THE COMPLETE CHLOROPLAST GENOMES OF *Leonurus turkestanicus* FROM THE CHATKAL AND NURATAU RIDGES (CENTRAL ASIA)

Umida TOJIBOEVA[*], Dilnoz AZIMOVA[**], Sardorbek MAMASOLIEV[***], Mokhistara SHARIPOVA[**], Ibrokhimjon ERGASHOV[*], Ulugbek KADIROV[*], Alijon ESANKULOV[****], Ziyoviddin YUSUPOV[*]

[*] *Institute of Botany, Academy of Sciences of Uzbekistan, Tashkent, Uzbekistan*
[**] *Jizzakh State Pedagogical University, Jizzakh, Uzbekistan*
[***] *Andijan State University, Andijan, Uzbekistan*
[****] *Academy of Sciences of the Republic of Uzbekistan, Tashkent Botanical Garden, Tashkent, 100053, Uzbekistan*
Corresponding author: Ziyoviddin Yusupov, Institute of Botany, Academy of Sciences of Uzbekistan, 32 Durmon yuli, Tashkent, Uzbekistan, e-mail: yusupov@mail.kib.ac.cn

Abstract. *Leonurus turkestanicus* (Turkestan Motherwort) is a perennial herb within the tribe *Leonureae* (Lamioideae, Lamiaceae), distributed mainly in Central Asia (Tien Shan, Pamir-Alay). It is a valuable herbal medicine with a calming effect on the nervous system and high blood pressure. In this study, the complete chloroplast genome of *Leonurus turkestanicus* from Chatkal (Tien Shan) and Nuratau (Pamir-Alai) ridges was sequenced and reported for the first time. The whole cp genome of *L. turkestanicus* ranged from 151,777 to 151,782 bp in length, contained a pair of inverted repeats (IR, 25,644–25,646 bp) regions separated by a small single copy (SSC, 17,653–17,654 bp) and the highest proportion is the large single copy (LSC, 82,831–82,841 bp). A total 129 genes were encoded including 87 protein-coding genes, 34 tRNA genes and 8 rRNA genes. A total of 220–226 simple sequence repeat (SSR) motifs were identified, in which A/T mononucleotide type is the most numerous. In this study, SSRs were mainly distributed in the LSC and SSC regions. A total of 40 to 42 long repeat sequences were identified in the direct, palindromic, and reverse categories, with the longest sequence being 46 base pairs. Sequence analysis indicated protein coding genes such as *rps*12, *mat*K, *ndh*F, *ndh*G, *rpl*20, *inf*A, *rps*15, *psb*K as hypervariable regions, which can be used as specific effective markers within *Leonurus* genus. The maximum-likelihood approach of phylogenetic analysis shows that *L.cardiaca* and *L. turkestanicus* are the most closely related taxa. Our study offers genetic data and provides valuable genomic resources for further investigations in valuable molecular markers suitable for identification of *Leonurus* species, on the phylogenetics and biodiversity of the genus *Leonurus*.

Keywords: chloroplast genome; Central Asia; comparative analysis; Lamiaceae; *Leonurus turkestanicus*; phylogenetic analysis.

INTRODUCTION

The genus *Leonurus* L. consists of approximately 25 species, with three commonly found in Uzbekistan. Among them, *Leonurus turkestanicus* V. Krecz. et Kupr (Turkestan Motherwort) is recognized as a promising medicinal plant [2]. *L. turkestanicus* is found in Iran, West Himalaya and Xinjiang [35], as well as the mountainous regions of Central Asia (Kyrgyzstan, Kazakhstan, Tajikistan, Turkmenistan, and Uzbekistan). It grows at 1000–2000 meters above sea level in the lower and middle mountain belts on stony and shallow-soil slopes, along riverbanks, in gorges, in shaded areas, and amid trees and shrubs [9, 30]. It grows 50–150 cm tall, with erect, sparingly branched stems. Leaves are simple, claw-shaped, and palmately lobed (5-lobed below, 3-lobed above). Ink-pink, bilabiate flowers are borne in axillary panicles. It flowers from June to August, producing fruit of four triangular, hairy, brown nutlets.

The contents of *Leonurus turkestanicus* include 0.8% flavonoids, 1.5% alkaloid stachydrine, 0.15% essential oil, and phenolic acids. Tannins (1.4–6.1%), organic acids (2.02–4.34%), and oil (34.4%) are present in the seeds and leaves, respectively. Leonurinine alkaloid (found in fruits), hyperoside, quercetin, 7-glucoside, quinqueloside, rutin, and ~0.4% stachydrine have been identified among the flavonoids.

Leonurus turkestanicus preparations have a sedative, hypotensive, and hypnotic effect in addition to having a relaxing impact on the central nervous system, slowing the heart rate and increasing the force of heart contractions. This plant's tincture is used in folk medicine to cure a variety of conditions, including heart defects, neuralgia, atherosclerosis, early-stage hypertension, cardiosclerosis, myocarditis, and myocardial dystrophy [1, 12, 16].

Substantial advances have been achieved in resolving the phylogenetic relationships within the Lamiaceae [20, 33, 34], and a temporal framework for the family's diversification has been established. Scheen et al. (2010) [26] clarified the Lamioideae framework using plastid DNA sequences, proposing that *Chaiturus*, *Lagochilus*, *Leonurus*, and *Panzerina* form the tribe Leonureae, with *Lagopsis* later added based on trnL-F data. Bendiksby *et al.* [4] expanded this analysis with four cpDNA regions, confirming *Lagopsis* and placing *Loxocalyx* within Leonureae. Roy and Lindqvist [24] further validated the tribe's boundaries using nuclear PPR markers, refining phylogenetic understanding within Lamioideae. However, the evolutionary relationships within *Leonurus,* as well as its taxonomic boundaries, are still not fully resolved, in part due to morphological similarity among species and the lack of extensive molecular data across its full geographic and taxonomic range [24].

Although there has been some progress, current studies remain fragmented and insufficient to construct a robust phylogeny for the genus. For instance, the chloroplast genomes of *L. japonicus* and *L. cardiaca* have been sequenced and analyzed, focusing on genome structure, repeat sequences, gene content, codon usage, and comparative genomic characteristics

(e.g., evolutionary development analysis) [31]. Furthermore, Yang *et al.* [32] amplified and sequenced the ITS region and *mat*K gene from ten *Leonurus* species in China including *L. japonicus* and *L. sibiricus* to explore their phylogenetic relationships. However, these efforts have largely concentrated on a limited number of species and genetic loci.

Given the limited scope of previous molecular investigations, especially in terms of taxon sampling and comprehensive genomic analysis, further research is necessary to clarify species boundaries and evolutionary relationships within *Leonurus.* Expanding the use of high-resolution genomic data, including complete chloroplast genomes and multilocus nuclear markers, will be essential in advancing our understanding of the phylogeny and diversification of this important genus.

In this study, we sequenced, assembled, and annotated the complete chloroplast genomes of *Leonurus turkestanicus* from two geographically distinct locations: the Nuratau ridge, a part of the Pamir-Alay system, and the Chatkal ridge, which is located in the southwestern part of the Tien Shan. Both regions known as an important global biodiversity [10, 29]. This study presents the first complete plastome report of *Leonurus turkestanicus* from two distinct localities. Through cross-species and cross-population comparisons of repeats and SSRs within *Leonurus*, we identify candidate hypervariable loci with potential as molecular markers. Furthermore, our results provide an incremental contribution to the understanding of phylogenetic relationships within the subfamily *Lamioideae.*

MATERIALS AND METHODS

Plant materials

Fresh leaves of *L. turkestanicus* were collected from the Western Tien Shan (sample L1. Western Tien Shan, Chatkal region. Right bank of Alatangasaya. 05.31.2020 *Tojibaev K. et al.,* E. 70° 04' 52.0" E, 41° 10' 45.3" N, *Gaziev A. 3105206*) and Nuratau ridge (sample L2. Nuratau Nature Reserve, Nuratau Nature Reserve, Khayatsai tract. 06.09.2020 *Ortikov E.*et al., 66° 44' 35.0" E, 40° 30' 49.5" N *Gaziev A. 9062020664*) (Fig. 1). All samples were stored in silica gel filled bags.

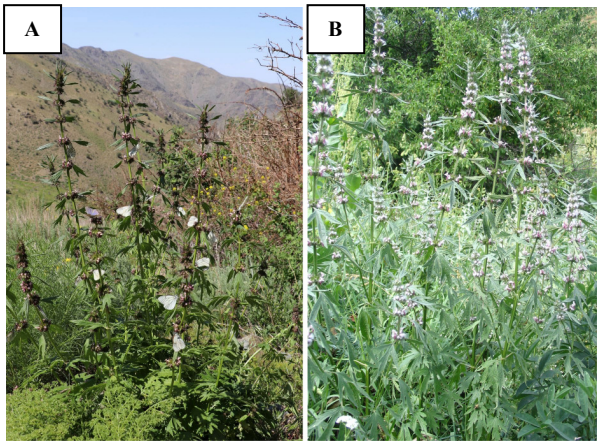

Figure 1. Living plants from their natural habitats: (A) Nuratau ridge (Photo by N. Beshko) and (B) Chatkal Ridge (Photo by K. Tojibaev)

Sequencing, assembly and annotation

Total genomic DNA was extracted from leaf tissues using the DP305 Plant Genomic DNA Kit (Tiangen, Beijing, China), following the manufacturer's protocol. DNA quality and quantity were assessed prior to library preparation. Sequencing libraries were constructed using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina (New England Biolabs, USA; Cat. No. E7370L), following the manufacturer's instructions. Briefly, DNA was fragmented by sonication to an average size of ~350 bp. Fragmented DNA underwent end repair, A-tailing, and ligation with Illumina-compatible adapters. Indexed adapters were added to each sample to enable multiplexing. Libraries were then amplified via PCR and purified using the AMPure XP system (Beckman Coulter, USA). Library quality was verified on an Agilent 5400 Fragment Analyzer, and concentrations were quantified using qPCR, targeting a final concentration of 1.5 nM.

Qualified libraries were pooled and sequenced using the Illumina HiSeq 4000 platform with 150 bp paired-end reads (PE150) at Novogene Bioinformatics Technology Co., Ltd. (Beijing, China). The raw reads were processed using the NGS QC Toolkit [22] to remove low-quality sequences and adapter contamination under default parameters. A total of 7,754,725 clean paired-end reads were obtained for sample L1 and 8,401,136 for sample L2.

Clean reads were assembled using NovoPlasty version 3.8.3 [7], and *Leonurus sibiricus* (GenBank accession: NC_067787) used as the reference. The assembly resulted in a single, complete circular contig representing the chloroplast genome for each sample. Read mapping was conducted in Geneious v10.0.2 to verify the accuracy of the assembly, showing even coverage and full genome representation.

Chloroplast genome annotation was performed using Geneious v10.0.2 [15] with *Leonurus sibiricus* (NC_067787) as a reference. Manual curation was applied to confirm gene boundaries, start and stop codons, and exon-intron junctions . Gene names were standardized according to the Chloroplast Genome Database [5]. Feature table files for GenBank submission were generated using GB2sequin [19].

The complete chloroplast genomes of *Leonurus turkestanicus* were deposited in GenBank under accession numbers PQ152226 (L1) and PQ152227 (L2). Metadata were submitted under BioProject PRJNA1140891, with BioSample IDs SAMN48099933 (L2) and SAMN48099932 (L1), and raw reads are available in the NCBI Sequence Read Archive under accession numbers SRR33253251 and SRR33253252, respectively.

Simple sequence repeats

The MIcroSAtellite (MISA) web tool [3] was used to identify cp simple sequence repeats (SSRs). The parameter to isolate perfect mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide motifs with a minimum of 8, 4, 3, 3, 3, and 3 repeats, respectively [35] REPuter software [17] was utilized to identify complement, palindrome, forward, and reverse repeats. To identify repetitions, the following parameters were used: (1) hamming distance = 3, (2) minimum repeat size = 30 bp, and (3) maximum computed repeats = 90 bp.

Nucleotide diversity (*Pi*) calculation

Pi diversity of of the chloroplast genomes of five samples, including two from *L. turkestanicus* (L1 and L2), as well as *L. cardiaca*, *L. japonicus*, and *L. sibiricus*, were conducted using DnaSP v6.12.03 [25], focusing on protein-coding genes. Genes with *Pi* values equal to 0 were excluded from the results. There was a 600 bp window length and a 100 bp step size [28]. A visual representation of the genome was created using OGDRAWv1.1 [21].

Phylogenetic Analysis

We have downloaded chloroplast genomes of of 28 species and 3 *Prostanthera* species (outgroup) of Lamiaceae family from NCBI (Table 1), Although *Prostanthera* (subfamily Prostantheroideae) is within Lamiaceae, it was selected as an outgroup because of its well-resolved position as an early-diverging lineage within the family and the availability of high-quality plastome data. The MAFFT algorithm was used to align the sequences [14]. The phylogenetic tree was constructed based on whole chloroplast genome (cp). Best-fit partitioning schemes and models were selected using the greedy search mode implemented in PartitionFinder v2.1.1 [18]. According to Edler *et al.* [8], the trees were built in raxmlGUI 2.0 using maximum likelihood (ML) and 1000 bootstrap replicates (BS metrics). Using the Akaike Information Criterion (AIC) in JModelTest V2.1.4, the most acceptable model for nucleotide substitutions was found to be the GTR+G model [6]. Tree visualization was done using FigTree v1.4.0 [23].

RESULTS

Genome features

This study analyzed the complete chloroplast genome of *L. turkestanicus*. The chloroplast genome sizes of the two *L. turkestanicus* populations were 151,777 bp and 151,782 bp, respectively. Each genome contains a pair of inverted repeat (IR) regions (IRa and IRb) measuring 25,644-25,646 bp, a large single-copy (LSC) region of 82,831-82,841 bp, and a small single-copy (SSC) region of 17,653-17,654 bp (Fig. 2, Table 2). The plastome is comprised of 129 genes, including 87 protein-coding genes, 34 tRNAsand 8 rRNAs (Table 3). Among these genes, 11 genes (*atp*F, *ndh*A, *ndh*B, *pet*B, *pet*D, *rpl*2, *rps*16, *rpo*C1, *trn*A-UGC,

Table 1. The list of 32 , their NCBI accession numbers and subfamilial taxonomic assignments used species for phylogenetic analysis.

No.	Species name	GenBank accession numbers	Subfamilies
1.	*Leonurus turkestanicus*	PQ152227*	
2.	*Leonurus turkestanicus*	PQ152226*	
3.	*Leonurus cardiaca*	MZ274170	
4.	*Leonurus sibiricus*	NC_067787	
5.	*Leonurus japonicus*	NC_038062	Lamioideae
6.	*Lamium takeshimense*	MN240520	
7.	*Lamium amplexicaule*	MT473770	
8.	*Phlomoides kirghisorum*	OR039725	
9.	*Phlomoides betonicoides*	NC_060416	
10.	*Scutellaria insignis*	NC_028533	
11.	*Scutellaria forrestii*	OP597817	Scutellarioideae
12.	*Scutellaria baicalensis*	MF521632	
13.	*Ajuga campylantha*	NC_084165	
14.	*Ajuga forrestii*	NC_048512	Ajugoideae
15.	*Ajuga nubigena*	NC_071844	
16.	*Thymus japonicus*	NC_046822	
17.	*Thymus quinquecostatus*	OP834031	
18.	*Thymus linearis*	NC_071852	
19.	*Thymus mongolicus*	MN482117	
20.	*Thymus villosus*	ON641262	
21.	*Thymus vulgaris*	ON641268	
22.	*Thymus mastichina*	NC_086485	
23.	*Mentha spicata*	NC_037247	Nepetoideae
24.	*Mentha longifolia*	NC_032054	
25.	*Mentha canadensis*	NC_044082	
26.	*Clinopodium chinense*	NC_050943	
27.	*Clinopodium euosmum*	OP311697	
28.	*Salvia japonica*	MW381778	
29.	*Salvia nanchuanensis*	MZ900990	
30.	*Salvia officinalis*	OM617846	
31.	*Prostanthera wilkieana*	NC_084412	
32.	*Prostanthera althoferi*	OR699053	Outgroups
33.	*Prostanthera campbellii*	OR699048	

*trn*K-TTT, *trn*I-GAU) have one intron and *clp*P and *ycf*3 have two introns (Supplementary table 1). The *trn*K-UUU gene contained the largest intron (from 1607 to 4166 bp). The GC content of the *L. turkestanicus* is 38%, while the corresponding values for the LSC, SSC, IR regions are 36.6%, 32.2%, and 43.4% respectively.

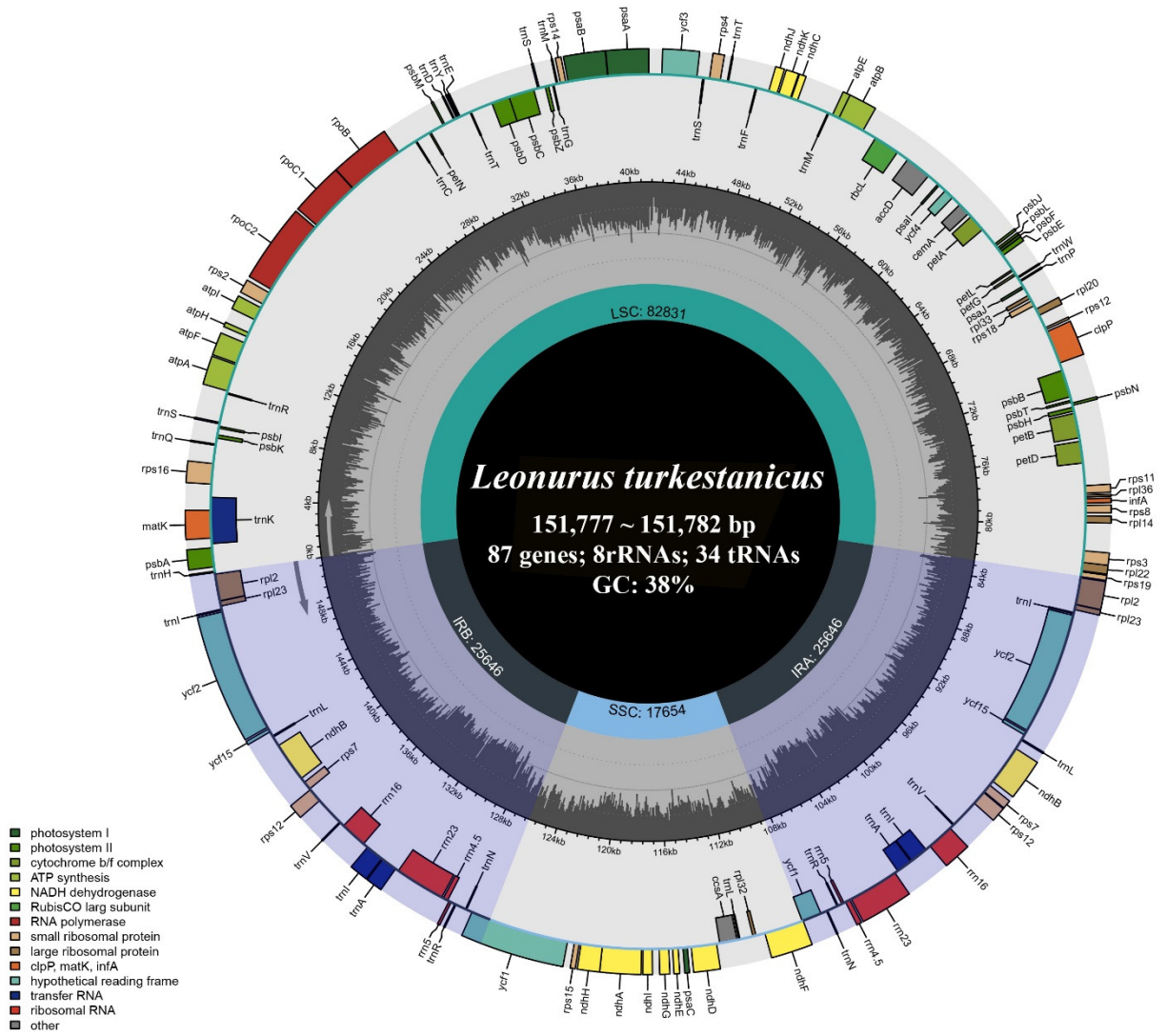

Figure 2. Gene map of the *Leonurus turkestanicus* chloroplast genome drawn by OGDRAW (the graph belongs to L1 sample). Genes shown inside the circle are transcribed clockwise, and those outside are transcribed counterclockwise. Genes belonging to different functional groups are color-coded. The darker gray color in the inner circle corresponds to the GC content, and the lighter gray color corresponds to the AT content. LSC, SSC, and IR are large single-copy regions, small single-copy regions, and inverted repeat regions, respectively.

Table 2. Comparison of plastome features of *L. turkestanicus* from two populations

Taxon	Total length (bp)	LSC (bp)	SSC (bp)	IR (bp)
Western Tian Shan (L1)	151.777	82.831	17.654	25.646
Nuratau ridge (L2)	151.782	82.841	17.653	25.644

Table 3. List of annotated genes in the chloroplast genomes of *Leonurus*

Category of genes	Group of genes	Genes
0	**1**	**2**
Genes for photosynthesis (44)	Subunits of photosystem I	*psa*A, *psa*B, *psa*C, *psa*I, *psa*J
	Subunits of photosystem II	*psb*A, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*I, *psb*J, *psb*H, *psb*K, *psb*L, *psb*M, *psb*N, *psb*T, *psb*Z
	Subunits of ATP synthase	*atp*A, *atp*B, *atp*E, *atp*F*, *atp*H, *atp*I
	Subunits of NADH-dehydrogenase	*ndh*A*, *ndh*B*(2), *ndh*C, *ndh*D, *ndh*E, *ndh*F, *ndh*G, *ndh*H, *ndh*I, *ndh*J, *ndh*K
	Subunits of cytochrome b/f complex	*pet*A, *pet*B*, *pet*D*, *pet*G, *pet*L, *pet*N
	RubisCO large subunit	*rbc*L

Continuation of Table 3:

0	1	2
Self-replication (57)	Large subunit of ribosome	*rpl2*(2), rpl14, rpl20, rpl22, rpl23 (2), rpl32, rpl33, rpl36*
	Small subunit of ribosome	*rps2, rps3, rps4, rps7(2), rps8, rps11, rps12**(2), rps14, rps15, rps16*, rps18*
	RNA polymerase	*rpo*B, *rpo*C1*, *rpo*C2
	Ribosomal RNAs	*rrn5(2), rrn4.5(2), rrn16(2), rrn23(2)*
	tRNA genes	*trn*A-*UGC*(2), trn*C-*GCA, trnD-GTC, trnE-TTC, trn*F-*GAA ,trn*G-*GCC, trnH-GTG, trn*I-*CAT (2), trn*I-*GAU* (2), trnK-TTT*, trnL-TAG, trnL-CAA(2), trn*M-*CAT(2), trnN-GTT(2), trnP-TGG, trnQ-TTG, trnR-ACG(2), trnR-TCT, trnS-GCT, trn*S-*GGA, trnS-TGA, trnT-GGT, trnT-TGT, trnV-GAC(2), trnW-CCA, trnY-GTA*
Other genes (5)	Subunit of acetyl-CoA-carboxylase	*accD*
	c-type cytochrome synthesis gene	*ccsA*
	Envelop membrane protein	*cemA*
	Protease	*clpP***
	Maturase	*matK*
Genes with unknown function (5)	hypothetical chloroplast reading frames (ycf)	*ycf1 (2), ycf2(2), ycf3**, ycf4, ycf15(2)*

*Indicates gene with one intron
**Indicates gene with two introns
(2) Indicates that the number of the repeat unit is 2
Counts were based on unique genes.

Repeat sequence analysis in *Leonurus* chloroplast genomes

A comparative analysis of chloroplast genomes across four *Leonurus* species, *L. turkestanicus* (L1 and L2), *L. cardiaca*, *L. japonicus*, and *L. sibiricus*, revealed notable patterns in repeat sequence distribution (Fig. 3). In *L. turkestanicus*, a total of 42 repeat sequences were identified, with the longest extending up to 46 base pairs (bp). These repeats were classified into four types based on their orientation and structural characteristics: forward, reverse, palindromic, and complementary. Forward and palindromic repeats were the most abundant across all examined species. The number of forward repeats ranged from 18 to 20, while palindromic repeats ranged from 21 to 22. Each genome contained only a single reverse repeat, and complementary repeats were entirely absent, suggesting that this type of repeat is not a feature of *Leonurus* chloroplast DNA.

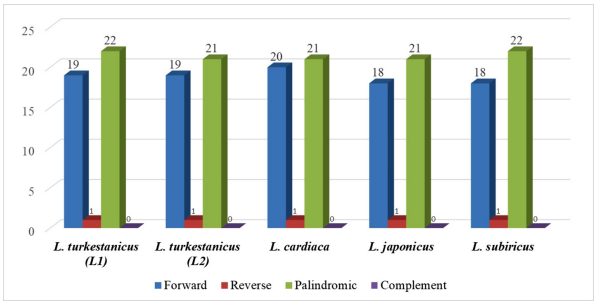

Figure 3. Distribution of simple sequence repeats (SSRs) across the chloroplast genomes of 4 *Leonurus* species, categorized into forward, reverse, palindromic, and complement types. The graph shows slightly different number of those repeats across species/samples.

SSR analysis and characterization of motif diversity in *Leonurus* chloroplast genomes

The analysis of SSRs indicated four categories of SSRs: mononucleotides, dinucleotide, trinucleotide, tetranucleotide (Table 4). The most abundant of the mononucleotide repeats is represented by the A/T motif. Among five *Leonurus* cp genomes, the most abundant repeats were the mononucleotides from 120 (*L. japonicus, L. sibiricus*) to 123–124 (*L. turkestanicus, L. cardiaca*). The AT/AT motifs are the most predominant representing dinucleotide repeat types. Trinucleotide repeats mainly represented by AAG/CTT (19 in all species) and AAT/ATT: 22 for *L. turkestanicus*, *L. cardiaca;* 24 for *L. japonicas and* 26 for *L. subiricus.* The tetranucleotide repeats are presented by AAAT/ATTT, AATC/ATTG, ACAG/CTGT. A unique tetranucleotide repeat types AAAC/GTTT, AGAT/ATCT were detected in *L. japonicus* and *L. sibiricus*. The most predominant motifs are mononucleotide (53%), followed by trinucleotide (28%), dinucleotide (15%), tetranucleotide (3%) repeats. The penta- and hexanucleotide repeats did not exhibit in *Leonurus* species. The identified SSRs were predominantly in the LSC region of the genome.

Comparison of genomic divergence

To determine the degree of divergence among protein-coding genes in five *Leonurus* samples, nucleotide diversity (*Pi*) values were calculated (Fig. 4). The analysis included two *L. turkestanicus* populations (L1 and L2), *L. cardiaca, L. japonicus,* and *L. sibiricus*. Among the genes analysed, *ndh*B exhibited the lowest nucleotide diversity (*Pi* = 0.00039), indicating high conservation across the samples. In contrast, *rps*12 (*Pi* = 0.01345), *mat*K (*Pi* = 0.01068), *ndh*F (*Pi* = 0.01026), and *ndh*G (*Pi* = 0.00904) displayed the highest variability, suggesting these genes may be subject to adaptive divergence.

Approximately 1.3% of the protein-coding genes exhibited high nucleotide diversity (*Pi* > 0.01), while around 59.74% showed moderate diversity with *Pi* values between 0.002 and 0.01. The remaining ~38.96% of genes displayed low diversity, with *Pi* values ranging from 0.0 to 0.002 (Supplementary Table 3).

Table 4. Types and number of short repeat sequences in *Leonurus* species chloroplast genomes.

Types of short repetitive sequences	**Types**	***L. turkestanicus* (L1)**	***L. turkestanicus* (L2)**	***L. cardiaca***	***L. japonicus***	***L. sibiricus***
Mono	A/T	112	112	111	115	115
	C/G	12	11	12	5	5
Di	AC/GT	3	3	3	4	4
	AG/CT	12	12	12	12	12
	AT/AT	19	19	18	19	19
Tri	AAC/GTT	7	7	7	8	8
	AAG/CTT	19	19	19	19	19
	AAT/ATT	22	22	22	24	26
	AGC/CTG	5	5	5	5	5
	AGG/CCT	1	1	1	1	1
	ATC/ATG	6	6	6	6	6
Tetra	AAAT/ATTT	2	2	2	2	2
	AATC/ATTG	1	1	1	1	1
	ACAG/CTGT	1	1	1	1	1
	AGAT/ATCT	-	-	-	1	1
	AAAC/GTTT	-	-	-	1	1
Total		**222**	**221**	**220**	**224**	**226**

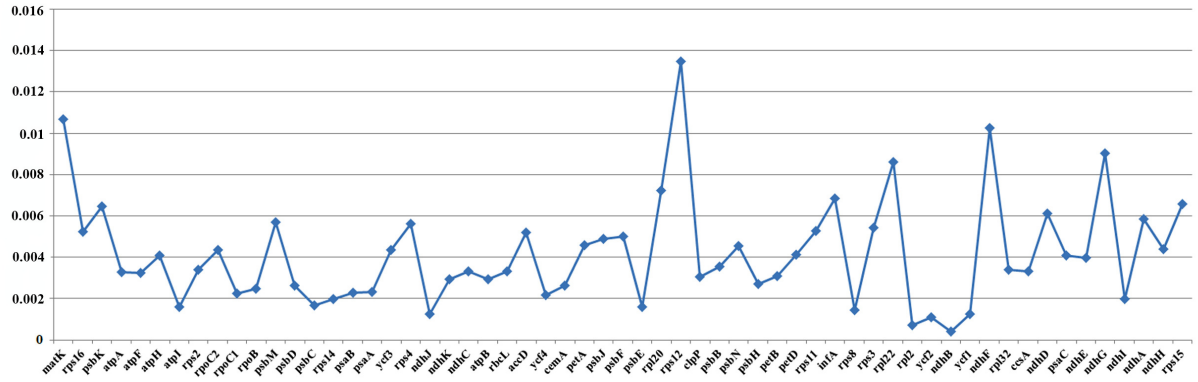

Figure 4. Nucleotide diversity (*Pi* in the y axis) in protein-coding genes (in the x axis) of *Leonurus* species, excluding genes with zero diversity. Data includes genes from *L. turkestanicus* (L1 and L2), *L. cardiaca*, *L. japonicus*, and *L. sibiricus*. Hypervirable genes are circularized

Identification of highly variable regions of two populations from Chatkal and Nuratau ridges

For the two samples of *Leonurus*, the average sequence diversity was 0.04% for the entire genome, and 0.01%, 0.01%, and 0.04% for SSC, LSC, and IR, in that orderrespectively. Diversity of protein-coding genes and intergenic spacer regions is presented in Table 5,. Because of the existence of several indels, the lengths of genes *pet*D, *pet*B, and *ndh*A differed when compared pairwise (4, 1, and 1). Within the IGS areas, short- and medium-sized insertions and deletions were more common than long-length indels.

The analysis identified multiple polymorphic sites across the studied sequences. Among these, both transitions (Ts) and transversions (Tv) were observed (Table 5). Transitions - substitutions between purines (A↔G) or between pyrimidines (C↔T) - were more frequent than transversions, which involve substitutions between purine and pyrimidine bases (e.g., A↔T or G↔C). (Ts) and (Tv) are the major causes for genome variation. There was a total of 6 polymorphic sites in the protein coding regions (1, 5, and 0 in the IR, LSC, and SSC, respectively), and 14 (4, 9, and 1 in the IR, LSC, and SSC, respectively) in the intergenic spacer regions. For the two groups of genetic regions, the ratios of Ts:Tv were the same (12:12). The overall number of polymorphic sites in all sections of the plastome was 48 (15 in the IR, 32 in LSC, and 1 in SSC), with 50% Ts and 50% Tv. The genomic regions *rbc*L/*acc*D, *ycf*1, *mat*K, *rpo*B, *pet*N/*psb*M, *psa*B, *ycf*3, *ndh*C/*atp*E, *psb*T/*psb*N, *rps*8/*rpl*14, and *ndh*D were shown to have the highest nucleotide variability overall (Table 5).

Phylogenetic analysis

Phylogenetic tree was built based on the cp genomes of five *Leonurus* and twenty-eight related (Fig. 5). A 100% bootstrap value was obtained for almost every clade in the tree based on ML analyses. According to the results, *Lamioideae, Scutellarioideae* and *Ajugoideae* formed a sister clade to *Nepetoideae*. Furthermore, *Leonurus* was demonstrated to be sister to *Lamium* in the tree (with 100% bootstrap). The clade including species of the genus *Lamium* was sister to *Leonurus*. The most close species to *L. turkestanicus* was *L. cardiaca.*

Table 5. Comparison of base substitutions and indels in protein coding between two populations of *Leonurus turkestanicus* (Western Tien Shan and Nuratau ridge).

Gene	Indel (bp)	Number of polymorphic sites			Nucleotide identity
		Total	Ts	Tv	
LSC					
*mat*K	0	2	1	1	99.9%
*atp*F/*atp*H	2	0	0	0	99.4%
*atp*H	0	1	1	0	99.6%
*atp*H/*atp*I	2	0	0	0	99.7%
*rps*2/*rpo*C2	1	0	0	0	99.5%
*rpo*C2	0	1	0	1	99.98%
*rpo*C1	0	1	1	0	99.95%
*rpo*B	0	2	0	2	99.95%
*pet*N/*psb*M	1	2	2	0	99.6%
*psb*M/*psb*D	1	1	0	1	99.9%
*psb*D	0	1	1	0	99.9%
*psb*Z/*rps*14	0	1	0	1	99.9%
*psa*B	0	2	2	0	99.9%
*ycf*3	0	2	1	1	100%
*rps*4/*ndh*J	0	1	1	0	99.96%
*ndh*K	0	1	0	1	99.9%
*ndh*C/*atp*E	0	2	1	1	99.9%
*rbc*L/*acc*D	0	5	0	5	99.3%
*psa*J/*rpl*33	1	1	1	0	99.7%
*rpl*20/*rps*12	1	0	0	0	99.9%
*psb*T/*psb*N	0	2	2	0	96.7%
*pet*B	1	0	0	0	100%
*pet*B/*pet*D	0	1	0	1	99.5%
*pet*D	4	0	0	0	100%
*rps*8/*rpl*14	0	2	1	1	99.1%
*rpl*14/*rps*3	0	1	1	0	99.9%
IR					
*rpl*22/*rpl*2	1	0	0	0	99.7%
*ycf*2	0	1	1	0	99.99%
*rps*12/*ycf*1	1	1	1	0	99.98%
*ndh*F	0	1	0	1	99.96%
*rpl*32/*ccs*A	0	1	1	0	99.9%
*ccs*A/*ndh*D	0	1	0	1	99.4%
*ndh*D	0	2	0	2	99.9%
*ndh*I/*ndh*A	0	1	0	1	98.8%
*ndh*A	1	1	0	1	100%
*ycf*1	0	5	3	2	99.9%
*ycf*1/*rps*7	3	1	1	0	99.96%
SSC					
*ycf*2	0	1	1	0	99.99%
*rpl*2/-	1	0	0	0	98.9%

DISCUSSION

This study is the first to provide a comprehensive analysis of the chloroplast genome features of *Leonurus turkestanicus*, an important medicinal plant species from Central Asia. The complete sequencing of this genome provides valuable insights into the genetic structure and evolutionary relationships within the genus Leonurus. The genome size and structure, including the presence of inverted repeats and single-copy regions, are consistent with those found in other members of the Lamiaceae family, indicating a conserved genomic structure [13, 34].

The genome sizes of *L. turkestanicus* range from 151,777 to 151,782 bp. Each genome contains 129 genes, including 87 protein-coding genes, 34 tRNA genes, and 8 rRNA genes, highlighting the complexity and functionality of the chloroplast genome, which is crucial for photosynthesis and other metabolic processes.

The low genetic diversity observed in *Leonurus* populations, particularly in the ndhB and other genes (Pi = 0),. low evolutionary pressure.However, in our analysis we found that 4 genes (*rps*12, *mat*K, *ndh*F and *ndh*G) exhibited the highest *Pi* diversity. Together, these six divergence hotspot regions should be useful for developing molecular markers for phylogenetic and phylogeographic analyses as well as plant identification of *Leonurus* species.

A phylogenetic analysis of 32 species strongly supports the monophyly of major clades within the Lamiaceae family, including the subfamilies *Lamioideae*, Scutellarioideae, Ajugoideae, and Nepetoideae. Moreover, as in other studies [24], the genera *Lamium* and *Phlomoides* were shown to be sister groups to the genus *Leonurus* within the *Lamioideae* subfamily. This suggests that these genera in the *Lamioideae* subfamily have a tight evolutionary relationship with one another. Three species of Leonurus, including *L. cardiaca*, *L. sibiricus,* and *L. japonicus*, constituted a clade within the *Leonurus* genus, together with the species *L. turkestanicus*. The analysis revealed a close relationship between *L. cardiaca* and *L. turkestanicus*, suggesting similar

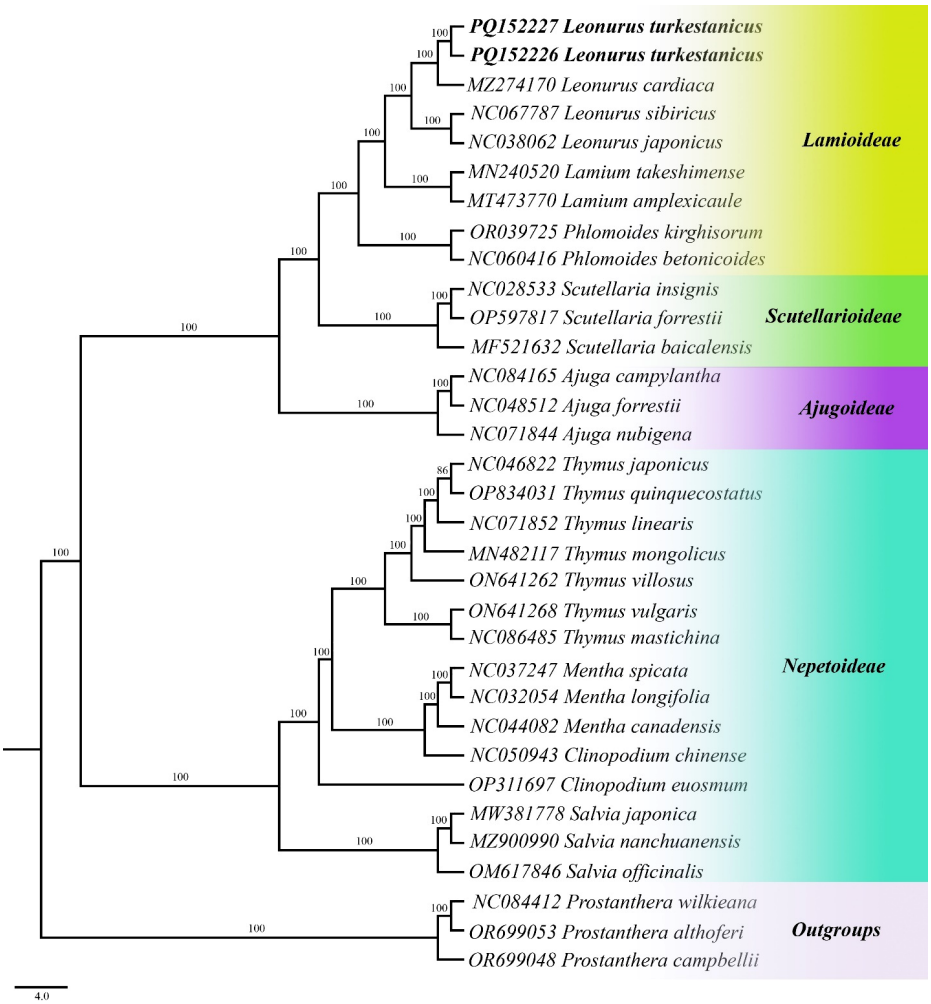

Figure 5. Phylogenetic tree of 33Lamiaceae species, including four *Leonurus* species, inferred using maximum likelihood (ML) analysis based on complete chloroplast genomes. Subfamilies are represented by distinct colors, and bootstrap values are indicated near the branches to assess node support. The NCBI accession numbers are given for each species before its name.

evolutionary pathways and potentially comparable medicinal properties. These results align with prior comparative studies of *Lamioideae* by Roy *et al.* [24], and Scheen *et al.* [26], yet also offer novel perspectives through the inclusion of *L. turkestanicus*, a previously underrepresented taxon. However, our tree includes only 4 *Leonurus* species, thus broader taxon sampling and nuclear data (e.g., target capture, UGP loci) are necessary to confirm this relationship. By integrating new genomic data, our findings refine the understanding of phylogenetic relationships within *Leonurus* and contribute to broader efforts in reconstructing the evolutionary history of Lamiaceae.

This study also provides a valuable genomic resource for future research on the phylogenetics and biodiversity of the genus *Leonurus.* The availability of the complete chloroplast genome sequence can facilitate comparative genomic studies, helping to identify genetic markers for species identification and conservation efforts [11, 27]. The creation of conservation measures to preserve *Leonurus* species in their native habitats can be facilitated by an understanding of the genetic diversity within the genus, particularly in light of changing environmental conditions and habitat degradation.

This study provides new plastome data and comparative insights into *Leonurus*, contributing to a clearer understanding of phylogenetic relationships within *Lamioideae*. Future work should expand taxon and population sampling across *Leonurus* to capture broader genomic diversity. Integrating plastome data with multilocus nuclear datasets will further resolve complex evolutionary histories. Moreover, investigating adaptive signals such as RNA editing patterns, Ka/Ks ratios, and selection in hypervariable plastid genes will enhance understanding of molecular evolution and environmental adaptation in the genus.

Acknowledgment. The authors would like to thank prof. Tojibaev K. and Dr. Ortikov E. for sharing collected samples and related informations. We also thank anonymous reviewers, for their comments. We appreciate the time and effort you put into reviewing our work.This research was funded by the project titled “Molecular-genetic identification of medicinal plant species in the flora of Uzbekistan and Belarus using DNA markers” (FL-7923051878). This research was also supported by the State Programs “Digital Nature: Development of a digital platform for the flora of Central Uzbekistan”, implemented by the Institute of Botany of the Academy of Sciences of the Republic of Uzbekistan for the period 2025-2029".

Conflict of interest. There is no actual or potential conflict of interest in relation to this article.

Supplementary Table 1. Location and length of intron-containing genes.

Gene	Start	End	ExonI	IntronI	ExonII	IntronII	ExonIII
*trn*K-TTT	1607	4166	36	2496	27	-	-
*rps*16	4889	6029	225	874	42	-	-
*atp*F	11734	12946	411	659	144	-	-
*rpo*C1	20590	23428	1611	796	432	-	-
*ycf*3	42075	44017	150	725	234	709	126
*clp*P	69019	70886	225	627	297	657	69
*pet*B	73782	75164	6	738	639	-	-
*pet*D	75359	76572	9	731	474	-	-
*rpl*2	82926	84417	470	628	394	-	-
*ndh*B	93088	95299	756	681	777	-	-
*trn*I-GAU	100584	101592	32	937	40	-	-
*trn*A-UGC	101657	102543	36	800	51	-	-
*ndh*A	117570	119739	540	1080	555	-	-

Supplementary Table 2. Forward, reverse, palindromic, and complement repeat regions of *L. turkestanicus.*

No.	Repeat type	The length of the first part	The starting point of the first part	The length of the second part	The starting point of the second part
1	Forward	46	90038	46	90056
2	F	46	144508	46	144526
3	F	41	96925	41	118146
4	F	39	43220	39	96927
5	F	39	43220	39	118148
6	F	34	90050	34	90068
7	F	31	42280	31	96949
8	F	30	105789	30	105820
9	F	30	128758	30	128789
10	F	30	43232	30	96939
11	F	31	8247	31	35156
12	F	31	31080	31	31103
13	F	31	90038	31	90074
14	F	31	144508	31	144544
15	F	30	9748	30	36041
16	F	30	37371	30	39607
17	F	30	43232	30	118160
18	F	30	87638	30	87680
19	F	30	146900	30	146942
20	Reverse	33	47592	33	47596
21	Palindromic	19690	82831	19690	132089
22	P	5957	102520	5957	126131
23	P	46	90038	46	144508
24	P	46	90056	46	144526
25	P	41	118146	41	137644
26	P	38	29356	38	29356
27	P	39	43220	39	137644
28	P	34	90050	34	144508
29	P	34	90068	34	144526
30	P	38	113319	38	113319
31	P	30	8248	30	44395
32	P	33	62631	33	62631
33	P	31	42280	31	137630
34	P	30	105789	30	128758
35	P	30	105820	30	128789
36	P	33	58704	33	58704
37	P	30	43232	30	137641
38	P	31	90038	31	144505
39	P	31	90074	31	144541
40	P	30	35157	30	44395
41	P	30	87638	30	146900
42	P	30	87680	30	146942

Supplementary Table 3. Individual characteristics of 77 protein-coding genes.

No.	Genes	Nucleotid diversity (*Pi* value)	Gene length (bp), after alignment	Gene type
0	**1**	**2**	**3**	**4**
1	*psb*A	0.00321	1059	reverse
2	*mat*K	0.01068	1536	reverse
3	*rps*16	0.00524	267	reverse
4	*psb*K	0.00645	186	forward
5	*psb*I	0	111	forward
6	*atp*A	0.00328	1524	reverse
7	*atp*F	0.00324	555	reverse
8	*atp*H	0.00407	246	reverse
9	*atp*I	0.00161	744	reverse
10	*rps*2	0.00338	711	reverse
11	*rpo*C2	0.00434	4191	reverse
12	*rpo*C1	0.00225	2043	reverse
13	*rpo*B	0.00249	3213	reverse
14	*pet*N	0	96	forward
15	*psb*M	0.00571	105	reverse
16	*psb*D	0.00264	1062	forward
17	*psb*C	0.00169	1422	forward
18	*psb*Z	0	189	forward
19	*rps*14	0.00198	303	reverse
20	*psa*B	0.00227	2205	reverse
21	*psa*A	0.00231	2253	reverse
22	*ycf*3	0.00434	510	reverse
23	*rps*4	0.00561	606	reverse
24	*ndh*J	0.00126	477	reverse
25	*ndh*K	0.00295	678	reverse
26	*ndh*C	0.00331	363	reverse
27	*atp*E	0	402	reverse
28	*atp*B	0.00294	1497	reverse
29	*rbc*L	0.00332	1446	forward
30	*acc*D	0.00518	1467	forward
31	*psa*I	0	111	forward
32	*ycf*4	0.00216	555	forward
33	*cem*A	0.00261	690	forward
34	*pet*A	0.00457	963	forward
35	*psb*J	0.00488	123	reverse
36	*psb*L	0	117	reverse
37	*psb*F	0.005	120	reverse
38	*psb*E	0.00159	252	reverse
39	*pet*L	0	96	forward
40	*pet*G	0	114	forward
41	*psa*J	0	135	forward
42	*rpl*33	0	201	forward
43	*rps*18	0	306	forward
44	*rpl*20	0.00724	387	reverse
45	*rps*12	0.01345	372	reverse
46	*clp*P	0.00305	591	reverse
47	*psb*B	0.00354	1527	forward
48	*psb*T	0	108	forward
49	*psb*N	0.00455	132	reverse
50	*psb*H	0.0027	222	forward
51	*pet*B	0.0031	645	forward
52	*pet*D	0.00414	483	forward
53	*rps*11	0.00528	417	reverse
54	*rpl*36	0	114	reverse
55	*inf*A	0.00684	234	reverse
56	*rps*8	0.00145	414	reverse
57	*rpl*14	0	369	reverse
58	*rps*3	0.00543	663	reverse
59	*rpl*22	0.0086	465	reverse
60	*rpl*2	0.00073	825	reverse
61	*rpl*23	0	282	reverse
62	*ycf*2	0.00108	6849	forward
63	*ycf*15	0	192	reverse
64	*ndh*B	0.00039	1533	reverse
65	*rps*7	0	468	reverse
66	*ycf*1	0.00125	1119	forward
67	*ndh*F	0.01026	2223	reverse
68	*rpl*32	0.00339	177	forward
69	*ccs*A	0.00332	963	forward
70	*ndh*D	0.0061	1410	reverse
71	*psa*C	0.00407	246	reverse

Continuation of Supplementary Table 3:

0	1	2	3	4
72	*ndh*E	0.00396	306	reverse
73	*ndh*G	0.00904	531	reverse
74	*ndh*I	0.00197	507	reverse
75	*ndh*A	0.00586	1095	reverse
76	*ndh*H	0.0044	1182	reverse
77	*rps*15	0.00659	273	reverse

REFERENCES

[1] Akopov, I.E., (1990): Most important domestic medicinal plants and their application. Meditsina, Tashkent, 378 p.

[2] Arabova, N.Z., Bahadirova, G.I.K., (2023): *Leonurus turkestanicus* – A promising medicinal plant for Uzbekistan. Bulletin of Science and Education, 6(137): 14-16.

[3] Beier, S., Thiel, T., Münch, T., Scholz, U., Mascher, M., (2017): MISA-web: a web server for microsatellite prediction. Bioinformatics, 33(16): 2583-2585.

[4] Bendiksby, M., Thorbek, L., Scheen, A.-C., Lindqvist, C., Ryding, O., (2011): An updated phylogeny and classification of Lamiaceae subfamily Lamioideae. Taxon, 60(2): 471-484.

[5] Cui, L., Veeraraghavan, N., Richter, A., Wall, K., Jansen, R.K., Leebens-Mack, J., Depamphilis, C.W., (2006): ChloroplastDB: the chloroplast genome database. Nucleic acids research, 34(suppl_1): D692-D696.

[6] Darriba, D., Taboada, G., Doallo, R., Posada, D., (2012): jModelTest 2: more models, new heuristics and parallel computing. Nature methods, 9(8): 772-772.

[7] Dierckxsens, N., Mardulyn, P., Smits, G., (2017): NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic acids research, 45(4): e18-e18.

[8] Edler, D., Klein, J., Antonelli, A., Silvestro, D., (2021): RaxmlGUI 2.0: a graphical interface and toolkit for phylogenetic analyses using RAxML. Methods in Ecology and Evolution, 12(2): 373-377.

[9] Eisenman, S.W., Zaurov, D.E., Struwe, L., (Eds.). (2013): Medicinal plants of central Asia: Uzbekistan and Kyrgyzstan. Springer Science & Business Media, 340 p.

[10] Ergashov, I., Yusupov, Z., Dolatyari, A., Khorasani, M., Eker, I., Turdumatova, N., Tojibaev, K., (2025): New insights into the molecular phylogeny and biogeographical history of *Allium* subgenus *Melanocrommyum* (Amaryllidaceae) based on plastome and nuclear sequences. Plant Diversity, 47(4): 561-575.

[11] Ergashov, I., Yusupov, Z., Toshtemirov, J., Mamajonov, S., Akbarova, M., Marupov, A., Esankulov, A. Evaluating the vulnerability of endangered *Allium* species (Amaryllidaceae) to climate change in Central Asia. Analele Universităţii din Oradea, Fascicula Biologie, 32(1): 87-104.

[12] Gammerman, A.F., Kadaev, G.N., Yatsenko-Hmelevsky, A.A., (1990): Medicinal plants. Vuishaya Shkola: Moscow, Russia, 103 p. [In Russian].

[13] Huang, Y.C., Men, W.X., Bian, C., Xue, H.F., Hou, W.J., Song, Y.Y., Xu, L., (2024): The complete chloroplast genome of *Leonurus sibiricus* Linnaeus (Labiatae, Leonurus Miller). Mitochondrial DNA Part B, 9(7): 934-938.

[14] Katoh, K., Standley, D.M., (2013): MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution, 30(4): 772-780.

[15] Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Drummond, A., (2012): Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 28(12): 1647-1649.

[16] Kosman, V.M., Pozharitskaya, O.N., Shikov, A.N., Makarov, V.G., (2002): Extraction of iridoid glycosides from motherwort grass using various solvents. Pharmaceutical Chemistry Journal, 36(2): 42-45.

[17] Kurtz, S., Choudhuri, J.V., Ohlebusch, E., Schleiermacher, C., Stoye, J., Giegerich, R., (2001): REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Research, 29(22): 4633-4642.

[18] Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., (2016): PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular Biology and Evolution, 34(3): 772-773.

[19] Lehwark, P., Greiner, S., (2019): GB2sequin-A file converter preparing custom GenBank files for database submission. Genomics, 111(4): 759-761.

[20] Li, B., Cantino, P.D., Olmstead, R.G., Bramley, G.L.C., Xiang, C.-L., Ma, Z.-H., Tan, Y.-H., Zhang, D.-X., (2016): A large-scale chloroplast phylogeny of the Lamiaceae sheds new light on its subfamilial classification. Scientific Reports, 6(1): 34343.

[21] Lohse, M., Drechsel, O., Bock, R., (2007): OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Current Genetics, 52: 267-274.

[22] Patel, R.K., Jain, M., (2012): NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PloS ONE, 7(2): e30619.

[23] Rambaut, A., (2012): FigTree v1. 4.0. University of Oxford. Available from: http://tree.bio.ed.ac.uk/software/figtree (accessed on 28 October 2021).

[24] Roy, T., Lindqvist, C., (2015): New insights into evolutionary relationships within the subfamily Lamioideae (Lamiaceae) based on pentatricopeptide repeat (PPR) nuclear DNA sequences. American Journal of Botany, 102(10): 1721-1735.

[25] Rozas, J., Ferrer-Mata, J.C., Sanchez-DelBarrio, P., Librado, P., Guirao-Rico, S.E., (2018). DnaSP version 6.12. 03: A software for comprehensive analysis of DNA polymorphism data.

[26] Scheen, A.C., Bendiksby, M., Ryding, O., Mathiesen, C., Albert, V. A., Lindqvist, C., (2010): Molecular phylogenetics, character evolution, and suprageneric classification of Lamioideae (Lamiaceae). Annals of the Missouri Botanical Garden, 97(2): 191-217.

[27] Shang, M., Wang, J., Dai, G., Zheng, J., Liao, B., Wang, J., Duan, B., (2023): Comparative analysis of chloroplast genome and new insights into phylogenetic relationships of *Ajuga* and common adulterants. Frontiers in Plant Science, 14: 1251829.

[28] Song, J.H., Yang, S., Choi, G., Park, I., (2023): Characterization of the complete chloroplast genome of the important herbal medicinal plant, *Thymus quinquecostatus* (Lamiaceae). Journal of Asia-Pacific Biodiversity, 16(1): 140-143.
[29] Tojibaev, K.S., Jang, C.G., Lazkov, G.A., Chang, K.S., Sitpayeva, G.T., Safarov, N., Oh, S.H., (2020): An annotated checklist of endemic vascular plants of the Tian-Shan Mountains in Central Asian countries. Phytotaxa, 464(2): 117-158.
[30] Vvedenskiy, A., (1961): Flora of Uzbekistan. Tashkent: Akademiya nauk Uzbekskoy SSR, 5: 297-306.
[31] Wang, Y., Wang, J., Garran, T.A., Liu, H., Lin, H., Luo, J., Guo, L., (2023): Genetic diversity and population divergence of *Leonurus japonicus* and its distribution dynamic changes from the last interglacial to the present in China. BMC Plant Biology, 23(1): 276.
[32] Yang, Z.Y., Pan, S.L., Huo, K.K., Wu, B.Y., Chao, Z., (2011): Molecular analysis of *Leonurus* species in China based on ITS and *matK* sequences. The American Journal of Chinese Medicine, 39(2): 411-422.
[33] Zahra, N.B., Shinwari, Z.K., (2016): What is done and what has to be done in Lamiaceae, a review of phylogenetics. PeerJ Preprints, 4: e2277v1.
[34] Zhao, F., Chen, Y.-P., Salmaki, Y., Drew, B.T., Wilson, T.C., Scheen, A.-C., Celep, F., Bräuchler, C., Bendiksby, M., Wang, Q., Min, D.-Z., Peng, H., Olmstead, R.G., Li, B., Xiang, C.-L., (2021): An updated tribal classification of Lamiaceae based on plastome phylogenomics. BMC Biology, 19(1): 1-27.
[35] Zhao, W., Guo, L., Yang, Y., Wang, Y., Yang, L., Wei, C., Li, Y., (2022): Complete chloroplast genome sequences of *Phlomis fruticosa* and *Phlomoides strigosa* and comparative analysis of the genus *Phlomis* sensu lato (Lamiaceae). Frontiers in Plant Science, 13: 1022273.

Received: June 26, 2025
Accepted: November 4, 2025
Published Online: November 8, 2025
Analele Universităţii din Oradea, Fascicula Biologie
https://www.bioresearch.ro/revistaen.html
Print-ISSN: 1224-5119
e-ISSN: 1844-7589
CD-ISSN: 1842-6433
University of Oradea Publishing House