

Appearance, molecule identification, and phylogenetic analysis of the three *Leonurus* species used for the treatment of gynecological diseases

Qing Du (✉ 2017001@qhmu.edu.cn)

Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College;
College of Pharmacy, Qinghai Minzu University <https://orcid.org/0000-0002-0732-3377>

Ziyi Rong

Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College;
College of Pharmacy, Xiangnan University <https://orcid.org/0000-0002-0577-1088>

Chang Zhang

Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College
<https://orcid.org/0000-0002-9384-7408>

Liqiang Wang

College of Pharmacy, Heze University <https://orcid.org/0000-0002-8366-7016>

Yanjuan Guan

Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College;
College of Pharmacy, Xiangnan University <https://orcid.org/0000-0001-7096-0072>

Haimei Chen

Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College
<https://orcid.org/0000-0001-7100-5915>

Bin Wang

College of Pharmacy, Xiangnan University <https://orcid.org/0000-0002-9939-7030>

Chang Liu (✉ cliu@implad.ac.cn)

Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College
<https://orcid.org/0000-0003-3879-7302>

Research Article

Keywords: *L. japonicus*, *L. cardiaca*, *L. sibiricus*, appearance traits, genome comparative, molecular markers, evolutionary analysis, nucleotide diversity, selective pressure, codon analysis

Posted Date: June 27th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3109565/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

We analyzed and compared the three *Leonurus* species from appearance characteristics, intrinsic gene contents and functions, gene transcribed spacer regions (ITS2) in the nuclear genome, and the process of evolutionary development. The chloroplast genomes of them were found to be common circular structures with the four regions, the length range from 151236bp to 151689bp including the identical number of 37 tRNA and 8 rRNA. The special trans-spliced *rps12* genes with two introns and three exons can be caught sight of *L. sibiricus*. The isoleucine (AUU) and cysteine (UGC) was the most and least abundant amino acid with the codons ending with most bases of A/U. The eleven hotspot divergent regions and four specific CDS genes were systematically counted with the highest nucleotide diversity. We cloned the DNA sequences of the two chosen IGS regions (*atpH-atpI* and *rps15-ycf1*) to develop the DNA barcodes, thus found that sixteen specific variable SNP sites and two Indel sites within the three *Leonurus* species. Furthermore, the ITS2 DNA sequences of 7 *Leonurus* species were varied after the 57th base in a total of 221 bases. In the aftermath of evolutionary analysis, the three *Leonurus* species were significantly clustered into one great clade, while they were located at the different sub-branch with similar topology and close relationship of *Phlomis rotata* based on the sequences of 64 shared nucleotide and ITS2 DNA. Thus above results can directly offer various evidence to better clarify the specific distinction among the three *Leonurus* species and their evolutionary history.

Significance Statements

The three *Leonurus* species were systematically distinguished through the extrinsic features, chloroplast genomics, and ITS2 regions of the ribosome components. Although they were clustered into one large clade, they have different evolutionary processes for their existence at the diverse sub-branch.

1. Introduction

The *Leonurus* species have an important medicinal value in the Lamiaceae family. Most of them are upright branches of herbs, about 20 species, with leaves near the palm division, and axillary verticillaster (Linn., 1754). The calyx is inverted-conical or tubular and bell-shaped. The corolla is white, pink to lavender with a flat top of a flower plate, and a small sharp prism of nuts. The *Leonurus* species are mainly distributed in Europe and the temperate zone of Asia, and a few species are found in America and Africa (Cardiaca, 1778). In China, there are almost 12 species, widely scattered, among which the species of *L. japonicus* Houtt. is used for official use with the name motherwort recorded in the monograph of Shengnong's Classic of Materia Medica (Jiao et al., 2020).

The three similar *Leonurus* species of *L. japonicus*, *L. cardiaca*, and *L. sibiricus* have distinct medical applications (Racg, 1961; Pitschmann et al., 2017). The whole grass of *L. japonicus* has the active ingredient of Leonurin, which is widely used for the treatment of women's amenorrhea, dysmenorrhea, irregular menstruation, excessive postpartum bleeding, lochia, postpartum uterine insufficiency, fetal movement, uterine prolapse, and the disorders of the red and white vaginal discharge (Zhao et al., 2022). The species of *L. cardiaca* was found in the Turkey of west Asia and Europe and has been applied to the treatment of heart disease, hypertension, hyperthyreosis, Graves' disease, dysphoria, insomnia, climacteric, postpartum complications, pruritus, acute posterior ganglionitis, flatulence, and irregular menstruation, etc (Angeloni et al., 2021). Whereas the species of *L. sibiricus* is distributed in inner Mongolia, northern of Shanxi and Hebei, Shanxi, Hei Longjiang, Jilin, and Liaoning in China. The whole grass and fruit have similar treated actions and also have the treated effects of edema and urine, small

abdominal distension pain, falling injury, liver heat headache, eye red swelling pain, and caligo corneae (Oliveira et al., 2017).

The structure of chloroplast ordinarily is composed of an outer chloroplast envelope, thylakoid, and stroma, which is an organelle containing chlorophyll energy for photosynthesis (Mohapatra et al., 2007). A variety of genes with different functions are included in the chloroplast and exerted distinct biological functions, that is to say, photosynthesis, gene expression, and unknown functions (Cui et al., 2020). Regarding the chloroplast genome of *Leonurus* species, several related types of research have been reported before. The chloroplast genomes of *L. japonicus* and *L. cardiaca* have been analyzed in the aspects of sequencing, assembling, repeats sequence, gene content, codon usages, and evolutionary developmental analysis.

Furthermore, the DNA sequences of ITS regions and the *matK* gene of ten *Leonurus* species including *L. japonicus* and *L. sibiricus* in China were amplified, sequenced, and investigated to find the phylogenetic relationship (Yang et al., 2011). The species populations of *L. cardiaca* genetics were comprehensively discriminated by AFLP, ISSR, RAPD, and IRAP molecular markers (Khadivi-Khub et al., 2014).

During the period of the plant seeding stage, the three kinds of plants are very similar in external shape until the mature flowering stage and they can be uneasy to be discriminated against. Once misidentified and misused, it will cause hidden dangers to medicinal effectiveness and safety. Therefore we systematically identified and studied the three plants from appearance traits, molecular markers, and evolutionary phylogeny given chloroplast genomes and ITS2 sequences in this study.

2. Materials and methods

2.1 Plant photo and materials

The identified photos of *L. japonicus*, *L. cardiaca*, and *L. sibiricus* were taken and provided by Dr. Lin Qinwen (Institute of Botany, Chinese Academy of Sciences, 112m, N39.99°, E116.20°, linqinwen83@163.com) from Beijing in China. The fresh three samples of *L. sibiricus* were sampled by Teacher Zhang Shumei, Zhang Keliang, and BCAS from Liaoning (geospatial coordinates: N41.75°, E120.13°), Inner Mongolia (geospatial coordinates: N39.37°, E111.50°), and Henan (geospatial coordinates: N31.83°, E114.08°) province and dried by silica gel. These donated voucher specimen was deposited at National Herbarium in China with the voucher number: BOP214746, BOP017659, and BOP017491, respectively (Contact person: CXU; Email: xuchao@ibcas.ac.cn). Furthermore, the three samples of *L. japonicus* and *L. cardiaca* by each were sampled by Professor Zhang Zhao and Mr. Zhang Chang at the Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing City (geospatial coordinates: E116.42°, N39.78°) saved the voucher number: impladzc22-1-1, impladzc22-1-2, impladzc22-1-3, impladzc22-2-1, impladzc22-2-2, and impladzc22-2-3, respectively (contact: Zhao Zhang; email: zhangzhao1962@tom.com). These above samples were used to extract the DNA genome for cloning the DNA barcode sequences.

2.2 Alignment and collinearity analysis of the three *Leonurus* species

The FASTA and GB files of three *Leonurus*' chloroplast genomes with registered No. containing the species of *L. japonicus* (MG673937.1 and NC_038062.1), *L. cardiaca* (NC_058592.1), and *L. sibiricus* (OP327561.1) were

downloaded from the website of the NCBI database (Table S1). The chloroplast genomes of the four *Leonurus* species were aligned by point to acquire the dot-plot graph on the website of NCBI database (<https://blast.ncbi.nlm.nih.gov/>) (Mohanty et al., 2022) and the collinearity comparisons of them were performed using Mauve multiplex genome alignment in Geneious software (v10.2.2, <https://www.geneious.com>, Buch et al., 2023; Darling et al., 2010).

2.3 Genome content and characteristics of 16 species' chloroplast genomes

The FASTA and GB files other 12 species were downloaded from the website of the NCBI database (Table S1). The basic contents and characteristics of genes from 16 species in this study were annotated by the CPGAVAS2 online tool (<http://www.herbalgenomics.org/chloroplastgavas2/>). The GC content of them was calculated by in-house Perl scripts (Table S1) (Arias et al., 2009). The comparison of genome structures and diverse splicing genes was visualized by the CPGview-RSG software (<http://www.1kmpg.cn/cpgview/>) (Liu et al., 2023).

2.4 Analysis of codon usage, Gene Oncology (GO), and Cluster of Orthologous Groups of proteins (KOG)

Relative Synonymous Codon Usage (RSCU), GO, and KOG of the three *Leonurus* species was calculated on the basis of the common proteins by using a script written in Perl procedure from the cloud platform of Genepioneer biotechnologies (<http://cloud.genepioneer.com>, Nanjing city, Jiangsu, China) and the results of them were visualized by the tools of CPJS draw in this platform (Li et al., 2023).

2.5 SSR and repeat analysis

The simple sequence repeats (SSRs) of the three *Leonurus* species were identified using MISA software (<https://webblast.ipk-gatersleben.de/misa/>) (Beier et al., 2017). The search parameters were referred to in the publication (Yang et al., 2022). Tandem repeats of the four cp genomes were predicted using the TRF software and REPuter program (<https://bibiserv.cebitec.uni-bielefeld.de/reputer>, Dr. Alexander Sczyrba, CeBiTec Bielefeld University), and the parameters were set as the references (Yang et al., 2022).

2.6 Analysis of inverted repeat (IR) boundary and genome comparison of the four *Leonurus* species

The boundaries of various regions and diverse genes from the four *Leonurus* species' chloroplast genome were visualized using the inverted-region (IR) scope software (<https://irscope.shinyapps.io/irapp/>) (Amiryousefi et al., 2018). The chloroplast genomes of the four *Leonurus* species were compared using the mVISTA procedure online with the chloroplast genome of *L. japonicus* (MG673937.1) as the reference genomics to find the potential divergence sequences (Mayor et al., 2000).

2.7 Analysis of nucleotide in the intergenic regions and ITS2 sequences

To explore the sequence divergence of nucleotides in the intergenic regions, encoding genes, and Internal Transcribed Spacer 2 (ITS2) sequences in the three *Leonurus* species, the informatics platform of Genepioneer (<http://cloud.genepioneer.com>) was used for computing nucleotide diversity (π) values. Moreover, the sequences of ITS2 in the seven *Leonurus* species (Registered Number: MN718258.1, OQ389953.1, EF395809.1, EF395808.1,

DQ903315.1, DQ903316.1, and EF395806.1) were downloaded from the NCBI database (<https://www.ncbi.nlm.nih.gov/>) and annotated by ITS2 database (<http://its2.bioapps.biozentrum.uni-wuerzburg.de/>) (Chen et al., 2010) to extract and confirm the correct sequences. After that, the nucleotide sequences of them were commonly compared to discriminate them by using Clustal X2.0 (Larkin, et al., 2007), GeneDoc software (Lanave et al., 2002), and depicted them using the software of Adobe photoshop2021 (McLean et al., 2002).

2.8 Identification of hypervariable regions

The genetic distances of intergenic spacers (IGSs) from the chloroplast genomes of the three *Leonurus* species were calculated by using the distmat program from EMBOSS (v6.3.1) (Sarachu et al., 2005) with the Kimura 2-parameters (K2p) evolutionary model (Tang et al., 2016). The highest K2p values of 30 hypervariable regions have been protracted in the graph. Moreover, the two selected fragments of divergence sequences with specific k2p values were chosen to develop the validation of the nucleotide diversity in the three *Leonurus* species.

2.9 DNA extraction, identification, and sequence cloning validation of the selected two hypervariable regions

Total genomic DNAs of *L. japonicus* and *L. cardiaca* were extracted using the plant genomic DNA kit (Tiangen Biotech, Beijing, China), while that of *L. sibiricus* was extracted by the CTAB method (Nishii et al., 2023). The DNA purity and amplification products were detected by 1.0% agarose gel electrophoresis stained with ethidium bromide alongside a 100 bp ladder (New England Biolabs, Ipswich, MA, USA) using the DNA marker as the reference (Takara) (Asad et al., 2023). The DNA concentration was determined by using a Nanodrop spectrophotometer 2000 (Thermo, Waltham, Massachusetts, USA) (Butler et al., 2022). Two pairs of polymerase chain reaction (PCR) amplification primers were identified from the two highest diverse IGS sequences using BioXM software (<http://202.195.246.60/BioXM/>) in the three *Leonurus* species. The PCR amplification system for the two hypervariable regions (*atpH-atpI* and *rps15-ycf1*) of each reaction included 12.5 μ L of 2 \times Taq PCR Master Mix (TransGen Biotech), 1.0 μ L of each primer (0.4 μ M), 2.0 μ L of extracted template DNA, and plusing ddH₂O to a final volume of 25.0 μ L (Du et al., 2022). A negative control (Milli-Q water in place of DNA template) was included in each PCR to ensure there was no contamination. All the amplifications were performed on a Pro-Flex PCR system (Applied Biosystems, Waltham, MA, USA) instrument with the amplification procedures as follows: degeneration 94°C for 2 min followed by 35 cycles of 94°C for 30 s, 56°C for 30 s, 72°C for 60 s, and a final extension step at 72°C for 2 min. The amplification products were saved at 4°C and sequenced using the Sanger sequencing platform (SinoGenoMax Co., Ltd., Beijing, China) with the same cloning primers on the ABI Prism 3730 Genetic Analyzer (Applied Biosystems, USA). The peak graph and sequences were apart analyzed by the Chromas and GeneDoc software (3.2) (Cittaro et al., 2016).

2.10 Analysis of selective pressure

The selection pressure of protein-coding genes was analyzed using the adaptive branch-site random effects likelihood (aBSREL) model implemented in the Hyphy software (Kosakovsky et al., 2020) and the informatics platform of genepioneer (<http://cloud.genepioneer.com>, Nanjing genepioneer Co., Ltd, Nanjing, Jiangsu, China). The values of Ka and Ks were calculated using KaKs_Calculator 2.0 and the results of Ka/Ks having P-value < 0.05 were selected as the suitable selective pressure (Wang et al., 2010).

2.11 Phylogenetic analysis

The chloroplast genomes of 16 species and the ITS2 DNA sequences of 19 species including four and seven *Leonurus* species were downloaded from the GenBank database, respectively with the two species of *Corydalis yanhusuo* and *Angelica sinensis* as outgroups. The shared 64 DNA sequences of cp genomes and the ITS2 DNA sequences were extracted by PhyloSuite (v1.2.2) (Zhang et al., 2020) software and aligned by using MAFFT(v7.313) (Kato et al., 2019) procedure. The phylogenetic tree was conducted based on the maximum likelihood (ML) method (Wascher et al., 2021) implemented in IQ-TREE(v1.6.8) (Nguyen et al., 2015) under the ccRev nucleotide substitution model (Janzen et al., 2022). The phylogenetic trees were visualized by using MEGA-5.0 (<https://www.megasoftware.net/> MEGA) (Tamura et al., 2011) and assessed by bootstrap analysis with 1000 replications.

3. Results

3.1 Plant characteristics of *L. japonicus*, *L. cardiaca*, and *L. sibiricus*

The species of *Leonurus* genus is an annual or biennial herb with dense roots and a shape stem erect is a blunt four prism. The lower leaf of the stem is oval and the middle leaf is diamond with mixed and stalked leaves. The leaves are fresh, tender, and green with juice. It has weak gas and a bitter taste. They are flowering from June to September and the nutlet fruits have been produced in the period from July to October. Because they have commonly been used in the treatment of gynecology, they have the name “benefit mother”.

The three *Leonurus* species have an obvious distinction in the flower color when they grow up to the ripe flowering stage. Their flowers have diverse colors as follows: *L. japonicus* (peach red), *L. cardiaca* (pink), and *L. sibiricus* (purple). In addition, the characteristics of tissues including stem, leaf, and flower has arguable feature. In the species of *L. japonicus*, the stem has inverted to the rough hair, especially dense in the section and edge and the base is sometimes close to hairless, with more branches (Fig. 1a). The leaf of the lower stem is ovate having the base wide wedge with 3 split palms. The corolla is pink to lavender, pubescent on the outside of the protruding calyx tube (Fig. 1a). While for the *L. cardiaca* species, the stem is blunt with four prisms, the smooth leaves are heart-shaped or wide oval, the apex is sharp-pointed, and usually 5–7 shallow cracked palms. The cover bud is axillary and the corolla is shaped like a lip (Fig. 1b). Meanwhile, in the species of *L. sibiricus*, the stem is blunt with four prisms with microgroove. The leaf of the middle stem is ovate with a wide wedge at the base and 3 deep splits as a palm. The flower is gradually spherical to the top and can be gathered into a long spike (Fig. 1c) (Pitschmann et al., 2017).

3.2 Assembly validation and collinearity analysis about chloroplast genomes of the four *Leonurus* species

To confirm the correctness of the assembly, we compared the four *Leonurus* chloroplast genomes with that of *L. japonicus* (MG673937.1) as the reference genome including the comparison. All the dot plots showed a diagonal line with two perpendicular lines, representing the IRs (Figure S1). The three other *Leonurus* species exhibited collinearity with that of *L. japonicus* (MG673937.1) and no rearrangement region could be found (Fig. 2). The result showed the high quality of assembly, sequence identity, and species relative, which have high uniformity regarding the nucleotide and regions (Brenner, 1966).

3.3 Differential gene contents among the chloroplast genomes in the four *Leonurus* species

After analysis of cp genomes in the four *Leonurus* species, in other words, *L. japonicus*, *L. cardiaca*, and *L. sibiricus*, they are all circular DNA molecules and conserved tetrad structure with the size is 151610bp, 151610bp, 151236bp, and 151689bp, respectively. Their genomes are commonly constructed by a LSC region, an SSC region, and a pair of IR regions with the lengths of 82827bp, 82827bp, 82294bp, and 82820bp; 17515bp, 17515bp, 17654bp, and 17619bp; 25634bp, 25634bp, 25644bp, and 25625bp; 25634bp, 25634bp, 25644bp, and 25625bp, respectively (Table S1 and Fig. 3). The length of CDS, tRNA rRNA, and non-coding regions in the four *Leonurus* species is 80391bp, 80391bp, 74508bp, and 80361 bp; 2834bp, 2834bp, 2833bp, and 2834bp; 9400bp, 9400bp, 9052bp, and 9400bp; 58985bp, 58985bp, 64843bp, and 59094bp. The overall GC content of them is 38.41%, 38.41%, 38.39%, and 38.41%, and that of the IR region (43.36%, 43.36%, 43.38%, and 43.37%) is higher than those of the LSC region (36.65%, 36.65%, 36.63%, and 36.65%) and the SSC region (32.23%, 32.23%, 32.16%, and 32.22%) (Table S1).

Moreover, the number of genes and proteins slightly changed from 132 to 133, from 87 to 88 in all 4 *Leonurus* species, respectively. While the number of tRNA and rRNA genes was commonly 37 and 8 (Table S1). Further comparison of cp genomes in the four *Leonurus* species, eight protein-coding genes (*rps12*, *rps7*, *rp12*, *rp123*, *ndhB*, *ycf1*, *ycf2*, and *ycf15*), seven tRNA-coding genes (*trnN*-GUU, *trnR*-ACG, *trnA*-UGC, *trnI*-GAU, *trnV*-GAC (×2), *trnL*-CAA, and *trnI*-CAU (×2)), and four types of rRNA-coding genes (*rrn16S*, *rrn23S*, *rrn5S*, and *rrn4.5S*) were located in the IR region (Table 1). In addition, twelve PCGs (*rps16*, *rps12* (×2), *atpF*, *rpoC1*, *petB*, *petD*, *rp12* (×2), *ndhB* (×2), and *ndhA*), and eight tRNA-coding genes (*trnK*-UUU, *trnG*-UCC, *trnL*-UAA, *trnV*-UAC, *trnI*-GAU, *trnA*-UGC, *trnA*-UGC, and *trnI*-GAU) contain one intron by each, whereas each of the other two PCGs (*ycf3* and *clpP*) contain two introns (Table S2, Figure S2). Differently, the *rp16* gene with one intron does not exist in the species of *L. cardiaca* (NC_058592.1, Figure S2), while the trans-spliced *rps12* genes with two introns and three exons can be found in the species of *L. sibiricus* (OP327561.1, Fig. 4).

Table 1

Comparison of the gene contents in the chloroplast genomes of *L. japonicus*, *L. cardiaca*, and *L. sibiricus*.

Species/Items		<i>L. japonicus</i>	<i>L. cardiaca</i>	<i>L. sibiricus</i>
Gene Function	Gene Type	Gene Name		
tRNA	tRNA genes	37 <i>trn</i> genes (include one intron in 6 genes)		
Photosynthesis	Large subunit of ribosome	<i>rpl14, rpl16, rpl2^a, rpl2^b, rpl20, rpl22, rpl23^a, rpl23^b, rpl32, rpl33, rpl36</i>		
	DNA dependent RNA polymerase	<i>rpoA, rpoB, rpoC1, rpoC2</i>		
	Small subunit of ribosome	<i>rps11, rps12^a, rps12^b, rps14, rps15, rps16, rps18, rps19, rps2, rps3, rps4, rps7^a, rps7^b, rps8</i>		
Gene expression	Subunits of ATP synthase	<i>atpA, atpB, atpE, atpF, atpH, atpI</i>		
	Subunits of photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ, ycf3</i>		
	Subunits of NADH-dehydrogenase	<i>ndhA, ndhB^a, ndhB^b, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>		
	Subunits of cytochrome b/f complex	<i>petA, petB, petD, petG, petL, petN</i>		
	Subunits of photosystem I	<i>psaA, psaB, psaC, psal, psaJ</i>		
	Subunit of rubisco	<i>rbcL</i>		
	Other genes	Subunit of acetyl-CoA-carboxylase	<i>accD</i>	
	C-type cytochrome synthase	<i>ccsA</i>		
	Protease	<i>clpP</i>		
	Translation initiation factor	<i>infA</i>		
	Mature enzyme	<i>matK</i>		
	Envelope membrane protein	<i>cemA</i>		
Unknown functions	Conservative open reading frame	<i>ycf1^a, ycf1^b, ycf15^a, ycf15^b, ycf2^a, ycf2^b, ycf4</i>		

Note: a: IRa region; b: IRb region.

3.4 Codon usage analysis

The codons were thought to result from a combination of natural selection, mutation, and genetic drift. Codon usage bias embodies that each gene of the species has its specific preferred amino acid codon (Subramanian et al., 2022). In this study, the number of the PCGs was 21328, 20864, and 20896 with 64 types in the three cp

genomes of *L. japonicus*, *L. cardiaca*, and *L. sibiricus*, respectively (Table S3). Among them, isoleucine was the most abundant amino acid encoded by the codons “AUU”, ranging from 880 to 898 counts with a percentage of 4.21%, compared to the least frequency amino acid of cysteine encoded by the codons “UGC”, whereas the stop codon encoded by “UAG” was the least abundant below 12 counts (0.05%). Total of these amino acids and codons, the RSCU value varied from 0.3336 in the serine (AGC) to 1.9302 in the Leucine (UUA), simultaneously, 31 codons having the RSCU > 1, of which 29 ended with A/U for the codon bias. In addition, the amino acids (codons) of Met (AUG) and Trp (UGG) do have not any codon usage bias because their RSCU values were similarly equivalent to one (Table S3 and Fig. 5). These results were consistent with most published species, for instance, the *Glycyrrhiza* species and *Agastache rugosa* (Dai et al., 2023; Liang et al., 2023).

3.5 GO and KOG analysis

Through the protein annotations from the database of GO and KOG function in the three *Leonurus* species, the results showed that diverse number functions of common proteins within the three main functions (cellular component, molecular function, and biological process, Table S4 and Figure S3) (Sangrador-Vegas et al., 2016; Mudado et al., 2006), that is to say, of which they existed at cell, membrane, and organelle including catalytic activity (37 proteins), structural molecule activity (25 proteins), transporter activity (6 proteins), binding (57 proteins), metabolic process (78 proteins), cellular process (73 proteins), response to stimulus (1, *psbA*), localization (17 proteins), biological regulation (2 proteins, *psbH* and *psbZ*), and cellular component organization or biogenesis (5 proteins, *ccsA*, *rpl23* (×2), *petN*, and *rps11*). Meanwhile, most of them had analogous functions with the homologous consensus sequence of genes and proteins, which mainly focused on the fruitful aspects, such as translation, ribosomal structure, and biogenesis; energy production and conversion; transcription; posttranslational modification, protein turnover, and chaperones; intracellular trafficking, secretion, and vesicular transport; transport and metabolism of carbohydrate, lipid, and inorganic ion (Table S5 and Figure S4).

3.6 SSR Polymorphism and repeat analysis

In the chloroplast genome of the three *Leonurus* species, there were four base types of SSR sequences (the base of A, T, AT, and TA) and three-component types (mononucleotides: p1, dinucleotide: p2, and compound: c, Table S6 and Fig. 6) (Iranawati et al., 2012). These chloroplast genomes included mononucleotides in the *L. japonicus* (the base of 8A or 13 T), the *L. cardiaca* (the base of 10A or 15 T), and the *L. sibiricus* (the base of 7A or 14T); while the dinucleotide (that is compound SSR) were found in the *L. japonicus* (the bases of 2 AT or 1 TA) and the *L. cardiaca* (the bases of 2 AT or 1 TA), added up to the number of SSR is 24, 25, and 24, respectively. The 20 SSRs were located in the LSC region of the three *Leonurus* species and 1 SSR was located in the former two species (*L. japonicus* and *L. cardiaca*, Table S6 and S7), whereas 3 SSRs in the SSC region of *L. sibiricus* (Table S7). The length of SSR sequences varied from 10 bp to 92 bp (Table S7-S10 and Fig. 6b). The four repeated sequences types (forward, reverse, complement, and palindromic) were observed in the species of *L. japonicus* (19 F, 5 R, 24 P, and 1 C, Table S7) and *L. cardiaca* (21 F, 3 R, 24 P, and 1 C, Table S7), nevertheless, the type of complement sequence was not detected in the species of *L. sibiricus* (20 F, 5 R, and 24 P, Table S7). For more detail, the number of tandem repeats found in the chloroplast genomes of *L. japonicus*, *L. cardiaca*, and *L. sibiricus* is 25, 26, and 24, respectively, which confirmed the total length of over 20 bp and the similarity $\geq 70\%$ between repeat units. The length of tandem repeats sequences varied from 29.4 bp to 86.1 bp (Tables S11-S13). The number of scattered repeats units in these three *Leonurus* species is 39, 40, and 39, respectively, and obtained using an e-value of fewer than 6.15×10^{-4} as the threshold. The length of scattered repeat sequences varied from 30 bp to 46 bp (Tables S14-S16).

3.7 Analysis of IR boundaries and comparative genomes

The illustration boundaries of the LSC, SSC, and IRs in the four cp genomes of *Leonurus* species showed the gene contents were uniform except for gene diversity of the sequence and length (Giurazza et al., 2016). Similarly, nine genes including across-the-border genes (*rps19*, *ycf1* ($\times 2$), and *ndhF*) and genes included in the regions (*rp12*, *rp12*($\times 2$), *trnH*, and *psbA*) commonly existed. The three genes of *trnH*, *rp12*, and *psbA* were located in the LSC region. The two *rp12* genes were apart and located within the IRb and IRa regions. The complete *rps19* genes were found at the border of the LSC/IRb junction. One of the two *ycf1* genes was present at the boundary between IRb and SSC region, the other one was across the IRa and SSC region, the same as the *ndhF* gene (Figure S5). After comparing the complete chloroplast genomes in the three *Leonurus* species with referred *L. japonicus* (MG673937.1), the four hotspot divergent regions were discovered to discriminate them. The regions were *trnC-GCA-petN* (A, LSC region), *petB-trnI-CAT* (B, LSC and SSC region) and *rp12-trnI-CAT* (C, IRb region), and *trnL-TAG-ndhF* (D, IRa region, Fig. 7).

3.8 Hypervariable region identification

Highly conserved regions and variable sequences in the chloroplast genomes (cpDNA) are suitable to construct DNA barcodes for effective application to interspecific discrimination and phylogenetic research (Powell et al., 1995). After calculation, 72 variable regions in the chloroplast genomes of 3 *Leonurus* species were found based on the intergenic and intron regions (NCRs) using the K2p model. The average value of the K2p in the 72 IGS regions ranged from 0 to 10.95 (Table S17 and Fig. 8). Among them, the five IGS regions (*trnC-GCA-petN*, *atpH-atpI*, *psaC-ndhE*, *rps15-ycf1*, and *petG-trnW-CCA*) showed evident variations at divergence sequences with a high K2p value. Several results showed agreement with that of the genome comparison (Fig. 7) in the main IGS regions of *trnC-GCA-petN* and *rps15-ycf1*. Therefore, these regions can be applied as potential variation regions to develop the distinction among the 3 *Leonurus* species.

3.9 Genus-specific DNA barcodes development

Given the results of hypervariable region identification (Coissac et al., 2016), we selected the nucleotide of the two IGS regions (*atpH-atpI* and *rps15-ycf1*) to design the specific primers (Table 2). Through amplifying these sequences, we develop the DNA barcodes to identify the 3 *Leonurus* species. The IGS regions (*atpH-atpI*, LSC region) was between the location of 13460 bp and 14389 bp, while IGS of *rps15-ycf1* (SSC region) was located between 121920 bp and 129540 bp in these chloroplast genomes. We cloned the two regions and acquired the three sequences of *atpH-atpI* (M1, Table 2, ~ 300 bp) and *rps15-ycf1* (M2, Table 2, ~ 300 bp) by each species using Sanger sequencing. Then, we comparatively analyzed the two molecular markers (MMs) among the three studied *Leonurus* species to determine the variations, including indels and single nucleotide polymorphisms (SNP) (Table S19, M1, and M2). The amplification products of the two IGS were checked and the strips were clearly shown on the agarose gel (Figure S6). From the peak map (up) and sequencing results (down) of the three studied *Leonurus* species with the pairs of primers related to the molecular markers M1 and M2 (Fig. 9), the marker M1 developed at the IGS *atpH-atpI* had six specific variables SNP sites (marked A, B, C, D, E, and F, Fig. 9a) and the marker M2 found at IGS *rps15-ycf1* had ten SNP sites (marked G, H, I, J, K, L, M, N, and Q, Fig. 10b) and two Indel sites (marked O and P, Fig. 9b). In detail, in the species of *L. japonicus*, the SNP site based at the positions 197/209/213 in the nucleotide sequence of YMC_M1 were G/A/A, and in the nucleotide sequence of YMC_M2, the Indel site was missing at the base at position 265 and the base at position 266 was T. In the species of *L. cardiaca*, the bases of the SNP site at positions 199/212/215 in the nucleotide sequence of OYMC_M1 were

C/A/G, while at positions 241/243/265/267 in the nucleotide sequence of OYMC_M2 were G/T/T/A. In the species of *L. sibiricus*, base positions 245/247/250/253/256/263/267 in the nucleotide sequence of XYYMC_M1 were the diverse base A/A/C/G/G/A, at position 265 in the nucleotide sequence of XYYMC_M2 was the InDel deletion site, and deletion of base T or A at the position 266 (Fig. 9). While using two labeled SNP and InDel sites for M1 and M2, the three *Leonurus* species were isolated and successfully discriminated against based on these SNP and indel loci by separately or unitedly using the two M1 and M2 molecular markers with these results.

Table 2
Primers for amplifying DNA barcodes to distinguish *Leonurus* species in the Lamiaceae family.

Number	Species	Conserved Sequences for Designing Forward Primers	Conserved Sequences for Designing Reverse Primers	IGS
M1	<i>L. japonicus</i> , <i>L. cardiaca</i> , <i>L. sibiricus</i>	CACACGTCGTAAAATTGAATCACAC	GCTTAGCGCAAAGCAATGTATGC	<i>atpH-atpI</i>
M2		CAGATATGGATTTTACCGATCAG	GCAAACAATACACAGATCAC	<i>rps15-ycf1</i>

3.10 Comparison of ITS2 sequences in the seven *Leonurus* species

In comparison with the chloroplast genome, Internal Transcribed Spacer2 (ITS2) is the non-coding region of the ribosome components in eukaryotes organisms (But et al., 2023). After annotation and contrast of the ITS2 sequences among the seven *Leonurus* species including the three studied species, we found that the most of nucleotide differences occur after the 57th base from the beginning of the sequence in a total of 221 bases except the species of *L. pseudomacranthus* showing the significant differentia, agreed with the discovery from the group of Dr. Chao and Yang (Yang et al., 2011) (Fig. 10 and Table S19). Whereas, the *Leonurus* species need to be further distinguished by the full-length ITS sequence or other distinctive ways.

3.11 Selective pressure analysis

Encoded codons can undergo evolution under a specific type of coding selective pressure by the comparative genomics approaches (Sheikh-Assadi et al., 2022). We checked the selection pressure of 80 protein-coding genes in the chloroplast genomes of three *Leonurus* species and found three genes in the three *Leonurus* species influenced by the distinct environment. Using aBSREL model implemented in the HyPhy software, we observed that the three genes were determined under positive selection, which included one Protease gene (*clpP*, *L. japonicus*), one gene for the subunits of NADH-dehydrogenase (*ndhB*, *L. cardiaca*), and one gene for the small subunit of ribosome (*rps12*, *L. sibiricus*). The optimized branch length, Likelihood Ratio Test (LRT), and p-value varied between 0.0005–0.0086, 8.3714–21.9128, and 0-0.0164, respectively. The ω_1 ratios of these genes were zero, indicating that they were under a strong purifying and weak positive selection. The branch-specific model showed that the above three genes were under positive selection in the three *Leonurus* clades along the phylogenetic tree including Clade ($\omega_2 = 371$) and Clade ($\omega_2 = 10000$). These genes in the three *Leonurus* species were probably impacted under extreme conditions, for instance, hyperthermia, high humidity, and high pressure (Table 3).

Table 3
Selection pressure analysis of the three *Leonurus* genus in the Lamiaceae family

Name	GenBank No.	B	Genes	LRT	Test p-value	Uncorrected p-value	ω distribution over sites
<i>Leonurus japonicus</i>	MG673937.1	0.0005	<i>clpP</i>	12.5857	0.0019	0.0006	$\omega_1 =$ 0.00(100%) $\omega_2 =$ 10000(0.25%)
<i>Leonurus cardiaca</i>	NC_058592.1	0.0086	<i>ndhB</i>	8.3714	0.0164	0.0055	$\omega_1 =$ 0.00(99%) $\omega_2 =$ 371(0.64%)
<i>Leonurus sibiricus</i>	OP327561.1	0.0006	<i>rps12</i>	21.9128	0.0000	0.0000	$\omega_1 =$ 0.00(97%) $\omega_2 =$ 10000(3.2%)

3.12 Phylogenetic analysis

In the phylogenetic analysis of the three *Leonurus* species in the Lamiaceae family, 80 protein sequences were extracted using the PhyloSuite software from the 16 chloroplast genomes in the study (Table S1). Among them, 64 shared nucleotide sequences of the common genes were found present in the 16 species. Using *Corydalis yanhusuo* (Papaveraceae family) and *Angelica sinensis* (Apiaceae family) as the outgroups, the phylogenetic tree was generated by the methods of maximum likelihood (ML) based on the 64 genes in these chloroplast genomes. The phylogenetic trees showed that all of the 14 species from the Lamiaceae family were clustered into a big branch with eight obvious clades (Fig. 11a). Among them, six species including the three *Leonurus* species and other two species (*Phlomoides rotata* and *Lagopsis supina*) were clustered into one great clade; the other eight species of *Phlomoides betonicoides* and *Eriophyton wallichii*, *Stachys byzantina*, *Colquhounia sequinii* and *Gomphostemma lucidum*, *Pogostemon septentrionalis*, *Holmskioldia sanguinea*, and *Gmelina hainanensis* were gathered into one branch solely in the Lamiaceae branch. Moreover, to find the affinity relationship among the more *Leonurus* species, we created the cladogram based on the ITS2 DNA sequences of the 19 evolutionary species including seven *Leonurus* species found in the NCBI database (Fig. 11b) by using the two same species as the outgroup. The results showed that the similar relationship of the three studied *Leonurus* apart were clustered into three different branches, which included the three embranchments of *L. japonicus* and *L. chaituroides*; *L. sibiricus*, and *L. pseudomacranthus*; *L. cardiaca*, *L. glaucescens*, and *L. turkestanicus*. Furthermore, the species of *Eriophyton wallichii*, *Phlomoides rotata*, *Phlomoides betonicoides*, and *Stachys byzantina* had a certain close relationship with the above three *Leonurus* branches. The specificity of *L. sibiricus* was clustered into the new sole clade in the light of cp genomes and ITS2 DNA sequences. Without a doubt, the outgroups of *Corydalis yanhusuo* (Papaveraceae family) and *Angelica sinensis* were gathered into a single branch by each and far more distantly related to the Lamiaceae species. The ML bootstrap showed strong support with bootstrap above 95% for all the nodes according to the chloroplast genomes (Fig. 11a). Thus the evolutionary

relationships are plausible to elucidate the close relationship of the *Leonurus* species in the Lamiaceae family from the ITS2 regions in the nuclear and chloroplast genomes.

4. Discussion

4.1 Gene difference of the cp genomes in the three *Leonurus* species

In the three *Leonurus* species, the ribosomal protein L16 (*rp16*) gene cannot be found in the species of *L. cardiaca*, while it commonly was considered to be the intron owned the fastest evolutionary rate which it was located in the non-coding region of the cp genome and used for the lower taxonomic order elements or among closely related groups (Downie et al., 1996). For instance, using the branch classification principle and method, based on the *rp16* sequence, the genetic diversity and genetic structure of 10 *Cardiocrinum giganteum* populations were analyzed and revealed the abundant genetic diversity and differentiation among them (ZHANG et al., 2019). Additionally, the ribosomal *rps12* gene in the chloroplast genome is the most specific gene of all chloroplast genes, consisting of two parts far apart, containing diverse exons, and encoding the ribosome small subunit S12 protein by trans-splicing (Zaita et al., 1987). In our study, the *rps12* gene consisted of three exons at both chains can be found in the cp genome of *L. sibiricus*, whereas only two exons in the *rps12* gene existed in the other two cp genomes of *L. japonicus* and *L. cardiaca*. Therefore exploring the molecular evolution pattern of the *rps12* gene is helpful to understand the evolutionary process and genomic characteristics. The evolutionary rate, selective pressure, and adaptive evolution of the *rps12* gene in a phylogenetic context of 78 gymnosperms and two ferns were analyzed (Ping et al., 2021). The results show that the *rps12* gene structure was stable during the evolution of gymnosperms and this provided more evidence that the reverse repeat region had a reduced substitution rate and revealed the evolutionary rate heterogeneity in gymnosperms. Furthermore, researchers found that the regions with the highest nucleotide diversity of the *rps12* gene can serve as potential markers for species identification and phylogeny in the Asparagaceae family (Munyao et al., 2020). As part of the 30S ribosomal subunit, the *rps12* protein plays an important role in translational accuracy together with ribosomal proteins S4 and S5 (Agarwal et al., 2015).

4.2 Variable sequences of divergence regions in chloroplast genomes of the three *Leonurus* species

Nucleotide diversity is commonly used to measure diversity within or between populations, or genetic variation among related species (Hua et al., 2023). Alternatively, nucleotide diversity can also be used to infer evolutionary relationships. From the results, the *Leonurus* species are highly conserved in gene number, sequence, and orientation, however, some mutation sites still were found. The four genes *ycf3*, *rps11*, *ndhD*, and *ndhA* showed variants in the sequences of nucleotide. For the *ycf3* gene with three exons and two introns, its splicing of intron 1, which has the function of ribozyme shear, was specifically required by the pentatricopeptide repeat protein, photosystem I biogenesis factor 2 (PBF2) through cooperating with CAF1 and CAF2 (Wang et al., 2020). Studies have shown that *rps11* and *ndhD* have the higher variation reported found in the chloroplast genomes of 16 *Clematis* species (WEI et al., 2022). The *rps11* gene is a component of the small ribosomal subunit 40S, which belongs to the S17p family of ribosomal proteins and is encoded by the *rps11* gene, and is predominantly present in eukaryotes (Stevenson et al., 1991). The chloroplast sequence of the *ndhD* gene was analyzed in the restriction site map of the *ndhD* gene region, and the occurrence of cytoplasmic male sterility in *Sorghum bicolor* was

explored (Fan et al., 2002). Similarly, the variable sequence of the *ndhD* gene can be used in the comparison of gene sequences in the chloroplast genome of *Betula platyphylla* (Wang et al., 2018). The mention of the *ndhA* gene and its introns can be used as DNA barcodes for the identification of 18 *Begonia* species and *polygonum* species (Wang et al., 2022). Regarding the above divergence regions usually can be used for the identification of the species and its adulterants, such as the IGS sequences of *atpH-atpI*, *ndhC-trnM*, *petA-psbJ*, and *rps15-ycf1* for the distinction of *Isodon rubescens*, *Musa*, and *Amana* species (Zhou et al., 2022; Song et al., 2022; Li et al., 2017); the IGS sequences of *ndhF-rpl32* and *trnC-GCA-petN* also applied for the discrimination of Nine *Musa* Species, *Polygonatum*, and Tribe *Polygonateae* (Wang et al., 2022; Song et al., 2022), respectively. Compared with IR regions, LSC and SSC regions had more variant sites, and the variants were mainly concentrated in non-coding regions. In general, different positional variation is not consistent in the cp genome and the hypermutation sections are not the same in different plant groups (Samuk et al., 2023).

4.3 Identification of *Leonurus* species by using the ITS2 sequences

The "DNA barcode" is based on a standardized short sequence of DNA within a small region of one species' genome, which can quickly and accurately distinguish diverse species (Antil et al., 2023). As a representative marker, the region of the internal transcriptional spacer 2 (ITS2) locates at the position of nuclear RNA cistron rapid evolution, is not only very different in the sequence, but also has the characteristics of easy amplification and strong generality, therefore it is very suitable reference as the "DNA barcode" (Schultz et al., 2005). For medicinal plants, ITS2 has been widely used as a highly efficient and DNA-based marker for species identification among the species or subspecies level in the Chinese Pharmacopoeia. For instance, the specific ITS2 sequences have become one part of the drug standard in the *Fritillaria cirrhosa* and *Dendrobium nobile* cloned by the polymorphism method of polymerase chain reaction-restriction enzyme length using the designated primer pair from the Chinese Pharmacopoeia (2020 edition, 3). In this study, the studied three *Leonurus* species and *L. pseudomacranthus* can be easily identified because of each own special ITS2 DNA sequence in the divergence positions. Nevertheless, the other three species (*L. glaucescens*, *L. chaituroides*, and *L. turkestanicus*) with identical sequences to the above two species of *L. cardiaca* and *L. japonicus* were required to be identified with the help of other methods, such as appearance or prescribed molecular fragments. Thus these ITS2 sequence diversity could provide direct proof that they can be certainly used to distinguish the plant resource of the three *Leonurus* species, also unavoidably becoming the basis for the differentiation between species and the identification of genetic traits.

4.4 Evolutionary comparison of the three *Leonurus* species with the before studies

The *Leonurus* species have a long historically medicinal value to gynecological diseases through the effect on promoting blood circulation, removing blood stasis, regulating menstruation, and relieving pain. Regarding the analysis of phylogenetic location and genetic relationship among the *Leonurus* species, the evolutionary tree or phylogenetic tree can be constructed according to protein sequences or structural difference relationships. The group of Dr. Liu found that the phylogenetic position of *L. japonicus* was in the Lamioideae and had a close genetic relationship between the *Leonurus* genera and *Stachys* genera based on chloroplast genomes (ZHANG et al., 2018), which had a certain related to the quality control of stachydrine hydrochloride and Leonurine hydrochloride extracted from the species of *L. japonicus*. The 22 accessions of *L. cardiaca* from different geographical locations worldwide formed four clades from different regions of the US (11 locations), China (1

location), US and European (1 plus 5 locations), and the US (4 locations) in the account of the chloroplast genomes (Sun et al., 2021), which illustrated the sequence polymorphism influenced the evolutionary relationship because of differences in geographic location. Dr. Yang et al. have found that the diverse original of *L. sibiricus*, *L. chaituroides*, *L. japonicus*, and *L. pseudomacranthus* was clustered into one branch (Clade I), while the other two species of *L. turkestanicus* and *L. glaucescens* were gathered into another branch (Clade II) based on the total sequences of ITS region including internal transcriptional spacer 2 (ITS1), 5.8S ribosomal RNA, and internal transcriptional spacer 2 (ITS2) (Yang et al., 2011; Yang et al., 2006). Nevertheless, the phylogenetic relationship in this study, the clustering development of four species (Clade I) above results was similarly compliant with the constructed tree given ITS2 DNA sequences, while the species of *L. cardiaca* newly supplemented showed the closely related to the above species of the Clade II. The separate branch of *L. sibiricus* illustrated its characteristics and distinctiveness. Therefore, in the evolutionary process, the seven *Leonurus* species kept the congruously coincident evolutionary relationship no matter what cp genomes, ITS, or ITS2 sequences were.

5. Conclusion

The three studied *Leonurus* species are one of the essential and medicinal genera in the Lamiaceae family. Especially, the species of *L. japonicus* is a vital traditional Chinese medicine (TCM) medicine included in Chinese pharmacopeia for the treatment of gynecological diseases. In the ripe stage of the three plants, they can be recognized from the outer features of diverse tissues. After comparing the sequences of the chloroplast genomes, the results showed that the gene size, content, and order of chloroplast genomes were all similar and had excellent collinearity. Though amplification of the two diversity regions at *atpH-atpI* and *rps15-ycf1* selected from the eleven hotspot divergent regions, the sixteen specific variable SNP sites and two Indel sites were identified for useful genetic diversity. Through sequence contrast of ITS2 DNA, the studied three *Leonurus* species possessed their specific contents to be significantly distinguished. Additionally, the phylogenetic analysis confirmed the previous evolutionary phylogeny, in which the three *Leonurus* species were gathered into the same great clade although they were located at the different sub-branch based on the shared CDS nucleotide or the ITS2 DNA sequences in the nuclear genome. Generally, this study provides valuable genetic information on *Leonurus* species which can aid in further species identification, planting resources breeding, quality evaluation, and evolutionary relationships.

Declarations

Accession numbers and data availability

Previously available sequencing data of chloroplast genomes from the studied 16 species having the related accession numbers are detailed in Table S1. Data will be made available on request.

CRediT authorship contribution statement

Qing Du: Formal analysis, Investigation, Writing-original draft, Writing-review & editing, Visualization. Ziyi Rong and Yanjuan Guan: Formal analysis, Investigation, Writing, editing, Visualization. Chang Zhang: Formal analysis, Writing-review & editing, Visualization. Liqiang Wang: Formal analysis, Writing-review & editing, Visualization. Bing Wang: review & editing, Funding acquisition. Haidong Gao: Formal analysis, Writing-review & editing, Visualization. Chang Liu: Conceptualization, Writing-review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by funds from the National Science & Technology Fundamental Resources Investigation Program of China [2018FY100705], Chinese Academy of Medical Sciences, Innovation Funds for Medical Sciences (CIFMS) [2021-I2M-1-022], National Science Foundation [81872966], Qinghai Provincial Key Laboratory of Phytochemistry of Qinghai Tibet Plateau [2022-ZJ-Y14], Hunan technological innovation guidance project (2018SK52001). We would like to thank Dr. Mei Jiang, Dr. Yang Ni, and Dr. Jinglin Li who have provided support for data analysis.

References

1. Agarwal D, Kamath D, Gregory ST. and O'Connor M. (2015) Modulation of decoding fidelity by ribosomal proteins S4 and S5. *J Bacteriol.* 197, 1017-1025.
2. Amiryousefi A, Hyvönen J. and Poczai P. (2018) IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics.* 34, 3030-3031.
3. Angeloni S, Spinozzi E, Maggi F, Sagratini G, Caprioli G, Borsetta G, Ak G, Sinan KI, Zengin G, Arpini S, Mombelli G. and Ricciutelli M. (2021) Phytochemical Profile and Biological Activities of Crude and Purified *Leonurus cardiaca* Extracts. *Plants (Basel).* 10, 195.
4. Antil S, Abraham JS, Sripoorna S, Maurya S, Dagar J, Makhija S, Bhagat P, Gupta R, Sood U, Lal R. and Toteja R. (2023) DNA barcoding, an effective tool for species identification: a review. *Mol Biol Rep.* 50, 761-775.
5. Arias RS, Ballard LL. and Scheffler BE. (2009) UPIC: Perl scripts to determine the number of SSR markers to run. *Bioinformatics.* 3, 352-360.
6. Asad N, Smith E, Shakya S, Stegman S. and Timmons L. (2023) Sustainable Methodologies for Efficient Gel Electrophoresis and Streamlined Screening of Difficult Plasmids. *Methods Protoc.* 6, 25.
7. Beier S, Thiel T, Münch T. and Scholz U, Mascher M (2017) MISA-web: a web server for microsatellite prediction. *Bioinformatics,* 33, 2583–2585.
8. Brenner S. (1966) Collinearity and the genetic code. *Proc R Soc Lond B Biol Sci.* 164, 170-180.
9. Subramanian K, Payne B, Feyertag F. and Alvarez-Ponce D. (2022) The Codon Statistics Database: A Database of Codon Usage Bias. *Mol Biol Evol.* 39, msac157.
10. Buch G, Schulz A, Schmidtman I, Strauch K. and Wild PS. (2023) A systematic review and evaluation of statistical methods for group variable selection. *Stat Med.* 42, 331-352.
11. But GW, Wu HY, Siu TY, Chan KT, Wong KH, Lau DT. and Shaw PC. (2023) Comparison of DNA extraction methods on CITES-listed timber species and application in species authentication of commercial products using DNA barcoding. *Sci Rep.* 13, 151.
12. Butler C, Matsumoto A, Rutherford C. and Lima HK (2022). Comparison of the Effectiveness of Four Commercial DNA Extraction Kits on Fresh and Frozen Human Milk Samples. *Methods Protoc.* 5, 63.
13. *Cardiaca* Lam.(1778) in Lam. et DC. *Fl. France:* 38.
14. Chen S., Yao H., Han J., Liu C., Song J., Shi L., Zhu Y., Ma X., Gao T., Pang X., Luo K., Li Y., Li X., Jia X., Lin Y. and Leon C. (2010) Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant

- species. PLoS One 5(1), e8613.
15. Cittaro D, Lazarevic D. and Provero P. (2016) Chromas from chromatin: sonification of the epigenome. *F1000Res.* 5, 274.
 16. Coissac E, Hollingsworth PM, Lavergne S. and Taberlet P. (2016) From barcodes to genomes: extending the concept of DNA barcoding. *Mol Ecol.* 25, 1423-1428.
 17. Cui N, Liao BS, Liang CL, Li SF, Zhang H, Xu J, Li XW. and Chen SL. (2020) Complete chloroplast genome of *Salvia plebeia*: organization, specific barcode and phylogenetic analysis. *Chin J Nat Med.* 18, 563-572.
 18. DAI Jiang-peng, CAI Yi-ming, LIU Qiao-zhen, GAO Xiao-xia. and ZHU Shuang. (2023) Analysis of codon usage bias of chloroplast genome in seven *Glycyrrhiza* species. *Chinese Traditional and Herbal Drugs.* 740, 2907-2916.
 19. Darling AE, Mau B. and Perna NT. (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One.* 5, e11147.
 20. Downie S R. and Katz-Downie D S. (1996) A molecular phylogeny of Apiaceae subfamily Apioideae: evidence from ribosomal DNA internal transcribed spacer sequence. *American Journal of Botany,* 83, 234-251.
 21. Du Q, Yang H, Zeng J, Chen Z, Zhou J, Sun S, Wang B. and Liu C. (2022) Comparative Genomics and Phylogenetic Analysis of the Chloroplast Genomes in Three Medicinal *Salvia* Species for Bioexploration. *Int J Mol Sci.* 23, 12080.
 22. Fan CF, Sun CY, Guo XC, Zhang FY, Sun Y, Niu TT. and Jia JF. (2002) Sequence variation of the chloroplast gene *ndhD* region in cytoplasmic male sterile sorghum. *Yi Chuan Xue Bao.* 29, 907-914.
 23. Giurazza F, Niola R, Albano G, Nazzaro G, Silvestre M, Sirimarco F. and Maglione F. (2016) Reply to: "Endovascular Management of Abnormal Placental Implantation Deliveries: Expanding IR Boundaries". *J Vasc Interv Radiol.* 27, 1103-1104.
 24. Hua Z, Jiang C, Song S, Tian D, Chen Z, Jin Y, Zhao Y, Zhou J, Zhang Z, Huang L. and Yuan Y. (2023) Accurate identification of taxon-specific molecular markers in plants based on DNA signature sequence. *Mol Ecol Resour.* 23, 106-117.
 25. Iranawati F, Jung H, Chand V, Hurwood DA. and Mather PB. (2012) Analysis of genome survey sequences and SSR marker development for Siamese Mud Carp, *Henicorhynchus siamensis*, using 454 pyrosequencing. *Int J Mol Sci.* 13, 10807-10827.
 26. Janzen T, Bokma F. and Etienne RS. (2022) Nucleotide Substitutions during Speciation may Explain Substitution Rate Variation. *Syst Biol.* 71, 1244-1254.
 27. Jingyao Ping, Peipei Feng, Jing Hao, Jinye Li, Yingjuan Su. and Ting Wang. (2021) The molecular evolution pattern of *rps12* gene in gymnosperms. *Chinese Science Bulletin,* 66, 3182-3193.
 28. Katoh K, Rozewicki J. and Yamada KD. (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* 20, 1160-1166.
 29. Khadivi-Khub A. and Soorni A. (2014) Comprehensive genetic discrimination of *Leonurus cardiaca* populations by AFLP, ISSR, RAPD and IRAP molecular markers. *Mol Biol Rep.* 41, 4007-4016.
 30. Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, Shank SD, Magalis BR, Bouvier D, Nekrutenko A, Wisotsky S, Spielman SJ, Frost SDW. and Muse SV. (2020) HyPhy 2.5-A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. *Mol Biol Evol.* 37, 295-299.

31. Lanave C, Licciulli F, De Robertis M, Marolla A. and Attimonelli M. (2002) Update of AMmtDB: a database of multi-aligned Metazoa mitochondrial DNA sequences. *Nucleic Acids Res.* 30, 174-175.
32. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ. And Higgins DG. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, 23, 2947-2948.
33. Li H, Guo Q, Xu L, Gao H, Liu L. and Zhou X. (2023) CPJSdraw: analysis and visualization of junction sites of chloroplast genomes. *PeerJ.* 11, e15326.
34. Li P, Lu RS, Xu WQ, Ohi-Toma T, Cai MQ, Qiu YX, Cameron KM. and Fu CX. (2017) Comparative Genomics and Phylogenomics of East Asian Tulips (*Amana*, Liliaceae). *Front Plant Sci.* 8, 451.
35. Liang Xianglan, Qin Yiming, Sun Xiaobo. and Guo Song. (2023) Analysis of Codon Usage Bias in the Chloroplast Genome of *Agastache rugosa*. *Molecular Plant Breeding*, 21, 809-818.
36. Linn. (1754) *Gen. Pl.* 254.
37. Liu S, Ni Y, Li J, Zhang X, Yang H, Chen H. and Liu C. (2023) CPGView: A package for visualizing detailed chloroplast genome structures. *Mol Ecol Resour.* 23, 694-704.
38. Mayor C., Brudno M., Schwartz J. R., Poliakov A., Rubin E. M., Frazer K. A., Pachter L. S. and Dubchak I. (2000) VISTA: Visualizing Global DNA Sequence Alignments of Arbitrary Length. *Bioinformatics*, 16,1046.
39. McLean D. (2002) Adobe Photoshop and Illustrator techniques. *J Audiov Media Med.* 25, 79-81.
40. Mohanty JN, Sahoo S. and Mishra P. (2022) NBLAST: a graphical user interface-based two-way BLAST software with a dot plot viewer. *Genomics Inform.* 20, e40.
41. Mohapatra A. and Tripathy BC. (2007) Differential distribution of chlorophyll biosynthetic intermediates in stroma, envelope and thylakoid membranes in *Beta vulgaris*. *Photosynth Res.* 94, 401-410.
42. Mudado Mde A. and Ortega JM. (2006) A picture of gene sampling/expression in model organisms using ESTs and KOG proteins. *Genet Mol Res.* 5, 242-253.
43. Munyao JN, Dong X, Yang JX, Mbandi EM, Wanga VO, Oulo MA, Saina JK, Musili PM. And Hu GW. (2020) Complete Chloroplast Genomes of *Chlorophytum comosum* and *Chlorophytum gallabatense*: Genome Structures, Comparative and Phylogenetic Analysis. *Plants (Basel).* 9, 296.
44. National Commission of Chinese Pharmacopoeia. (2015) *Pharmacopoeia of the People's Republic of China*. Beijing: China Medical Science Press.
45. Nguyen LT, Schmidt HA, von Haeseler A and Minh BQ. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32, 268-274.
46. Nishii K, Möller M, Foster RG, Forrest LL, Kelso N, Barber S, Howard C. and Hart ML. (2023) A high quality, high molecular weight DNA extraction method for PacBio HiFi genome sequencing of recalcitrant plants. *Plant Methods.* 19, 41.
47. Oliveira AS, Cercato LM, de Santana Souza MT, Melo AJO, Lima BDS, Duarte MC, Araujo AAS, de Oliveira E Silva AM. and Camargo EA. (2017) The ethanol extract of *Leonurus sibiricus* L. induces antioxidant, antinociceptive and topical anti-inflammatory effects. *J Ethnopharmacol.* 206, 144-151.
48. Pitschmann A, Waschulin C, Sykora C, Purevsuren S, Glasl S. (2017) Microscopic and Phytochemical Comparison of the Three *Leonurus* Species *L. cardiaca*, *L. japonicus*, and *L. sibiricus*. *Planta Med.* 83(14-15):1233-1241.

49. Powell W, Morgante M, Andre C, McNicol JW, Machray GC, Doyle JJ, Tingey SV. and Rafalski JA. (1995) Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. *Curr Biol.* 5, 1023-1029.
50. Racg G, et al. (1961) *Acta Bot Horti Bacurestiensis*, 62, 525.
51. Samuk K. (2023) Average nucleotide diversity should be weighted by per-site sample size. *Mol Ecol Resour.* 23, 355-358.
52. Sangrador-Vegas A, Mitchell AL, Chang HY, Yong SY. and Finn RD. (2016) GO annotation in InterPro: why stability does not indicate accuracy in a sea of changing annotations. *Database (Oxford)*. 2016, baw027.
53. Sarachu M, Colet M. (2005) wEMBOSS: a web interface for EMBOSS. *Bioinformatics*, 21, 540-541.
54. Schultz J, Maisel S, Gerlach D, Müller T. and Wolf M. (2005) A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *RNA*. 11, 361-364.
55. Sheikh-Assadi M, Naderi R, Kafi M, Fatahi R, Salami SA. and Shariati V. (2022) Complete chloroplast genome of *Lilium ledebourii* (Baker) Boiss and its comparative analysis: lights into selective pressure and adaptive evolution. *Sci Rep.* 12, 9375.
56. Song W, Ji C, Chen Z, Cai H, Wu X, Shi C. and Wang S. Comparative Analysis the Complete Chloroplast Genomes of Nine *Musa* Species: Genomic Features, Comparative Analysis, and Phylogenetic Implications. *Front Plant Sci.* 13, 832884.
57. Stevenson JK, Drager RG, Copertino DW, Christopher DA, Jenkins KP, Yepiz-Plascencia G. and Hallick RB. (1991) Intercistronic group III introns in polycistronic ribosomal protein operons of chloroplasts. *Mol Gen Genet.* 228, 183-192.
58. Subramanian K, Payne B, Feyertag F. and Alvarez-Ponce D. (2022) The Codon Statistics Database: A Database of Codon Usage Bias. *Mol Biol Evol.* 39, msac157.
59. Sun J, Wang Y, Garran TA, Qiao P, Wang M, Yuan Q, Guo L. and Huang L. (2021) Heterogeneous Genetic Diversity Estimation of a Promising Domestication Medicinal Motherwort *Leonurus Cardiaca* Based on Chloroplast Genome Resources. *Front Genet.* 12, 721022.
60. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S. (2011) *Molecular Biology and Evolution*, 28:2731-2739.
61. Tang YL, Wu YS, Huang RS, Chao NX, Liu Y, Xu P, Li KZ, Cai DZ. and Luo Y. (2016) Molecular identification of *Uncaria* (Gouteng) through DNA barcoding. *Chin Med.* 11, 3.
62. Wang D, Zhang Y, Zhang Z, Zhu J. and Yu J. (2010) KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics*, 8, 77-80.
63. Wang J, Qian J, Jiang Y, Chen X, Zheng B, Chen S, Yang F, Xu Z. and Duan B. (2022) Comparative Analysis of Chloroplast Genome and New Insights Into Phylogenetic Relationships of *Polygonatum* and Tribe Polygonateae. *Front Plant Sci.* 13, 882189.
64. Wang S, Yang C, Zhao X, Chen S. and Qu GZ. (2018) Complete chloroplast genome sequence of *Betula platyphylla*: gene organization, RNA editing, and comparative and phylogenetic analyses. *BMC Genomics*, 19, 950.
65. Wang X, Yang Z, Zhang Y, Zhou W, Zhang A. and Lu C. (2020) Pentatricopeptide repeat protein PHOTOSYSTEM I BIOGENESIS FACTOR2 is required for splicing of *ycf3*. *J Integr Plant Biol.* 62, 1741-1761.

66. Wascher M. and Kubatko L. (2021) Consistency of SVDQuartets and Maximum Likelihood for Coalescent-Based Species Tree Estimation. *Syst Biol.* 70, 33-48.
67. WEI Hai-zhong, PAN Li-qin, TIAN Sheng-ye, TANG Zi-yi, HE Hai-ye, ZHANG Hui-juan. and JIANG Ming. (2022) Chloroplast genome sequence characterization and phylogenetic analysis of *Clematis henryi*. *Chinese Traditional and Herbal Drugs*, 53, 3766-3773.
68. Yang L, Li J. and Zhou G. (2022) Comparative chloroplast genome analyses of 23 species in *Swertia L.* (Gentianaceae) with implications for its phylogeny. *Front Genet.* 13, 895146.
69. Yang ZY, Chao Z, Huo KK, Wu BY. and Pan SL. (2006) Nuclear ribosomal DNA internal transcribed spacer 1 sequences of 4 *Leonurus* species. *Nan Fang Yi Ke Da Xue Xue Bao.* 26, 1593-1595.
70. Yang ZY, Pan SL, Huo KK, Wu BY. and Chao Z. (2011) Molecular analysis of *Leonurus* species in China based on ITS and matK sequences. *Am J Chin Med.* 39, 411-422.
71. Zaita N, Torazawa K, Shinozaki K. and Sugiura M. (1987) Trans splicing in vivo: Joining of transcripts from the 'divided' gene for ribosomal protein S12 in the chloroplasts of tobacco. *FEBS Lett.* 210, 153-156.
72. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX. and Wang GT. (2020) PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour.* 20, 348-355.
73. ZHANG Hui , HE Shuaibing, KONG Fande, CHEN Haimei, TANG Taishan, XU Shufei , MIAO Li. and LIU Chang. (2018) Sequence of Chloroplast Genome and the Phyletic Evolution within *Leonurus artemisia*. *Information on Traditional Chinese Medicine*, 35, 21-27.
74. ZHANG Xiao-yun, YANG Miao-qin, XU Ying, ZHANG Yong-hong. and LI Hong-fei. (2019) Study on the Genetic Diversity and Genetic Structure of *Cardiocrinum giganteum* Based on rpl16 Sequence Analysis. *Journal of Plant Genetic Resources*, 20, 199-206.
75. Zhao ZH, Yao ZH, Lin SJ, Chu G, Mu KQ, Wang Y, Bi KS, Wang TJ, Li Q. and Liu R. (2022) *Leonurus japonicus* Houtt. (Motherwort): Systematic research through chemical profiling, stability under controlled conditions and pharmacokinetic analysis on screening Q-markers for quality control. *J Pharm Biomed Anal.* 213, 14707.
76. Zhou Z, Wang J, Pu T, Dong J, Guan Q, Qian J, Shi L. and Duan B. (2022) Comparative analysis of medicinal plant *Isodon rubescens* and its common adulterants based on chloroplast genome sequencing. *Front Plant Sci.* 13, 1036277.

Supplementary Tables

Supplementary Tables S1-S19 are not available with this version.

Table S1. Gene contents and characteristics of 16 species' chloroplast genomes in the study.

Table S2. The lengths of introns and exons for the splitting genes in the chloroplast genomes of the three *Leonurus* species.

Table S3. The type, number, and RSCU value of codon in the three *Leonurus* species cp genomes.

Table S4. The type and number of gene functions annotated in the database of Gene Oncology.

Table S5. The type and number of gene and protein functions annotated in the database of Gene Oncology.

Table S6. The SSRs types and repeat length in the chloroplast genomes of the three *Leonurus* species.

Table S7. SSR features and tandem repetition sequences in the chloroplast genomes of the three *Leonurus* species.

Table S8. SSR characteristic analysis in the chloroplast genome of *Leonurus japonicus*.

Table S9. SSR characteristic analysis in the chloroplast genome of *Leonurus cardiaca*.

Table S10. SSR characteristic analysis in the chloroplast genome of *Leonurus sibiricus*.

Table S11. The Statistics of tandem repeats identified in the chloroplast genome of *Leonurus japonica*.

Table S12. The Statistics of tandem repeats identified in the chloroplast genome of *Leonurus cardiaca*.

Table S13. The Statistics of tandem repeats identified in the chloroplast genome of *Leonurus sibiricus*.

Table S14. The Scattered repeats unit identified in chloroplast genome of *Leonurus japonica*.

Table S15. The Scattered repeats unit identified in chloroplast genome of *Leonurus cardiaca*.

Table S16. The Scattered repeats unit identified in chloroplast genome of *Leonurus sibiricus*.

Table S17. K2p distances among the intergenic spacer regions from the four chloroplast genomes of *Leonurus* species.

Table S18. Sequence results of molecular markers M1 and M2 in the chloroplast genomes of the three *Leonurus* species.

Table S19. Sequence results of ITS2 DNA in the seven *Leonurus* species.

Figures

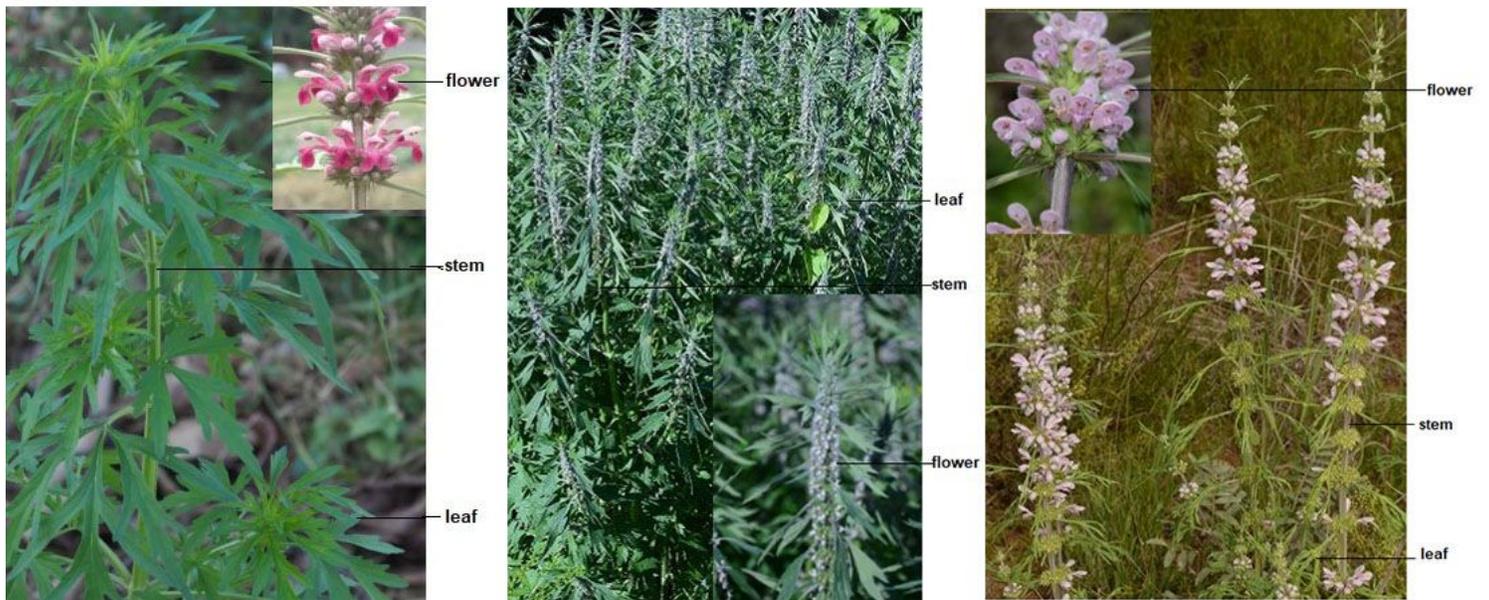


Figure 1

The morphologic characteristics of *Leonurus japonicus* (a), *Leonurus sibiricus* (b), and *Leonurus cardiaca* (c). The characteristics differed in the shape of stem, leaf and flower has been marked in the figure.

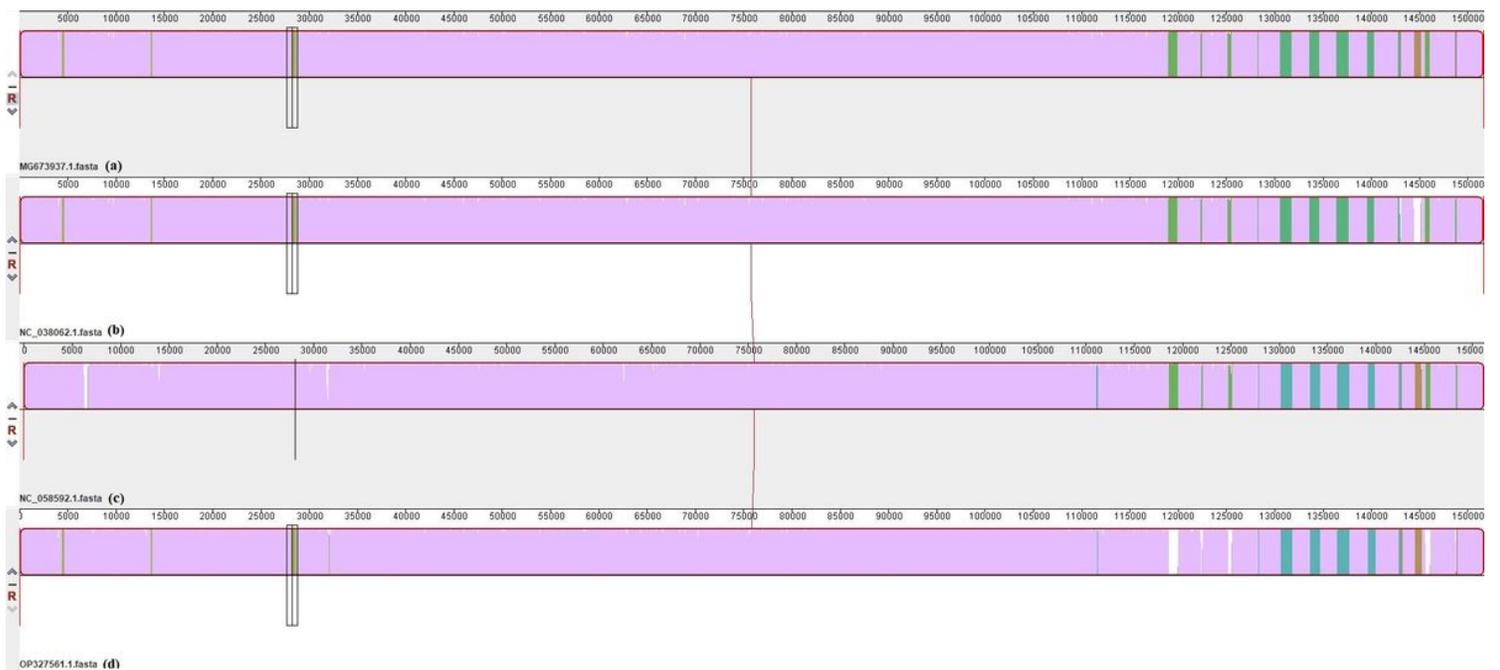
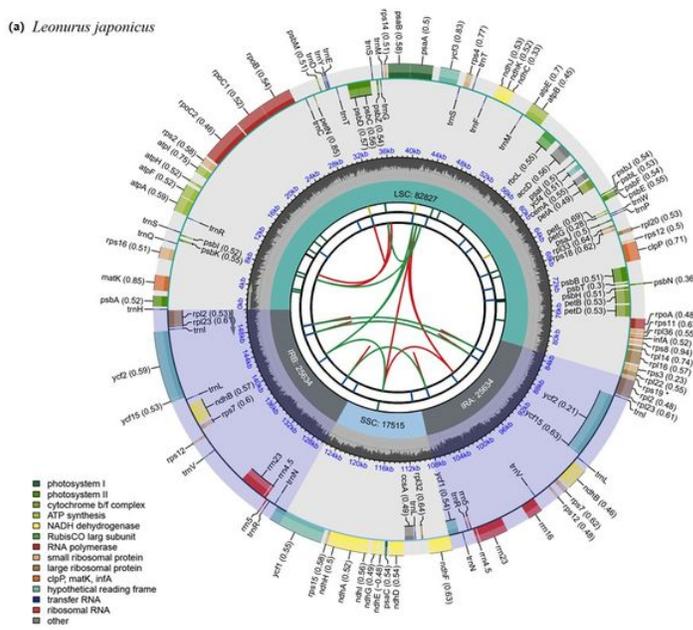


Figure 2

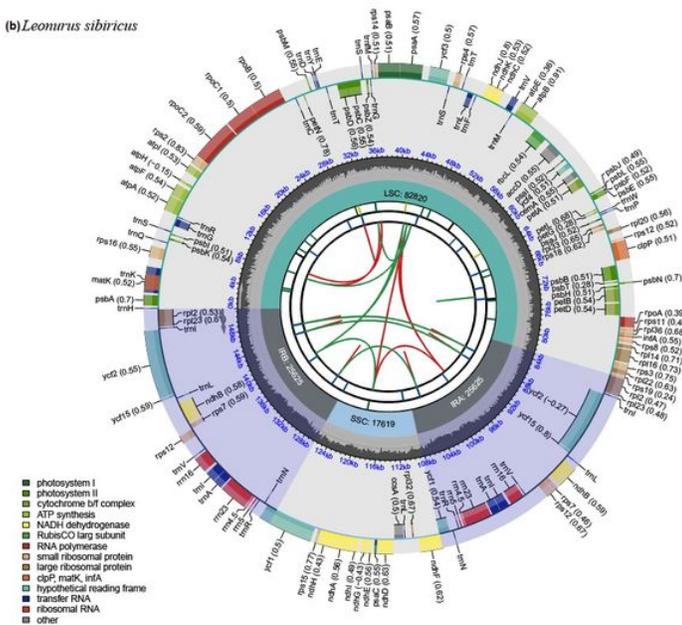
The Collinearity of the chloroplast genomes in the four *Leonurus* species.

L. japonicus (a: MG673937.1 and b: NC_038062.1), *L. cardiaca* (c: NC_058592.1), and *L. sibiricus* (d: OP327561.1) from up to down in order.

(a) *Leonurus japonicus*



(b) *Leonurus sibiricus*



(c) *Leonurus cardiaca*

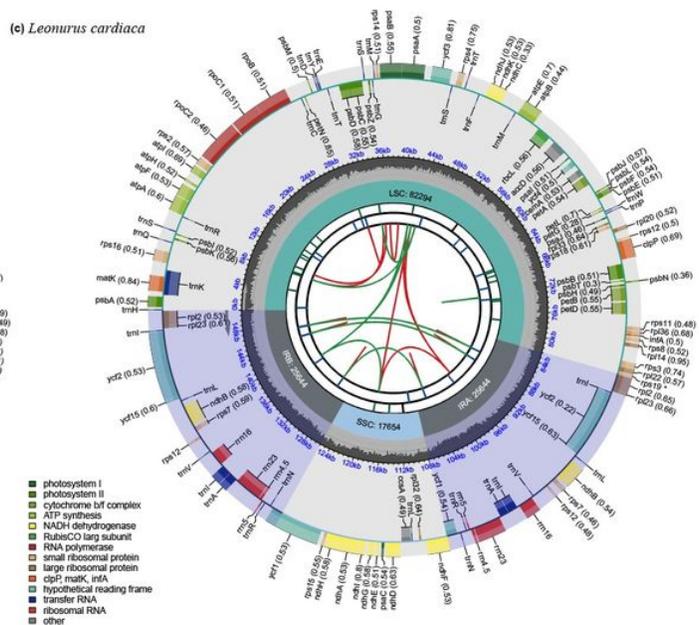


Figure 3

Graphic representation of features identified in the three chloroplast genomes of *Leonurus* species by using CPGView (<http://www.1kmpg.cn/cpgview>). The map contains seven tracks. From the center going outward, the first track shows the distributed repeats connected with red (the forward direction) and green (the reverse direction) arcs. The next track shows the tandem repeats marked with short bars. The third track shows the microsatellite sequences as short bars. The fourth track shows the size of the LSC and SSC. The fifth track shows the IRA and IRB. The sixth track shows the GC contents along the chloroplast genome. The seventh track shows the genes having different colors based on their functional groups.

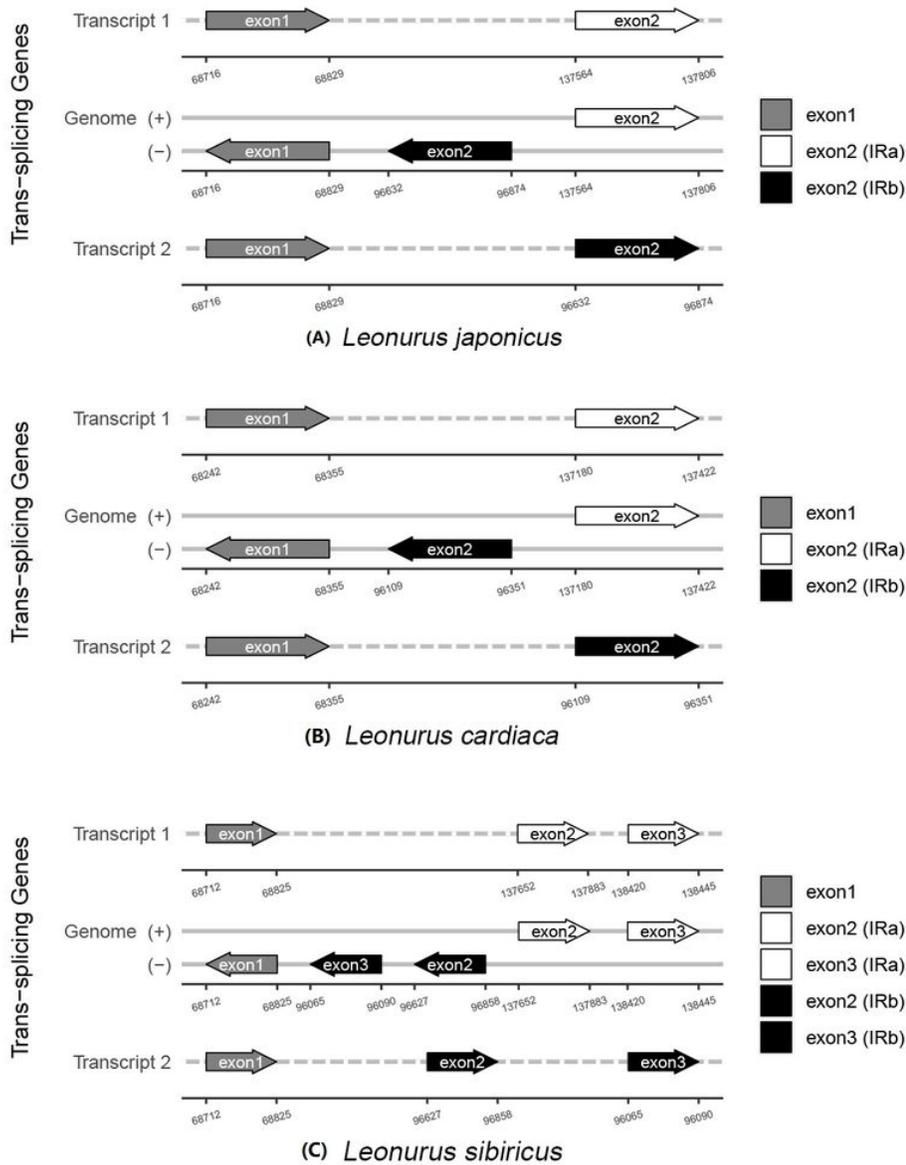


Figure 4

The trans-splicing genes structures in the *Leonurus japonicus* chloroplast genomes

(<http://www.1kmpg.cn/cpgview>). The white area is exon 2 in IRa, the black area is exon2 in IRb and the grey area is exon 1. The arrow shows the sense direction of the genes.

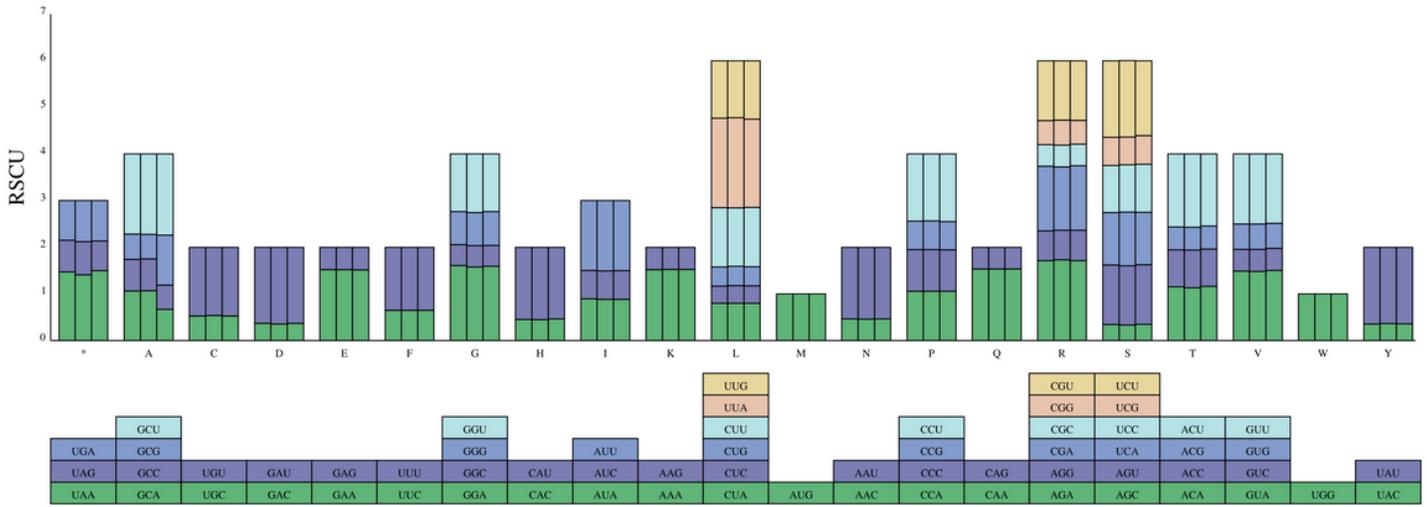


Figure 5

The codon usage bias in all protein-coding genes of chloroplast genomes in the three *Leonurus* species from left to right: *Leonurus japonicus*, *Leonurus sibiricus*, and *Leonurus cardiaca*. The different RSCU values has been shown in the ordinate.

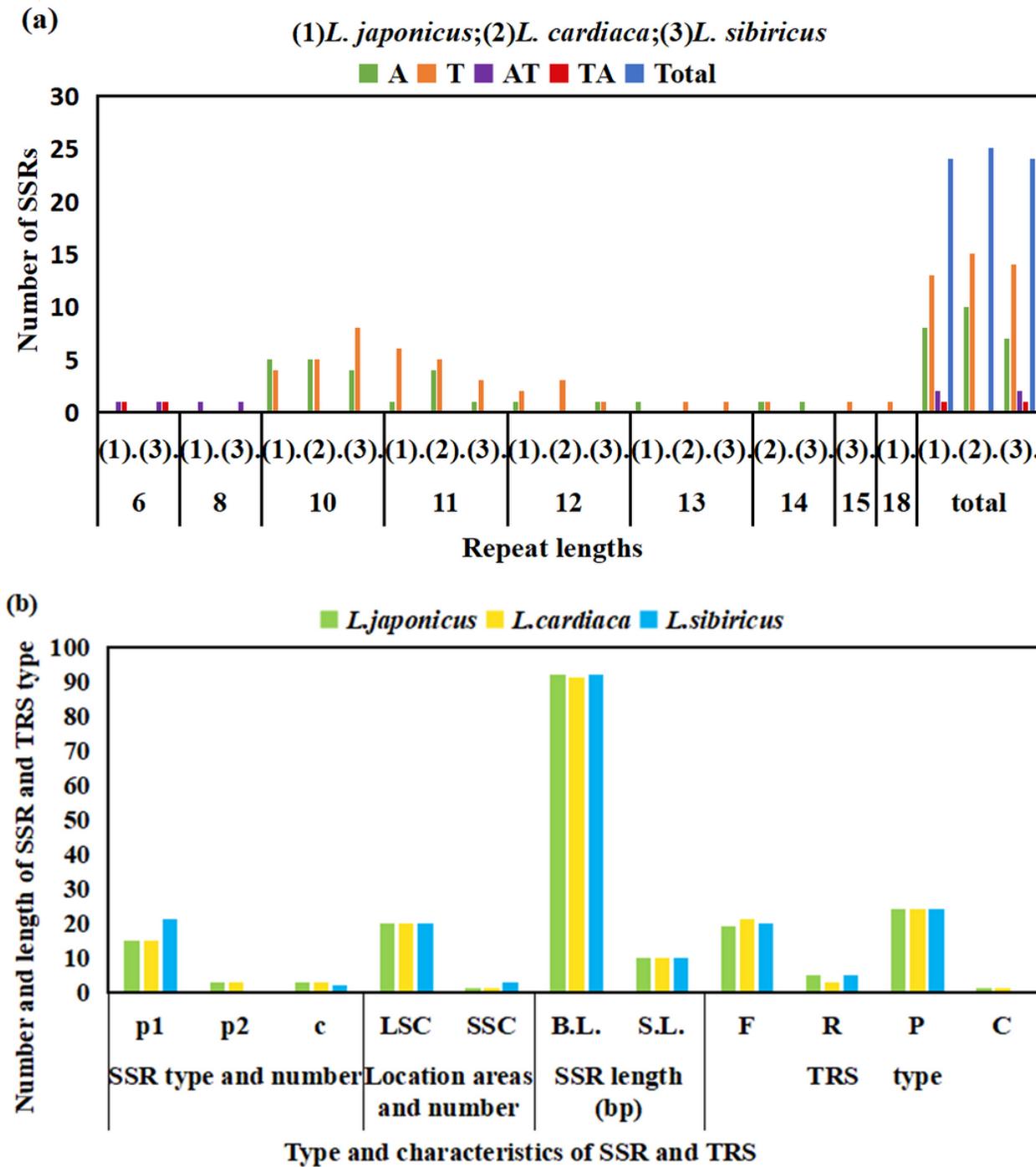


Figure 6

Analysis of SSR and TRS repeats in the chloroplast genomes of the three *Leonurus* species. *Leonurus japonicus* (1), *Leonurus sibiricus* (2), and *Leonurus cardiaca* (3). (a) Repeat lengths (abscissa) and SSR number (ordinate) in the three *Leonurus* species were shown with the different colors of the mononucleotides (green and orange), dinucleotide (purple and red), and total nucleotide (blue). (b) The type and characteristics (abscissa), number, and length (ordinate) of SSR and TRS were shown with the diverse colors of *L. japonicus* (green), *L. cardiaca* (yellow), and *L. sibiricus* (blue).

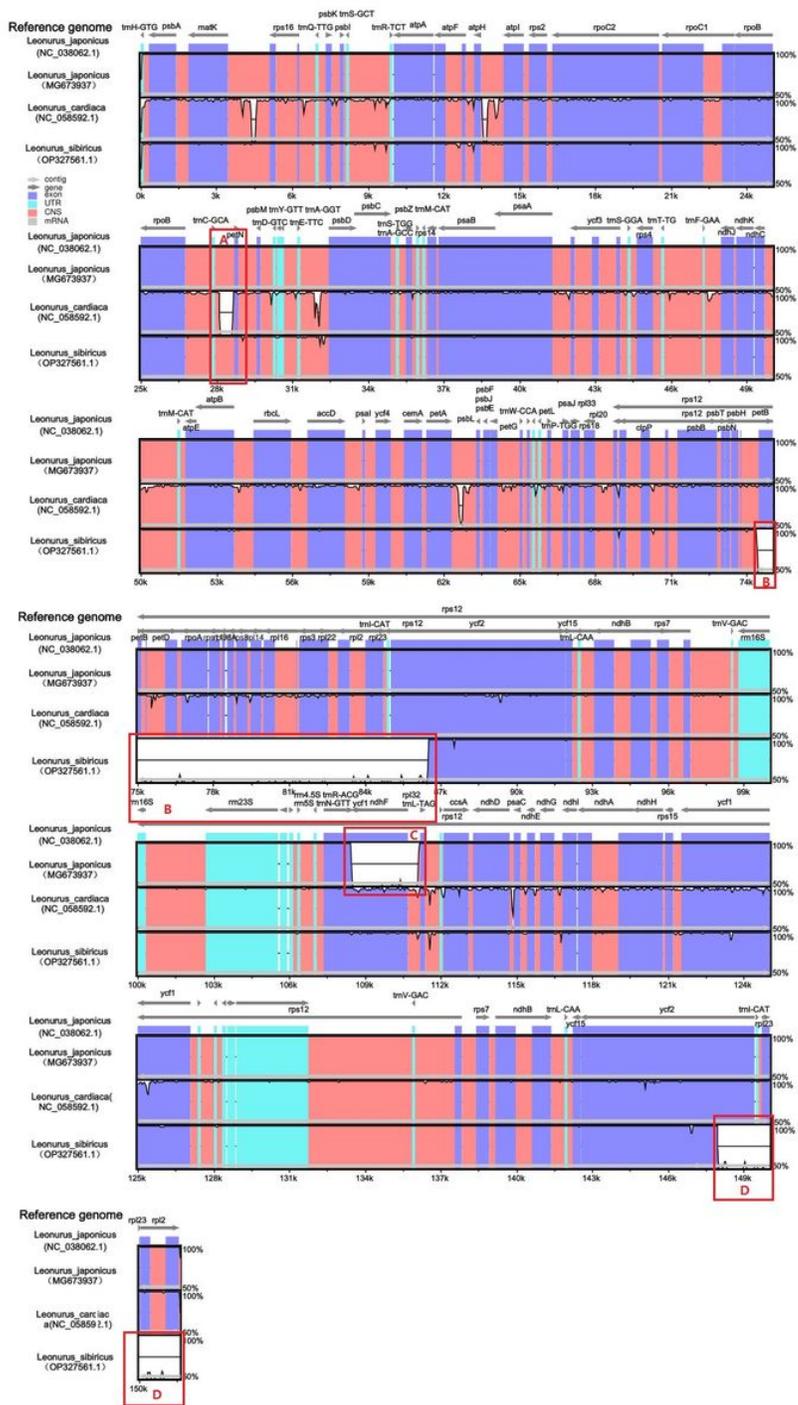


Figure 7

Structure comparison of the chloroplast genomes from the four *Leonurus* species by using the mVISTA program. Gray arrows and thick black lines above the alignment indicate genes with their orientation and the position of the IRs, respectively. A cut-off value of 70% identity was used for the plots, and the Y-scale represents the percent identity between 50% and 100%. UTR: Untranslated Regions; CNS: Conserved Non-coding Sequences. A: IGS (*trnC-GCA-petN*); B: IGS (*petB-trnI-CAT*); C: IGS (*rpl2-trnI-CAT*); D: IGS (*trnL-TAG-ndhF*).

K2P Distance for Various IGS

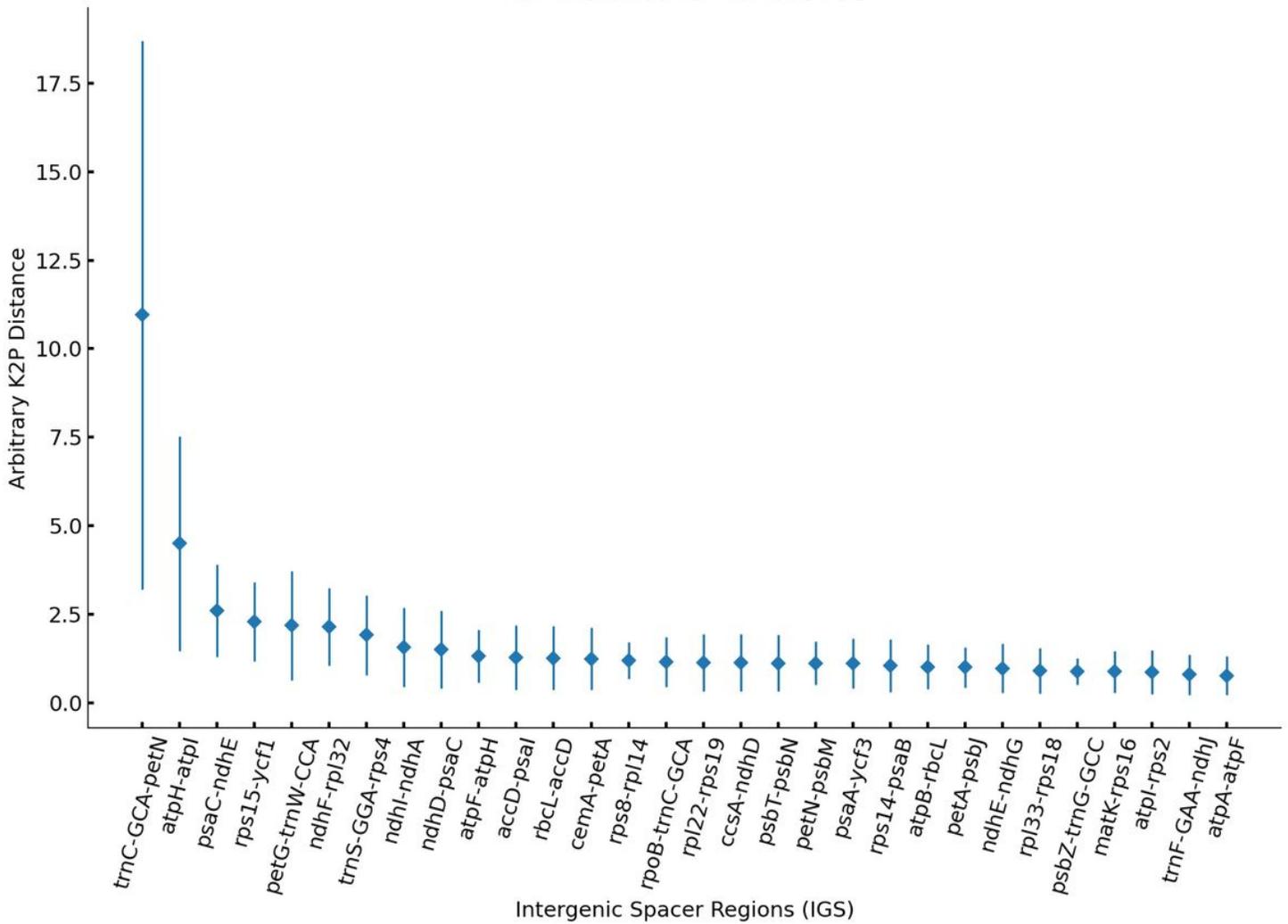


Figure 8

K2p distances among the intergenic spacer regions from 4 chloroplast genomes of *Leonurus* species. The K2p distances were calculated in pairs. The abscissa shows the different IGS regions and the ordinate shows the arbitrary K2P distance. The blue dots within the coordinate axis represent the average value of the three pairs. The Error bars represent the standard error among the three pairs.

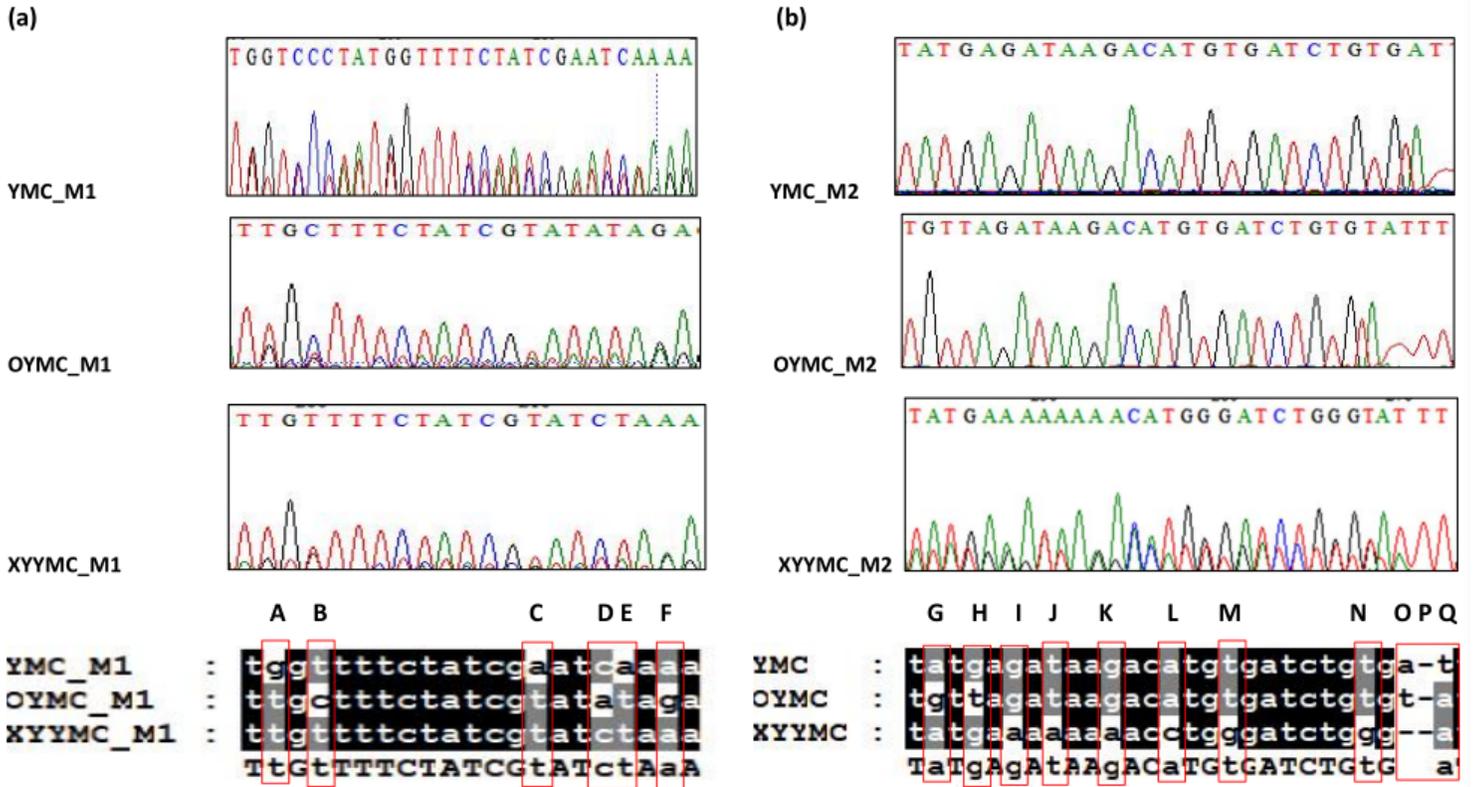


Figure 9

The peak map (up) and sequencing results (down) of the three studied *Leonurus* species with the pairs of primers M1 (a) and M2 (b). The symbols of YMC_M1, OYMC_M1, and XYMYC_M1 (a) are the sequencing results and peak maps of the M1 DNA barcode from one sample of *L. japonicus*, *L. cardiaca*, and *L. sibiricus*, respectively. The symbols of YMC_M2, OYMC_M2, and XYMYC_M2 (b) are the sequencing results and peak maps of the M2 DNA barcode from one sample of *L. japonicus*, *L. cardiaca*, and *L. sibiricus*, respectively. The variant bases and Indel sites have been marked A-Q in the red frame of the sequences according to the Arabian alphabet.

```

*          20          *          40          *          60          *          80
Leonurus_japonicus_ITS2 : CGCATCGCGTCGCCCCCTCCGCCGGGGTGGGGCGGAGATTGCCCCCGCTGCCACCGAGCGGC-GGGCCGGGCCAAATG : 84
Leonurus_pseudomacranthus_ITS2 : CGCATCGGTTCGCCCCCT--CCCGTGGGGTGTGGCGGAGATTGCCCTCTCCACATGCT-CGTGCAC-GCAGCCGGCCAAATG : 81
Leonurus_sibiricus_ITS2 : CGCATCGCGTCGCCCCCTCCGCCGGGGTGGGGCGGAGATTGCCCCCGCTGCC-CGATCGCGGC-GCGCCGGGCCAAATG : 83
Leonurus_chaituroides_ITS2 : CGCATCGCGTCGCCCCCTCCGCCGGGGTGGGGCGGAGATTGCCCCCGCTGCCACCGAGCGGC-GGGCCGGGCCAAATG : 84
Leonurus_turkestanicus_ITS2 : CGCATCGCGTCGCCCCCTCCGCCGGGGTGGGGCGGAGATTGCCCCCGCTGCCCGTACAGCCGAGCGCCGGGCCAAATG : 85
Leonurus_cardiaca_ITS2 : CGCATCGCGTCGCCCCCTCCGCCGGGGTGGGGCGGAGATTGCCCCCGCTGCCCGGTACAGCCGAGCGCCGGGCCAAATG : 85
Leonurus_glaucescens_ITS2 : CGCATCGCGTCGCCCCCTCCGCCGGGGTGGGGCGGAGATTGCCCCCGCTGCCCGGTACAGCCGAGCGCCGGGCCAAATG : 85
CGCATCGeGTCGCCCCCTcCCGCCGGGGTgGGCGGAGATTgGC CcCgTgCg c cGcG CgGCGGCCAAATG

*          100         *          120         *          140         *          160         *
Leonurus_japonicus_ITS2 : CGAATCCCGCTCGCGCCGCGTCGCGACCAGTGGTGGTTGATGATTCAACTCGCGTGTCTCGCGCCCGCGTCCGTCGGCAA : 169
Leonurus_pseudomacranthus_ITS2 : CGAATCCCTCTATCAGTCCCGCTCGCGACCAGTGGTGGTTGATAATTCAACTCGCGTGTCTCGGTATCCCGCATGTCGTCGGTTC : 166
Leonurus_sibiricus_ITS2 : CGAATCCCGCTCGCGCCGCGTCGCGACCAGTGGTGGTTGACGATTCAACTCGCGTGTCTCGCGTCCCGCGTCCCGTCCCGTCC : 168
Leonurus_chaituroides_ITS2 : CGAATCCCGCTCGCGCCGCGTCGCGACCAGTGGTGGTTGATGATTCAACTCGCGTGTCTCGCGCCCGCGTCCCGTCCCGTCC : 169
Leonurus_turkestanicus_ITS2 : CGAATCCCGCTCGAGCAGCTCGCGACCAGTGGTGGTTGAACATTCAACTCGCGTGTCTCGCGTTCATCCGGTCCGTCGGTTG : 170
Leonurus_cardiaca_ITS2 : CGAATCCCGCTCGAGCAGCTCGCGACCAGTGGTGGTTGAACATTCAACTCGCGTGTCTCGCGTTCATCCGGTCCGTCGGTTG : 170
Leonurus_glaucescens_ITS2 : CGAATCCCGCTCGAGCAGCTCGCGACCAGTGGTGGTTGAACATTCAACTCGCGTGTCTCGCGTTCATCCGGTCCGTCGGTTG : 170
CGAATCCgCgTcG cG CgTCCGACCAGTGGTGGTTGA ATTCAACTCGCGTGTCTCGcg C Cg cG CGTCgG

*          180         *          200         *          220
Leonurus_japonicus_ITS2 : GGAAACGATTCGAAACCCAACGGCGGAGCA---TCGTGCCACGACCG : 216
Leonurus_pseudomacranthus_ITS2 : GGAAACGACTACGAAACCCAACGGCGGAGCACGAAATCGTGCCACACCG : 217
Leonurus_sibiricus_ITS2 : GGAAACGATCACGAAACCCAACGGCGGAGCACGCATCGTGCCACGACCG : 219
Leonurus_chaituroides_ITS2 : GGAAACGATTCGAAACCCAACGGCGGAGCA---TCGTGCCACGACCG : 216
Leonurus_turkestanicus_ITS2 : GGAGACGATAAAGAAACCCAACGGCGGAGCACGCATCGTGCCACGACCG : 221
Leonurus_cardiaca_ITS2 : GGACACGATAAAGAAACCCAACGGCGGAGCACGCATCGTGCCACGACCG : 221
Leonurus_glaucescens_ITS2 : GGACACGATAAAGAAACCCAACGGCGGAGCACGCATCGTGCCACGACCG : 221
GGAACGAt GAAACCCAACGGCGGAgCA TCGTGCCACGACCG

```

Figure 10

The nucleotide comparison of the ITS2 DNA sequences in the seven *Leonurus* species. The left column shows the name of the seven *Leonurus* species. The black and grey background shows the same bases from these species. The white background shows the parts of the bases that exist the differentia in some species.

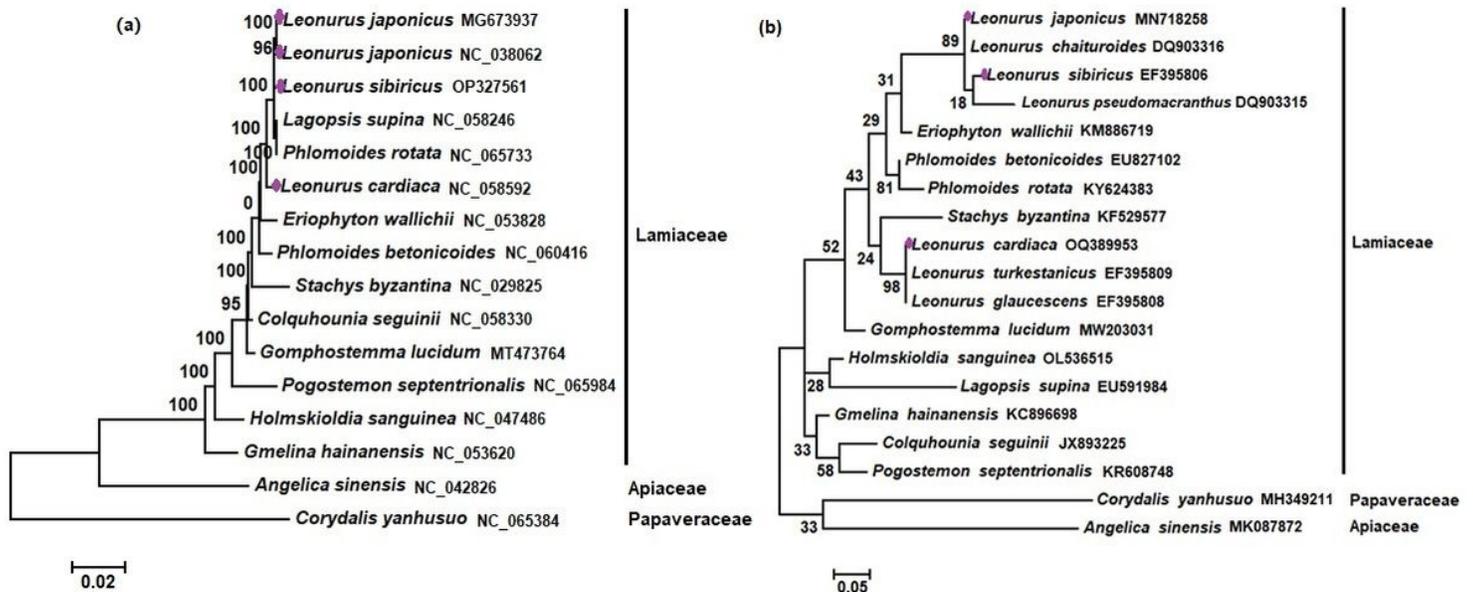


Figure 11

The phylogenetic relationships of nucleotide in the 16 studied species (a) for the cp genomes and 19 species based on the ITS2 DNA sequences. It includes 14 Lamiales species with the two species *Angelica sinensis* and *Corydalis yanhusuo* used as the outgroup. The tree was constructed with the nucleotide sequences of 64 CDS shared in the 16 species (a) and 19 ITS2 DNA by using the Maximum likelihood (ML) method. Bootstrap supports were calculated from 1000 replicates in the Fig. 11a.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Fig.S1.png](#)
- [Fig.S2.jpg](#)
- [Fig.S3.png](#)
- [Fig.S4.png](#)
- [Fig.S5.jpg](#)
- [Fig.S6.png](#)