, based on high–throughput sequencing technology, we sequenced the cp genomes of 13 Lamiaceae species from the Tibet Autonomous Region of China, analysed and compared their structural characteristics using bioinformatics methods, and constructed phylogenetic trees, aiming to provide new ideas for the classification and identification of the Lamiaceae species.

2. Material

2.1 Plant Materials

The plant materials of 13 species of the Lamiaceae used in this study were collected from the Tibet Autonomous Region of China (Tibet) (Table 1, Fig. 1) and identified by Professor Guoyue Zhong of Jiangxi University of Chinese Medicine (JUTCM). The certificate specimens were all stored in the herbarium of JUTCM.

2.2 DNA Extraction, Quality Examination and Sequencing

Leaves dried with silica gel were used to extract DNA using a Plant Genomic DNA Kit (Tiangen Biochemical Technology (Beijing) Co., Ltd.). The quality of genomic DNA was detected by a spectrophotometer and 1% agarose gel electrophoresis. According to the standard genomic DNA library preparation procedure provided by Illumina, intact DNA samples with concentrations greater than 20 ng·μL^{-1} were lysed by ultrasonic treatment, and fragmented DNA was purified and end–repaired. Libraries with an insertion size of 350 bp were prepared and then sequenced using the Illumina NovaSeq 6000 high–throughput sequencing platform (Illumina, San Diego, CA, USA). Sequencing was performed by Novgene Biotech Co., Ltd.(Nanjing, Jiangsu Province, China).

2.3 Assembly and Annotation of cp Genome

Trimmomatic v0.36 software [27] (Aachen and Institute of Bio- and Geosciences: Plant Sciences, Forschungszentrum Jülich, Leo-Brandt-Straße, Jülich, Germany) was used for quality filtering of raw data in Illumina sequencing. Single bases with a quality score lower than 20 were deleted from both ends of the sequence, as well as regions of sequences with more than three consecutive uncalled bases. All reads less than 40 bp were discarded. Filtered data were mapped to available cp genomes of the closest species in the Lamiaceae using Bowtie2 v.2.2.3 software [28] (Department of Computer ScienceUniversity College

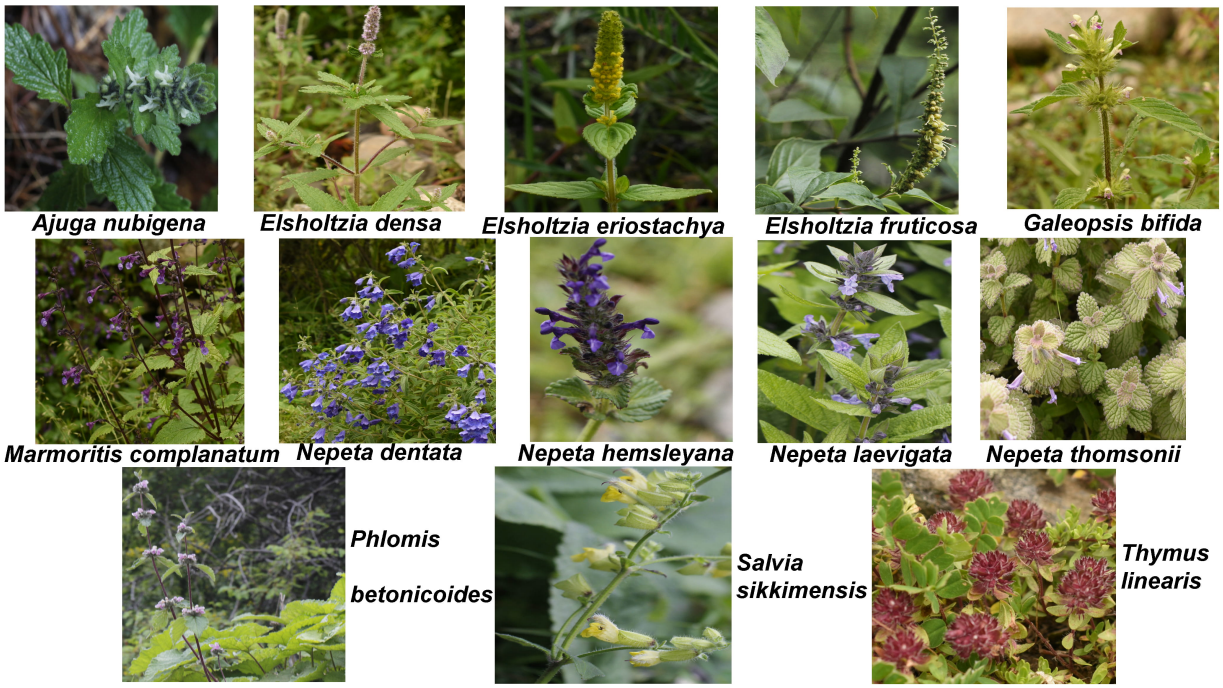

Fig. 1. The collected 13 Lamiaceae species.

London, Gower Street, London, UK), and reads from nuclear and mitochondrial origins were excluded. The chloroplast genomes were then assembled and reconstructed using GetOeller 1.7.5 software [29] (Chinese Academy of Sciences, Kunming, Yunnan, China). CPGAVAS2 software [30] (Faculty of Technology, University of Bielefeld, Germany) (http://47.96.249.172:16019/analyzer/annotate/) was used to complete automatic annotation. Geneious 11.0.5 software [31] (Oxford, England) was used for manual correction, and OGDRAW V1.1 software [32] (Max-Planck-Institut für Molekulare Pflanzenphysiologie, Am Mühlenberg, Germany) (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) was used to map the physical structure of the chloroplast genome.

2.4 SSR Sequence Analysis

MISA software [33] was used to screen microsatellite sequences, and the parameters were set as ten repeats for single–nucleotide SSRs, five repeats for dinucleotide single sequence repeats (SSRs), four repeats for trinucleotide SSRs, and three repeats for four– and five–nucleotide SSRs.

Forward, reverse, complementary and palindrome repeat sequences were identified using REPuter software [34] (Faculty of Technology, Bielefeld, Germany) (http://bibiserv.techfak.uni-bielefeld.de/reputer/).

2.5 Codon Preference and Selection Pressure Analysis

CodonW software 1.4.2 [35] (Paul Sharp lab, Dept of Genetics, University of Nottingham, UK) was used to identify codon usage patterns and calculate codon bias (RSCU). The CODEML program in PAML V4.973 software [36] (Department of Biology, Galton Laboratory, London, UK) was used to calculate ratio of nonsynonymous nucleotide substitution rate to synonymous nucleotide substitution rate (Ka/Ks) for each gene. Pairwise comparisons were performed, dN/dS ratio of each protein–coding gene were calculated using the yn00 program in PAML and the detailed parameters were: icode = 10, weighting = 0, commonf3x4 = 0, and other parameters in the CODEML control file were left at default settings.

2.6 Nucleotide Diversity Analysis

Nucleotide variability (Pi) of cp genomes was assessed by alignment of sequences using MAFFT V7 software [37] (Bioinformatics Center, Institute for Chemical Research, Kyoto University Uji, Kyoto 611-0011, Japan), followed by manual adjustment of sequences using BioEdit software [38] (Borland company, Scotts Valley, California, USA). Sliding window analysis was completed using DnaSP version 5.1 software [39] (Departament de Genètica, Facultat de Biologia and Institut de Recerca de la Biodiversitat, Universitat de Barcelona, Barcelona, Spain). The step size was set to 200 bp, and the window length was set to 600 bp.

2.7 IR Boundary Analysis

Using IRSCOPE software [40] (https://irscope.shinyapps.io/irapp/), the IR (Inverted repeat) boundaries were drawn, and the gene type and location in the border area were identified. The expansion and contraction of genes were also discovered.

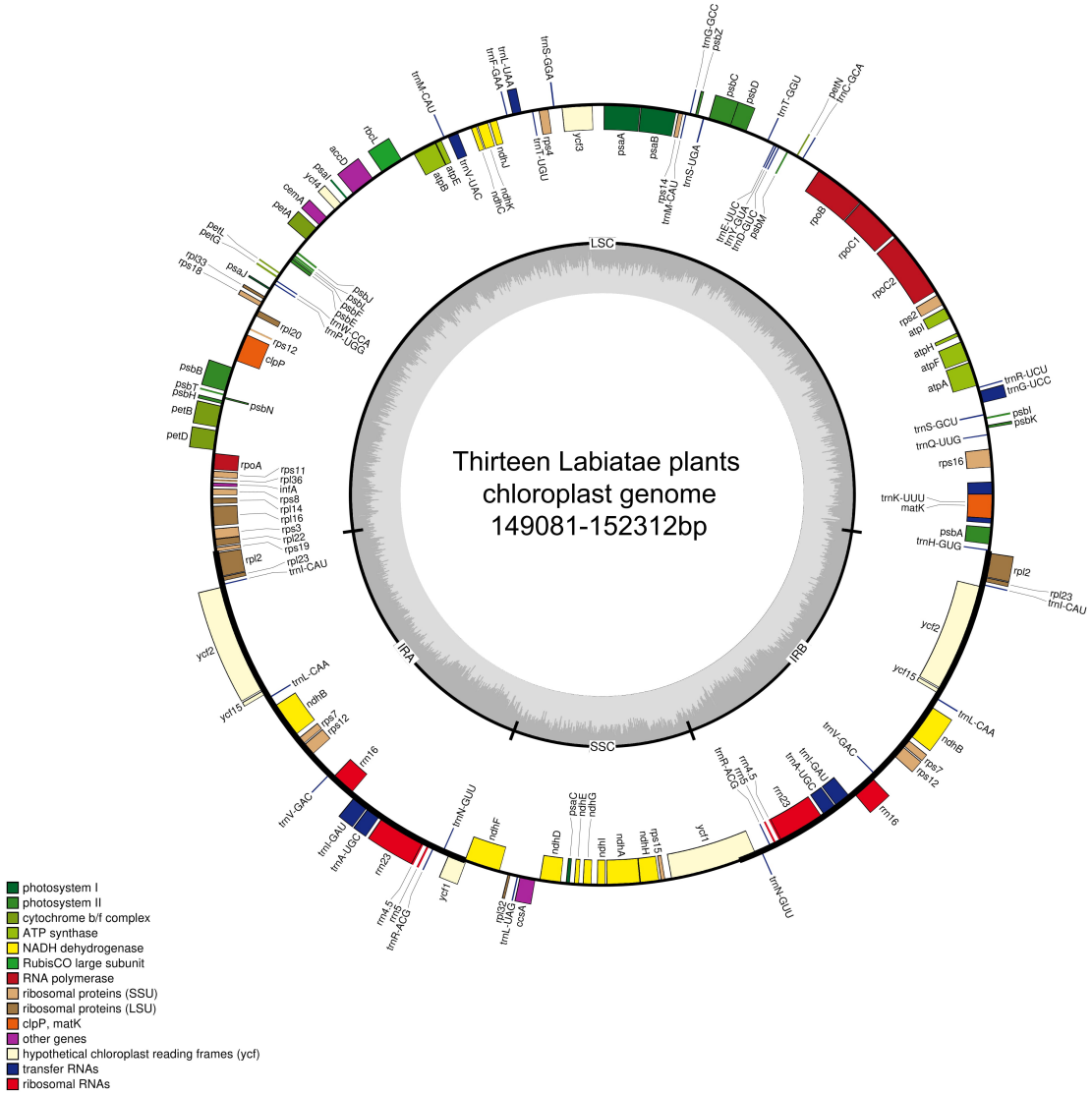

Fig. 2. The cp genome map of 13 Lamiaceae species. Note: The black thick line indicates that the large single copy area (LSC) and small single copy area (SSC) were separated by two reverse repeat areas (IRa and IRb). The dark inner circle indicates the GC content in the chloroplast genome. Genes drawn inside the circle are transcribed clockwise, and those drawn outside the circle are transcribed counter clockwise.

2.8 Sequence Variation Analysis

The differences in cp genome sequences of the 13 Lamiaceae species evaluated in this study were compared using mVISTA software [41] (National Energy Research Scientific Computing Center Genome Sciences Department, Berkeley, USA) (http://genome.lbl.gov/vista/mvista/submit.shtml).

2.9 Phylogenetic Relationship Analysis

Selecting the published cp genomes of 84 Lamiaceae species and the 13 cp genomes from the Lamiaceae species in Tibet, the MAFFT V7 program [37] was used to compare the entire cp genomes, with manual adjustment when necessary. After being estimated in Modelfinder, the general time reversible (GTR) + r model of nucleotide substitution was used for phylogeny. Using the cp genome of *L. ruthenicum* (NCBI accession number NC039651.1) as the outgroup, a maximum likelihood (ML) phylogenetic tree was constructed using PhyloSuite v1.2.2 [42] (Key Laboratory of Aquaculture Disease Control, Chinese Academy of Sciences, Wuhan, China) with bootstrap values of 10,000 replicates.

3. Results and Analysis

3.1 Basic Characteristics of cp Genomes

Consistent with most land plants, the cp genomes of the 13 Lamiaceae species showed a conserved quadripartite structure [43] (Fig. 2), with lengths ranging from 149,081 bp to 152,312 bp, indicating that the lengths of these cp genomes were conserved [23]. The LSC region (80,908–

Table 2. Cp genome characteristics of 13 Lamiaceae species.

Species	Size	Length (LSC)	Length (SSC)	Length (IR)	GC%	PCGs	tRNAs	rRNAs	Genes
A. nubigena Diels	150,490	82,153	17,177	25,580	38.27	88	37	8	133
E. densa Benth.	149,081	81,483	17,364	25,117	37.92	86	37	8	131
E. eriostachya Benth.	148,288	80,908	17,468	24,956	38.18	88	37	8	133
E. fruticosa (D. Don) Rehd.	151,583	82,812	17,495	25,638	37.96	88	37	8	133
G. bifida Boenn.	152,210	83,302	17,572	25,668	37.88	88	37	8	133
M.s complanatum (Dunn) A. L. Budantzev.	151,834	82,917	17,671	25,623	38.47	88	37	8	133
N. dentata C. Y. Wu et Hsuan.	152,312	83,416	17,692	25,602	37.85	88	37	8	133
N. hemsleyana Oliver ex Prain.	151,893	83,258	17,407	25,614	37.95	88	37	8	133
N. laevigata (D. Don) Hand.–Mazz.	152,177	83,214	17,611	25,676	37.86	88	37	8	133
N. thomsonii Benth. ex Hook. f.	151,985	82,790	17,615	25,790	37.83	87	38	8	133
P. betonicoides (Diels) Kamelin & Makhm.	151,794	83,205	17,389	25,600	38.49	88	37	8	133
S. sikkimensis Stib.	151,104	82,369	17,559	25,588	37.97	88	37	8	133
T. linearis Benth.	151,828	82,941	17,675	25,606	37.86	88	37	8	133

Table 3. Genes in the cp genome of 13 Lamiaceae species.

Category	Group	Genes
Photosynthetic	Subunits of photosystem I	*psaA,psaB,psaC,psaI,psaJ,ycf1(x2),ycf15(x2),ycf2(x2),ycf3**,ycf4*
	Subunits of photosystem II	*psbA,psbB,psbC,psbD,psbE,psbF,psbH,psbI,psbJ,psbK,psbL,psbM,psbN,psbT,psbZ*
	Subunits of NADH dehydrogenase	*ndhA*,ndhB*(x2),ndhC,ndhD,ndhE,ndhF,ndhG,ndhH,ndhI,ndhJ,ndhK*
	Subunits of cytochrome b/f complex	*petA,petB*,petD*,petG,petL,petN*
	Subunits of ATP synthase	*atpA,atpB,atpE,atpF*,atpH,atpI*
	large subunit of RubisCO	*rbcL*
Self–replication	Large subunit of ribosomal	*rpl14,rpl16*,rpl2*(x2),rpl20,rpl22,rpl23(x2),rpl32,rpl33,rpl36*
	Samll subunit of ribosomal	*rps11,rps12**(x2),rps14,rps15,rps16*,rps18,rps19,rps2,rps3,rps4,rps7(x2),rps8*
	Subunits of RNA polymerase	*rpoA,rpoB,rpoC1*,rpoC2*
	Ribosomal RNAs	*rrn16(x2),rrn23(x2),rrn4.5(x2),rrn5(x2)*
	Transfer RNAs	*trnA–UGC*(x2),trnC–GCA,trnD–GUC,trnE–UUC,trnF–GAA,trnG–GCC,trnG–UCC*,trnH–GUG#,trnI–CAU(x2),trnI–GAU*(x2),trnK–UUU*,trnL–CAA(x2),trnL–UAA*,trnL–UAG,trnM–CAU(x2),trnN–GUU(x2),trnP–UGG,trnQ–UUG,trnR–ACG(x2),trnR–UCU,trnS–GCU,trnS–GGA,trnS–UGA,trnT–GGU,trnT–UGU,trnV–GAC(x2),trnV–UAC*,trnW–CCA,trnY–GUA*
	Tanskational initiation factor	*infA*
Other	Protease	*clpP***
	Maturase	*matK*
	Envelope membrance protein	*cemA*
	c–type cytochrome synthesis gene	*ccsA*
	Subunit of Acetyl–CoA–carboxylase	*accD*

Note: The parentheses stand for multi–copy genes; * and ** refer to genes containing 1 intron and 2 introns, respectively; # represents the specific gene of *E. miltiorrhiza*.

83,416 bp) and SSC region (17,177–17,692 bp) were separated by two IRs (IRa and IRb, 24,956–25,790 bp), and the GC content was 37.83–38.49%, much lower than the AT content.

The cp genome sequence of each species encoded 131–133 genes, including 88 protein–coding genes, 37 tRNA genes and 8 rRNA genes (Table 2). Taking the cp genome of *A. nubigena* as an example, the encoded genes can be divided into various types according to their product functions. There were 49 genes associated with photosynthesis, of which four genes (*ycf1*, *ycf15*, *ycf2* and *ndhB*) had two copies. Fifty–nine genes were involved in self–replication, 16 of which had two copies. There were also five additional genes (Table 3). In 12 of the species, except for *E. densa*, there were 20 multicopy genes, and 18 genes (*trnK–UUU*, *rps16*, *trnG–UCC*, *atpF*, *rpoC1*, *ycf3*, *trnL–UAA*, *trnV–UAC*, *clpP*, *petB*, *petD*, *rpl16*, *rpl2*, *ndhB*, *rps12*, *ndhA*, *trnA–UGC*, *trnI–GAU*) contained one or two introns. Among the 18 genes, three (*rps12*, *clpP*, *ycf3*) contained two introns. *E. densa* has an extra gene (*trnH–GUG*) that contains an intron, which is consistent with previous studies [22].

3.2 SSR Analysis

A total of 34–57 simple repeat sequences of 6 types were detected in the 13 cp genomes, among which *E. fruticosa* had the highest SSR number and *M. complanatum* and *N. hemsleyana* had the lowest SSR numbers (Fig. 3). Among the six types of repeats, single–nucleotide repeats accounted for the highest proportion (61.07%), followed by dinucleotide (16.42%) and tetranucleotide (16.78%) repeats. Trinucleotide, pentanucleotide and hexanucleotide repeats accounted for 3.50%, 2.02% and 0.18%, respectively. Only one hexanucleotide repeat was detected in *G. pseudogentiana*. The length of single–nucleotide tandem repeats accounted for 41.88% of the total sequence length, followed by the trinucleotide sequence length (30.26%), dinucleotide sequence length (15.87%) and pentanucleotide sequence length (9.96%). The tetranucleotide sequence length accounted for 2.03%. In single–nucleotide sequences, A/T accounted for 97.58%.

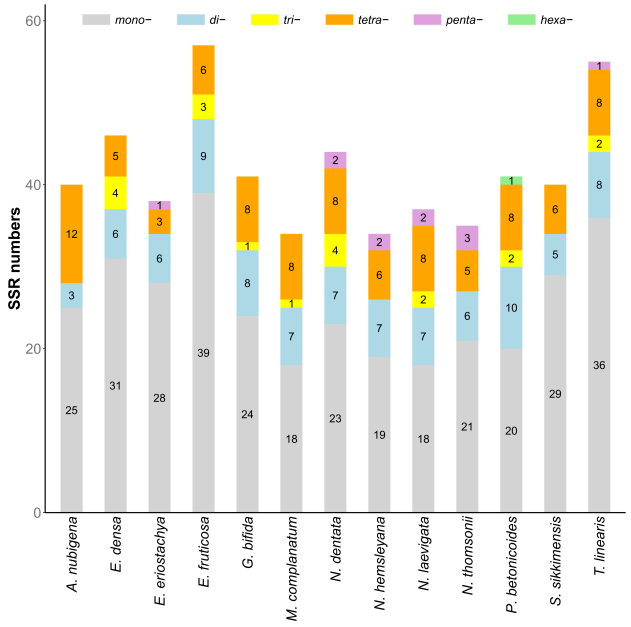

Fig. 3. Type and number of SSRs in cp genome of 13 Lamiaceae species.

SSR sequences were not evenly distributed in these cp genomes. They were distributed mostly in the LSC region, accounting for approximately 65–85%, while the remaining 15–35% were distributed in the LSC and IR regions. The SSR sequences of *E. fruticosa*, *M. complanatum* and *P. betonicoides* were not distributed in the IR region. SSR loci showed high polymorphism in the cp genomes of the 13 Lamiaceae plants, which provided a basis for the subsequent development of SSR molecular markers [22].

In the scattered long repeats in the cp genome, a total of 477 long repeats were detected in the 13 species analysed by REPuter software. *E. fruticosa* had the largest number of long repeats, while *E. densa* had the smallest number (Fig. 4). There were four types of long repeats, including forward, palindromic, reverse and complement. Among them, *E. fruticosa* and *E. densa* had all four types. *A. nubigena*, *E. eriostachya*, *G. bifida* and *S. sikkimensis* only had forward and palindromic types; and the other seven species had forward, palindromic and reverse types.

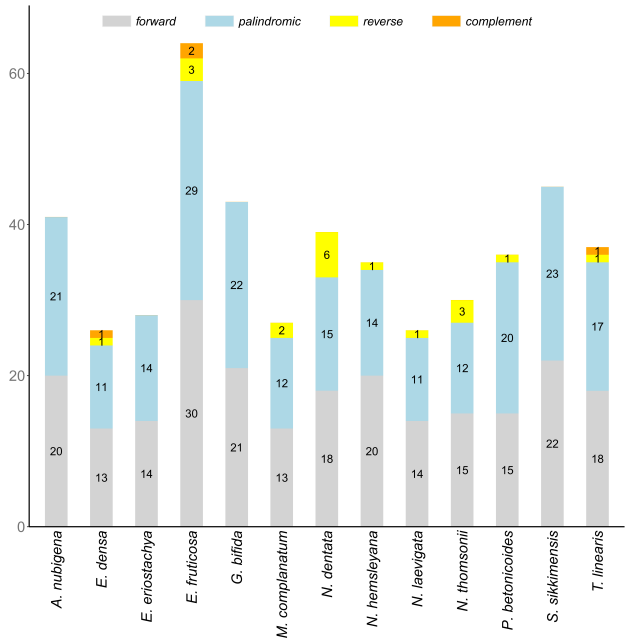

Fig. 4. Long repeats in the cp genomes of 13 Lamiaceae species.

3.3 Analysis of Codon Preference

The RSCU value refers to the ratio between the actual usage frequency of a certain codon and its theoretically expected usage frequency, which is often used as an important parameter to measure codon bias. An RSCU value >1 indicates that the codon usage bias is strong. Such codons are high–frequency codons [44]. A total of 819 codons were found in the 13 species of the Lamiaceae. In these codons, 403 codons were high–frequency codons with RSCU values >1, among which 61 codons ended in A/T(U), accounting for 98.13% (**Supplementary Fig. 1**), indicating that high–frequency codons preferred to end in A/T(U). The lowest RSCU value was 0.30, and the highest RSCU value was 3.97.

3.4 Ka/Ks Analysis

During plant evolution, genes are affected by nonsynonymous substitutions (Ka) and synonymous substitutions (Ks). The ratio Ka/Ks can be used to indicate the evolutionary rate to determine whether gene sequences have selective pressure in the process of evolution [45]. When Ka = Ks, the gene is not under selective pressure; when Ka

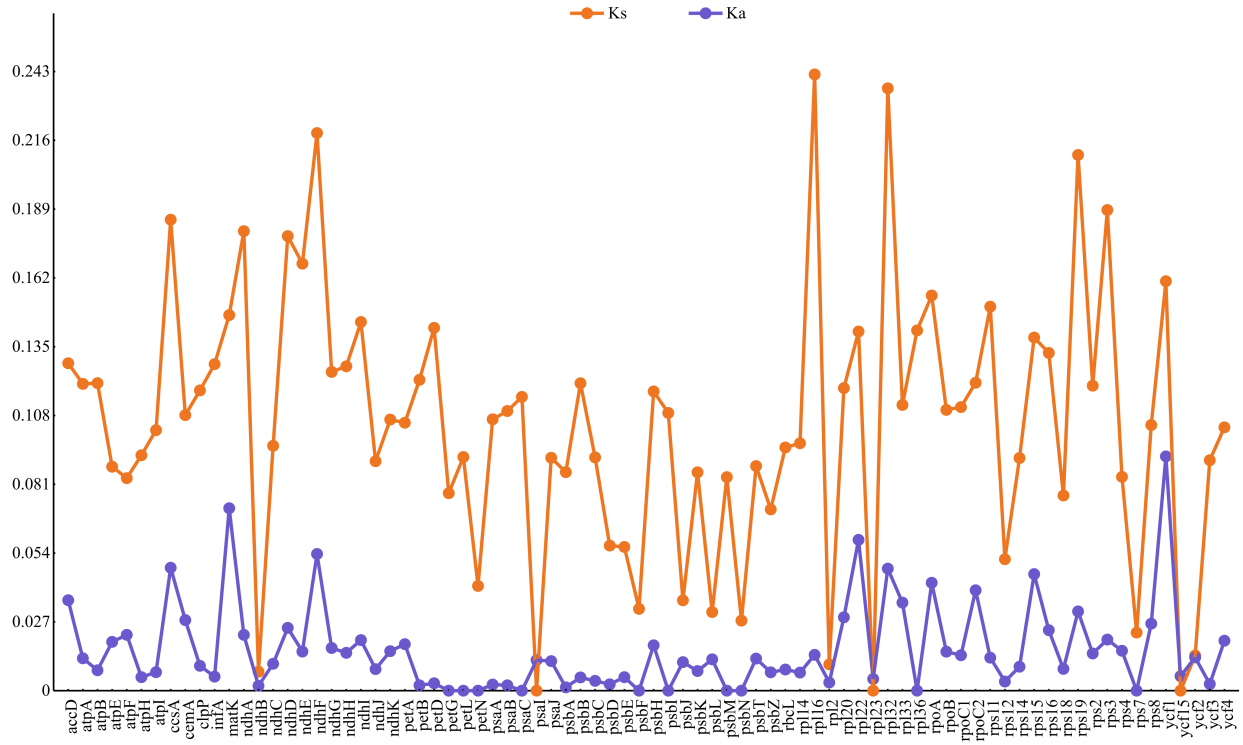

Fig. 5. Analysis of gene selection pressure of 13 Lamiaceae species. Note: The x–axis represents the gene, and the y–axis represents the value of Ka/Ks.

> Ks, the gene is under positive selective pressure. The Ka/Ks ratio of genes in the cp genomes of the 13 species was 0–0.58 (Fig. 5), indicating that most genes are subjected to purification selective pressure. The Ka/Ks ratios of 8 photosynthesis–related genes (*petG*, *petL*, *petN*, *psaC*, *psbF*, *psbI*, *psbM*, *psbN*), a ribosomal large gene (*rpl36*) and a small subunit gene (*rps7*) were all 0. The Ks values of 3 genes (*psaI*, *rpl23* and *ycf15*) were also 0.

3.5 Nucleotide Diversity Analysis

Nucleotide diversity (Pi) values were calculated to identify divergence hotspots. With Pi values ranging from 0.000 to 0.230, a total of 754 variation sites were detected (Fig. 6). In addition, there were 8 variation sites with Pi values >0.12. Between 4810–5400, 7001–7600, 8001–8600, 71,801–72,400, 133,601–134,200, 135,401–136,000, 135,601–136,200, and 135,801–136,400, four genes were distributed in the LSC region, two in the SSC region and two in the IRb region. Higher levels of genetic variation were detected in the LSC and SSC regions, suggesting that rapid nucleotide substitution in the Lamiaceae species may be involved, which may provide further consideration for the identification of the Lamiaceae species and the construction of phylogeny.

3.6 IR Boundary Analysis

The expansions and contractions of the IR regions often result in genome size variations among various plant lineages. Therefore, the analysis of the boundary changes of the cp genome can reflect the evolutionary process [46]. In the cp genomes of *N. laevigata*, *rpl22/trnH* and *trnH/psbA* were located on the two sides of the LSC/IRb and LSC/IRa boundaries, respectively (Fig. 7). In the cp genomes of other species, *rps19/rpl2* and *rpl2/trnH* were discovered on the two sides of the two boundaries. The expansion of the IR boundary allows it to enter the coding regions of nearby genes, which will lead to the formation of pseudogenes in the other IR copy region due to the inverted repeat characteristic of IR [23]. In this study, the LSC/IRb boundary entered the coding region of the *rps19* gene, and the SSC/IRa boundary entered the coding region of the *ycf1* gene; therefore, both *ycf1* and *rps19* became pseudogenes. The IR boundaries of the cp genomes of the other 12 Lamiaceae species except *N. laevigata* were essentially conserved, suggesting that the evolutionary direction of *N. laevigata* may be different from those of the other 12 species.

3.7 Collinearity Analysis

The order of genes in the 13 cp genomes were essentially identical, no rearrangement or inversion were detected among these genomes (**Supplementary Fig. 2**).

3.8 Sequence Variation Analysis

To investigate intergeneric differences among cp genomes, the identity percentage of these species was plotted using the mVISTA program, taking the cp genome of

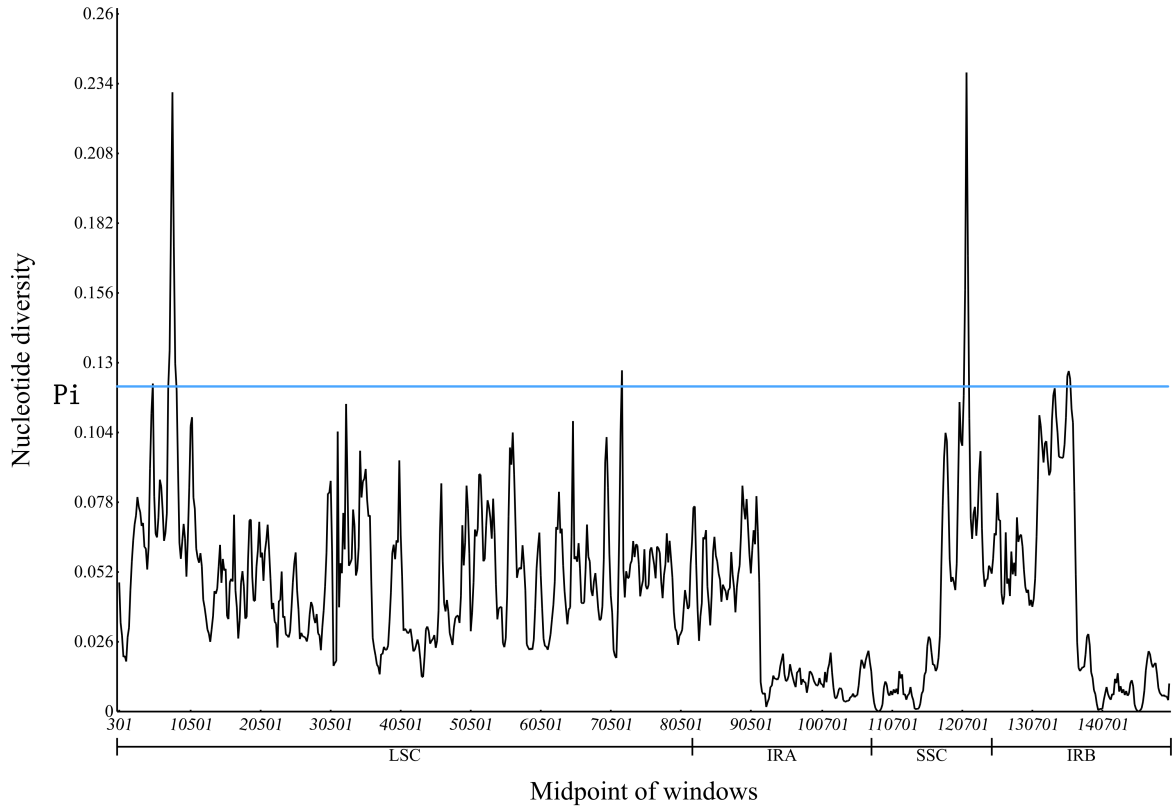

Fig. 6. Nucleotide diversity analysis of 13 Lamiaceae species.

A. nubigena as a reference. A high degree of similarity was detected among the 13 cp genomes. More variations were detected in the LSC/SSC region than in the IR region. The variation in the non–coding region was significantly higher than that in the coding region, indicating that the 13 cp genomes were conserved (**Supplementary Fig. 3**).

3.9 Phylogenetic Relationship Analysis

In this study, we constructed an ML phylogenetic tree (Fig. 8) and analysed the genetic relationship. During sequence alignment, the alignment length was 18,755,965 bp. The results showed that all species comprise eight clades, which consisted of the Nepetoideae subfamily, Lamioideae subfamily, Ajugoideae subfamily, Scutellarioideae subfamily, Premnoideae subfamily, Viticoideae subfamily, *Tectona* genus, and *Callicarpa* genus (Fig. 8). The major clade made from the Nepetoideae subfamily was composed of species of 12 genera, e.g., the *Salvia*, *Melissa* and *Mentha* genera; the Viticoideae subfamily was composed of the *Vitex* genus; the Lamioideae subfamily was composed of species of 13 genera, e.g., the *Stenogyne*, *Phyllostegia* and *Haplostachys* genera; the Scutellarioideae subfamily was composed of the *Scutellaria* and *Holmskioldia* genera; the Ajugoideae subfamily was composed of species of four genera, e.g., the *Ajuga*, *Teucrium* and *Caryopteris* genera; and the Premnoideae subfamily was composed of the *Gmelina* and *Premna* genera. These results are consistent with the findings of Li [47]. Except for three branch nodes, the bootstrap value of each node was greater than 70, indicating that the constructed ML tree was reliable. In addition, 97 species could also be divided into eight tribes (Fig. 8). Except for *Lamium galeobdolon* (NC036972.1) which was not clustered with the *Lamium* genus, the phylogenetic relationships of other species in the ML tree at the tribe level were consistent with those in Li's study [47].

4. Discussion

The cp genomes of 13 Lamiaceae plants in Tibet were sequenced and characterized for the first time. Their structure, length and GC content were similar to those of other Lamiaceae species [22–26]. Similar to previous reports [15,47], the *rps12*, *clpP* and *ycf3* genes all contained two introns.

In the cp genomes of 12 species, 133 genes were annotated. However, the cp genome of *E. densa* had only 131 genes because one *trnM–CAU* gene and two *ycf2* genes were missing, and one *trnfM–CAU* gene was added, which was consistent with a previous report on *E. densa* [10]. Compared with the published cp genome of *E. densa* [10], one more *rps16* gene was detected in the cp genome assembled in this study. Although the deletion of the *rps16* gene has also been found in some angiosperms [48], it was found in the cp genomes of other species in this study and other published cp genomes of the Lamiaceae. Structural vari-

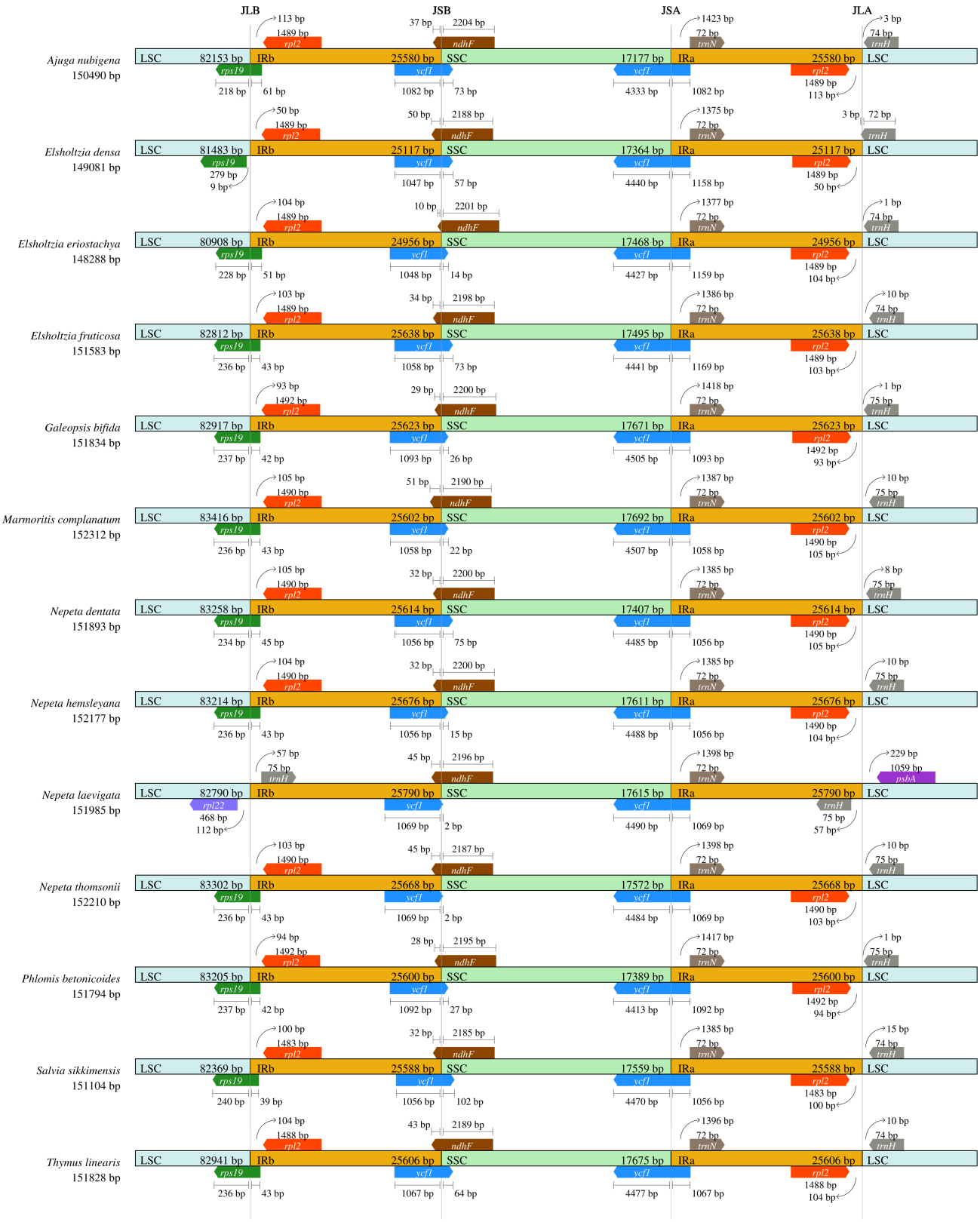

Fig. 7. Comparative analysis of cp genome boundary regions (LSC, SSC, IR) in 13 Lamiaceae species. Note: The coloured squares indicate that the large single copy (LSC) area and the small single copy (SSC) area are separated by two reverse repeat areas (IRa and IRb). JLB (LSC and IRb boundary), JSB (IRb and SSC boundary), JSA (SSC and IRa boundary), JLA (IRa and lSC boundary).

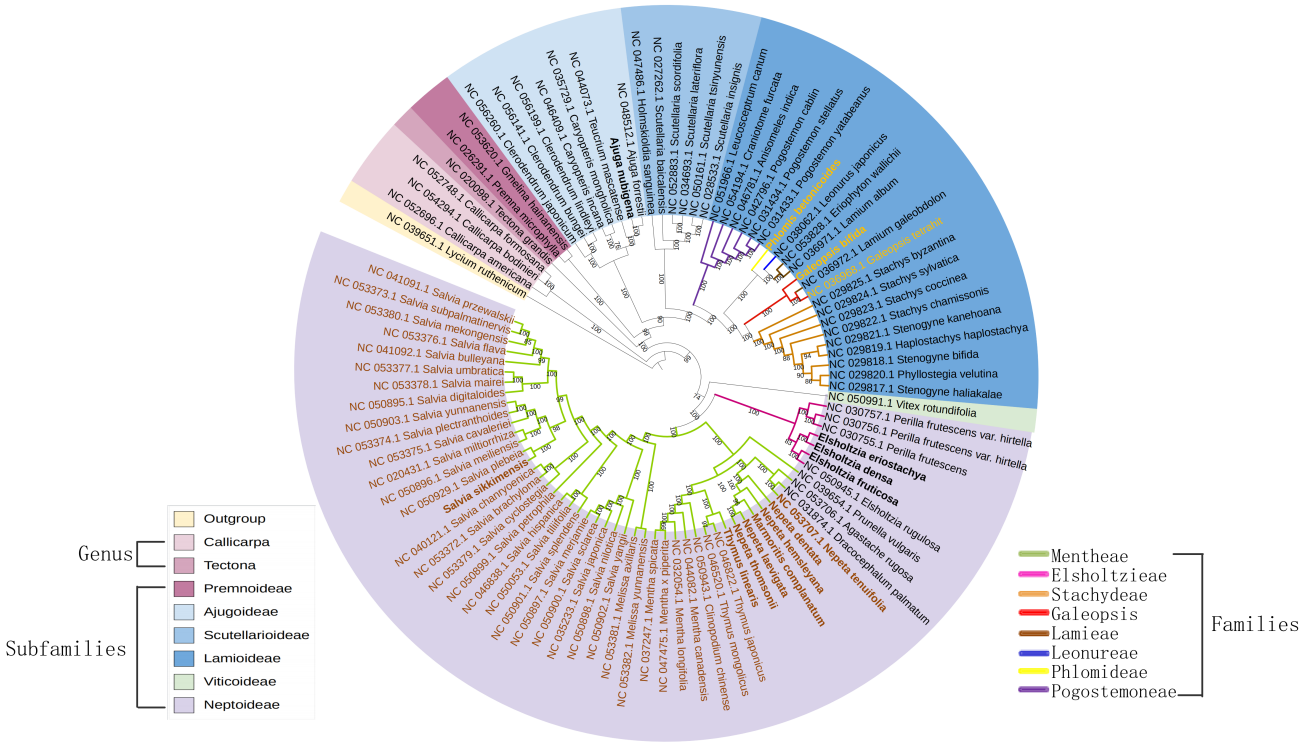

Fig. 8. The ML phylogenetic tree of cp genomes of 97 Lamiaceae species. Note: Coloured squares represent subfamilies and genera, coloured horizontal lines represent families. Brown fonts represent sister groups and orange fonts represent the complex group.

ations and gene loss of cp genomes are important for the study of plant phylogeny.

The contraction and expansion at the boundary of the IR region can cause length changes in cp genomes [49]. In this study, *N. laevigata* showed a deletion of the *rpl2* and *rps19* genes, which was different from the other 12 species. The discovery of hypermutated regions in cp genomes is of great value in phylogenetic and population studies of related species [50]. Eight hypervariable regions (4810–5400, 7001–7600, 8001–8600, 71,801–72,400, 133,601–134,200, 135,401–136,000, 135,601–136,200, and 135,801–136,400) were identified by analysing nucleotide diversity, which can provide potential targets for the development of molecular markers and phylogenetic studies [51].

In this study, cp genomes were used to construct an ML phylogenetic tree of the Lamiaceae, and the results showed that the eight major clades corresponded to the subfamily classification of Li [47]. Except for one species, the phylogenetic relationships at the tribe level were also consistent with those in Li's study [47]. Bendiksby [52] reconstructed the phylogenetic tree of the Lamiaceae subfamily using DNA fragments from the cp genomes and added a new clade, *Galeopsis*, based on the work of Li [47], which was also consistent with the classification results of this study. Among the species in this study, the genera *Salvia*, *Thymus* and *Nepeta* have a common ancestor and can be considered "sister groups". The genera *Galeopsis* and *Phlomoides* belong to the subfamily Lamioideae and form a complex group, indicating that they have closer genetic relationships than other species, which is consistent with previous studies [47,53,54].

Morphological and geographical data can provide a better understanding of molecular phylogenetic classification. Among the species sequenced in this study, *N. laevigata* and *N. thomsonii* have round heart–shaped leaves and spikes. *N. dentata* has ovate–oblong leaves, *N. hemsleyana* has linear lanceolate leaves, and *Marmoritis complanatum* has ovate–round or kidney–like leaves. The five species of plants have the common morphological characteristics of the *Nepeta* genus, namely, the stems are prismatic, and the leaves are white pilose. In this study, the ML phylogenetic tree showed that *N. laevigata* and *N. thomsonii* constituted sister groups, and *M. complanatum*, *N. dentata* and *N. hemsleyana* formed sister groups with both of them. Therefore, the morphological characteristics verified the rationality of the molecular phylogenetic classification. From a geographical point of view, *M. complanatum* is mainly found in the mountains and rocks in the eastern area of Tibet. The special morphological characteristics of *M. complanatum* were caused by the parallel evolution of the *Nepeta* genus due to its geographical location [55]. They did not form sister groups, which was consistent with the relative relationships inferred from leaf morphology. *Lamium galeobdolon* and other species of the *Lamiaceae* tribe were not grouped together, though the cause remains to be determined.

5. Conclusions

This study shows that it is feasible to use cp genomes to classify species in the Lamiaceae at the subfamily and tribe levels, which is basically consistent with morphological classification and can provide a molecular basis for the study of the evolutionary relationships of the Lamiaceae species.

Availability of Data and Materials

The datasets used and/or analysed during the current study are available from the first author upon reasonable request. Raw data for the cp genomes sequencing can be accessed by logging on to the website https://www.ncbi.nlm.nih.gov/.

Author Contributions

These should be presented as follows: XW and ZM designed the research study. YN, QQ, YD, and SZ, performed the research. XW, QQ, YD, and SZ provided help and advice on paper writing. YN, XW, and QQ analysed the data. YN, QQ and XW wrote the manuscript. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript.

Ethics Approval and Consent to Participate

The plant materials of 13 species of the Lamiaceaeused in this study were collected from the Tibet Autonomous Region of China (Tibet) and identified by Professor Guoyue Zhong of Jiangxi University of Chinese Medicine (JUTCM). The certificate specimens were all stored in the herbarium of JUTCM.

Acknowledgment

Not applicable.

Funding

This research was funded by the Jiangxi University of Chinese Medicine Science and Technology Innovation Team Development Program (Grant CXTD22002).

Conflict of Interest

The authors declare no conflict of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.31083/j.fbl2806110.

References

[1] Editorial Committee of Flora Reipublicae Popularis Sinicae, Chinese Academy of Sciences. Flora Reipublicae Popularis Sinieae65(2) (pp. 3). China Science Publishing & Media Lid: BeiJing. 1977. (In Chinese)

[2] Biao L, Sheng LX, Dong GH, Ma GH. Investigation on Traditional Tibetan Medicine Plant Resources of Labiate in Gannan County of Qinghai–Tibet Plateau. Journal of Mudanjiang Normal University. 2016: 43–46. (In Chinese)

[3] Li WS, Jian L, Lan XZ. Diversity of medicinal plant resources in the Lhasa River Reaches, Tabet. Lishizhen Medicine and Materia Medica Research. 2013; 24: 1480–1483. (In Chinese)

[4] Sun TJ, Fan HY. On the Present Situation and Suggestions of Conservation and Development of Tibetan Medicinal Materials. Journal of Medicine & Pharmacy of Chinese Minorities. 2013; 19: 7–11. (In Chinese)

[5] Patrignani F, Prasad S, Novakovic M, Marin PD, Bukvicki D. Lamiaceae in the treatment of cardiovascular diseases. Frontiers in Bioscience-Landmark. 2021; 26: 612–643.

[6] Keshavarzi M, Jahandideh R, Bokaee ZN. Morphological and anatomical studies on Ziziphora clinopodioides Lam. (Labiatae). Pakistan Journal of Biological Sciences. 2008; 11: 2599–2605.

[7] Furukawa M, Makino M, Ohkoshi E, Uchiyama T, Fujimoto Y. Terpenoids and phenethyl glucosides from Hyssopus cuspidatus (Labiatae). Phytochemistry. 2011; 72: 2244–2252.

[8] Muñoz O, Peña RC, Montenegro G. Iridoids from Stachys grandidentata (Labiatae). Zeitschrift Fur Naturforschung. C, Journal of Biosciences. 2001; 56: 902–903.

[9] Zhao F, Chen YP, Salmaki Y, Drew BT, Wilson TC, Scheen AC, *et al*. An updated tribal classification of Lamiaceae based on plastome phylogenomics. BMC Biology. 2021; 19: 2.

[10] Fu G, Liu J, Li JQ. Characterization of chloroplast genome structure and phyletic evolution of Elsholtzia densa. Chinese Traditional and Herbal Drugs. 2022; 53: 1844–1853. (In Chinese)

[11] Pu F, Zhang DQ. Chloroplast Genome Characteristics of a Tibetan Endangered Plant Lamiophlomis rotata. Journal of Dali University. 2022; 7: 22–28. (In Chinese)

[12] Du Q, Wang LQ, Chen ZE, Jiang M, Chen HM, Zeng J, *et al*. Characterization and phylogenetic analysis of the complete chloroplast genome of Lycopus europaeus. Acta Pharmaceutica Sinica. 2022; 57: 2206–2215. (In Chinese)

[13] Shen LQ. Complete Chloroplast Genomes of Three Lamiaceae Medicinal Plants: Comparative and Evolutionary Analysis in Lamiaceae [master's thesis]. Zhjiang University. 2018. (In Chinese)

[14] Su T. Phylogenetic Study on Chloroplast Genome of Salvia subgenus of Labiatae [master's thesis]. Guizhou University. 2021. (In Chinese)

[15] Wang Y, Wang S, Liu Y, Yuan Q, Sun J, Guo L. Chloroplast genome variation and phylogenetic relationships of Atractylodes species. BMC Genomics. 2021; 22: 103.

[16] Li G, Zhang L, Xue P. Codon usage pattern and genetic diversity in chloroplast genomes of Panicum species. Gene. 2021; 802: 145866.

[17] Pervaiz T, Zhang C, Faheem M, Mu Q, Fang J. Chloroplast based genetic diversity among Chinese grapes genotypes. Mitochondrial DNA. Part A, DNA Mapping, Sequencing, and Analysis. 2017; 28: 565–569.

[18] Chávez-Pesqueira M, Núñez-Farfán J. Genetic diversity and structure of wild populations of Carica papaya in Northern Mesoamerica inferred by nuclear microsatellites and chloroplast markers. Annals of Botany. 2016; 118: 1293–1306.

[19] Zhang TT, Yang Y, Song XY, Gao XY, Zhang XL, Zhao JJ, *et al*. Novel Structural Variation and Evolutionary Characteristics of Chloroplast tRNA in *Gossypium* Plants. Genes. 2021; 12: 822.

[20] Zhu B, Qian F, Hou Y, Yang W, Cai M, Wu X. Complete chloroplast genome features and phylogenetic analysis of Eruca sativa (Brassicaceae). PLoS ONE. 2021; 16: e0248556.

[21] Liu F, Movahedi A, Yang W, Xu D, Jiang C, Xie J, *et al*. The complete chloroplast genome and characteristics analysis of Musa basjoo Siebold. Molecular Biology Reports. 2021; 48: 7113–7125.

[22] Liu F, Movahedi A, Yang W, Xu L, Xie J, Zhang Y. The complete chloroplast genome and characteristics analysis of Callistemon rigidus R.Br. Molecular Biology Reports. 2020; 47: 5013–5024.
[23] Zhao C, Xu W, Huang Y, Sun Q, Wang B, Chen C, *et al*. Chloroplast genome characteristics and phylogenetic analysis of the medicinal plant Blumea balsamifera (L.) DC. Genetics and Molecular Biology. 2021; 44: e20210095.
[24] Zhang L, Wang S, Su C, Harris AJ, Zhao L, Su N, *et al*. Comparative Chloroplast Genomics and Phylogenetic Analysis of *Zygophyllum* (Zygophyllaceae) of China. Frontiers in Plant Science. 2021; 12: 723622.
[25] Jung J, Do HDK, Hyun J, Kim C, Kim JH. Comparative analysis and implications of the chloroplast genomes of three thistles (*Carduus* L., Asteraceae). PeerJ. 2021; 9: e10687.
[26] Lin Z, Zhou P, Ma X, Deng Y, Liao Z, Li R, *et al*. Comparative analysis of chloroplast genomes in Vasconcellea pubescens A.DC. and Carica papaya L. Scientific Reports. 2020; 10: 15799.
[27] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30: 2114–2120.
[28] Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology. 2009; 10: R25.
[29] Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, *et al*. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biology. 2020; 21: 241.
[30] Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, *et al*. CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Research. 2019; 47: W65–W73.
[31] Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, *et al*. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012; 28: 1647–1649.
[32] Lohse M, Drechsel O, Kahlau S, Bock R. OrganellarGenomeDRAW–a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Research. 2013; 41: W575–W575–81.
[33] Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik. 2003; 106: 411–422.
[34] Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Research. 2001; 29: 4633–4642.
[35] Peden JF. Analysis of codon usage [PhD thesis]. University of Nottingham: UK. 1999.
[36] Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution. 2007; 24: 1586–1591.
[37] Katoh K, Kuma KI, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Research. 2005; 33: 511–518.
[38] Hall T. A.BioEdit:A User–Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT.Nucleic Acids Symposium Series. 1999; 41: 95–98.
[39] Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; 25: 1451–1452.
[40] Amiryousefi A, Hyvönen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics. 2018; 34: 3030–3031.
[41] Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, *et al*. VISTA: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics. 2000; 16: 1046–1047.
[42] Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, *et al*. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Molecular Ecology Resources. 2020; 20: 348–355.
[43] Oldenburg DJ, Bendich AJ. Most chloroplast DNA of maize seedlings in linear molecules with defined ends and branched forms. Journal of Molecular Biology. 2004; 335: 953–970.
[44] Anue MR, Sun, XL, Cheng CZ, Lai ZX. Analysis of co–don usage pattern of banana basic secretory protease gene. Plant Diseases and Pests. 2019; 10: 1–4, 9.
[45] Nguyen PAT, Kim JS, Kim JH. The complete chloroplast genome of colchicine plants (Colchicum autumnale L. and Gloriosa superba L.) and its application for identifying the genus. Planta. 2015; 242: 223–237.
[46] Liu LX, Li R, Worth JRP, Li X, Li P, Cameron KM, *et al*. The Complete Chloroplast Genome of Chinese Bayberry (*Morella rubra*, Myricaceae): Implications for Understanding the Evolution of Fagales. Frontiers in Plant Science. 2017; 8: 968.
[47] Li B, Cantino PD, Olmstead RG, Bramley GLC, Xiang CL, Ma ZH, *et al*. A large-scale chloroplast phylogeny of the Lamiaceae sheds new light on its subfamilial classification. Scientific Reports. 2016; 6: 34343.
[48] Khan AL, Asaf S, Lubna, Al-Rawahi A, Al-Harrasi A. Decoding first complete chloroplast genome of toothbrush tree (Salvadora persica L.): insight into genome evolution, sequence divergence and phylogenetic relationship within Brassicales. BMC Genomics. 2021; 22: 312.
[49] Zhang J, Lu JH, Wang QQ, Nan LM, Xu K. Charateristics of the chloroplast genome of Glycyrrhiza eurycarpa P.C.Li from Xinjiang with comparison and phylogenetic analysis of the chloroplast genomes of the medicinal plants of Glycyrrhiza. Acta Pharmaceutica Sinica. 2022; 57: 1516–1525. (In Chinese)
[50] Ren F, Wang L, Li Y, Zhuo W, Xu Z, Guo H, *et al*. Highly variable chloroplast genome from two endangered Papaveraceae lithophytes *Corydalis tomentella* and *Corydalis saxicola*. Ecology and Evolution. 2021; 11: 4158–4171.
[51] Yang JP, Zhu ZL, Juan FY, Zhu F, Chen YJ, Niu ZT, *et al*. Comparative plastomic analysis of three Bulbophyllum medicinal plants and its significance in species identification. Acta Pharmaceutica Sinica. 2020; 55: 2736–2745. (In Chinese)
[52] Bendiksby M, Thorbek L, Scheen AC, Lindqvist C, Ryding O. An updated phylogeny and classification of Lamiaceae subfamily Lamioideae. Taxon. 2011; 60: 471–484.
[53] Munyao JN, Dong X, Yang JX, Mbandi EM, Wanga VO, Oulo MA, *et al*. Complete Chloroplast Genomes of *Chlorophytum comosum* and *Chlorophytum gallabatense*: Genome Structures, Comparative and Phylogenetic Analysis. Plants. 2020; 9: 296.
[54] Scheen AC, Bendiksby M, Ryding O, Mathiesen C, Albert AA, Lindqvist C. Molecular phylogenetics, character evolution, and suprageneric classification of Lamioideae (Lamiaceae). Annals of the Missouri Botanical Garden. 2010; 97: 191–217.
[55] Chen J. On the eurasian genus glechoivia Linn. and its relationship with allied genera. Plant Diversity. 1979: 81–89. (In Chinese)