Plant Diversity of Central Asia

PDCA ISSN 2181-2403
PDCA ISSN 2181-239X

RESEARCH ARTICLE

Phylogenomics and biogeographic patterns of *Scutellaria* L. (Lamiaceae): southern Asian origins and worldwide dispersal

Bobur Karimov[1], Elena Nikitina[1], Kobiljon A. Bobokalonov[2], Vera A. Cheryomushkina[3], Alexey Yu. Astashenkov[3], Ziyoviddin Yusupov[1]✉, Komiljon Sh. Tojibaev[1]

[1]*Institute of Botany, Academy of Sciences of Uzbekistan, Tashkent, Uzbekistan*

[2]*Institute of Botany, Physiology and Genetics of Plants, National Academy of Sciences of Tajikistan, Dushanbe, Tajikistan*

[3]*Central Siberian Botanical Garden SB RAS, Novosibirsk, Russia*

✉ Corresponding author: Mail: ziyo-nur87@mail.ru; ORCID: https://orcid.org/0000-0003-2278-542X

ABSTRACT

Scutellaria L. (Lamiaceae) is a diverse and pharmacologically important genus with significant species richness in Central Asia. Here, we sequenced and analyzed complete chloroplast genomes of 14 *Scutellaria* species from Uzbekistan and adjacent regions, integrating them with 184 publicly available plastomes to investigate genome evolution, phylogenetic relationships, and biogeographic history. All plastomes exhibited a conserved quadripartite structure, with limited variation in genome size (151.6–152.3 kb), GC content (38.23–38.34%), and gene content. Codon usage showed a consistent bias toward A/T-ending codons. A total of 32–55 SSRs were identified per genome, with elevated variability in several Central Asian taxa, highlighting their utility as molecular markers. Nucleotide diversity analyses revealed highly variable loci, primarily in noncoding regions, suitable for DNA barcoding and phylogenetic inference. Phylogenomic reconstruction strongly supported the monophyly of *Scutellaria* and resolved three major evolutionary lineages. Newly sequenced taxa provided strong support for the cohesion of the *Anaspis* lineage, reinforcing its recognition at the subgeneric level. Ancestral area analyses indicate a southern Asian origin followed by diversification in Central and Eastern Asia and multiple dispersal events into Europe, Africa, and the Americas.

Key words: plastome structure; comparative genomics; codon usage bias; hypervariable regions; lineage diversification; evolutionary history.

Introduction

Scutellaria L. (Lamiaccea; Scutellarioideae) (Paton 1990; Safikhani et al. 2018) is a sub-cosmopolitan genus of 479 accepted species, as reported by Plants of the World Online (POWO, 2025). Taxonomic treatments differ, with several authors treating a number of names as synonyms and recognizing approximately 360 accepted species (Paton 1990a, b; Paton et al. 2016). The primary center of diversity of *Scutellaria* L. lies in Central Asia (Hindukush and Pamir-Alay), with secondary centers in the eastern Mediterranean and the Andes. The genus likely dispersed from Asia to Europe and North America via Beringia during the early–mid Tertiary (Paton, 1989).

Species of *Scutellaria* exhibit therapeutic potential due to diverse bioactive secondary

http://doi.org/10.54981/PDCA/vol4_iss2/a2

Received: 26 March 2025; Accepted: 13 April 2025; Editor: David E. Boufford

Plant Diversity of Central Asia. 2 (2025) 1–22

metabolites, including flavonoids, phenolic acids, phenylethanoid glycosides, and terpenoids, which may act synergistically (Georgieva et al., 2021). Accordingly, they have been widely employed in traditional medicine across Asia and North America (Shen et al., 2021; Rodideal et al. 2024). Moreover, numerous species of *Scutellaria* are economically important in horticulture, owing to their diverse coloration and broad ecological adaptability across habitats from wetlands and coastlines to rocky, alpine regions (Paton 1990a)

Molecular phylogenetic studies have substantially advanced understanding of *Scutellaria* taxonomy and evolution (Zhao et al., 2017; Safikhani et al., 2018; Seyedipour et al., 2020; Salimov et al., 2021; Zhao et al., 2020; Shan et al., 2021). Analyses based on nuclear and plastid markers reveal complex infrageneric relationships, including paraphyly of subgenera (*Scutellaria*, *Apeltanthus*) and non-monophyly of several sections. Despite these inconsistencies, the monophyly of the genus is strongly supported (Li et al., 2016).

The chloroplast genome, due to its maternal inheritance, structural conservation, low mutation and recombination rates, and compact size, serves as an effective system for investigating plant taxonomy and evolutionary relationships (Wang et al., 2024; Bobur et al., 2025; Nikitina et al., 2025; Jiang et al., 2017).

Recent phylogenomic analysis based on extensive chloroplast genome sampling across more than 200 taxa has provided a robust framework and significantly improved resolution of infrageneric relationships in *Scutellaria* (Wang et al., 2024), highlighting the importance of comprehensive genomic data for resolving the evolutionary history of the genus.

Despite recent advances in phylogenomics, including large-scale plastome analyses (Wang et al., 2024), significant gaps remain in the representation of Central Asian taxa, particularly from biodiversity hotspots such as the western Tian Shan and Pamir-Alai. These regions represent key centers of diversification for *Scutellaria*, yet remain underrepresented in global phylogenetic frameworks. Consequently, the evolutionary history and biogeographic patterns of many taxa from this region remain insufficiently resolved.

To address these gaps, the present study employs complete chloroplast genome sequencing and comparative plastome analyses of 14 species of *Scutellaria* from Uzbekistan and adjacent regions. By integrating newly generated plastome data with a broad dataset of publicly available sequences, we aimed to (i) characterize plastome structure and variation, (ii) identify highly variable genomic regions, (iii) reconstruct robust phylogenetic relationships within *Scutellaria* and across Lamiaceae, (iv), critically evaluate the validity of recognizing sect. *Anaspis* at the subgeneric rank and (v) infer biogeographic patterns and centers of diversification.

Materials and Methods

Plant material. Plant material used for DNA extraction consisted of leaf samples obtained from herbarium specimens, with authorization granted by the curator of the TASH National Herbarium of Uzbekistan. The study included 14 species of *Scutellaria* collected from the western Tian Shan, Pamir-Alai, Hissar, and Turkestan mountain ranges in Uzbekistan and neighboring regions. Both old (1929–1991) and more recently collected specimens were incorporated into the dataset (Table 1).

DNA extraction, library preparation, and sequencing. High-quality DNA extraction, library preparation, and sequencing were conducted by Novogene (Beijing, China). Sequencing libraries were prepared using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina® following the manufacturer's instructions. The libraries were sequenced on the Illumina HiSeq 4000 platform to generate paired-end reads (150 bp × 2).

Quality control of sequencing data. Raw sequencing reads were processed using fastp v1.1.0 (Chen et al., 2018) for adapter trimming,

Table 1. Herbarium voucher information for species of *Scutellaria* sequenced in this study.

Species	Subgenus / Section	Voucher ID	Locality	Date	Collector(s)	Identifier
S. urticifolia Juz. & Vved.		–	Tashkent Region, Pskem Range, upper Aksarsay	08 Jun 2023	Natalya Beshko	Natalya Beshko
S. pycnoclada Juz.		TASH066770	western Tian Shan, Pskem Range, Oygaining River valley	10 Aug 2019	Tojibaev et al.	Lazkov (2024)
S. colpodea Nevski		TASH066596	Baysun District, near Pulkhakim gorge, gravelly slopes	08 Aug 1940	Popova & Vasilchenkova	Abdullaeva
S. striatella Gontsch.		TASH069895	SW Pamir-Alai, Baba-tag Mts.	21 Jul 1936	Lepeshkin	–
S. glabrata Vved.	*Euscutellaria / Lupulinaria*	TASH136047	Malguzar Range, upper river Tashkeskensai	02 Jul 2021	Mustafina F. et al.	Gaziev A.
S. tschimganica Juz. / *S. supina* L		TASH068627	Tashkent Alatau, Parkent District	21 Jun 1953	Maylun et al.	–
S. intermedia Popov		TASH068975	western Pamir-Alai, Hissar Range, Aksu River. h-2100	16 Jul 1991	Malsev I.I.	Malsev I.I.
S. phyllostachya Juz.		TASH069402	Angren valley, between Kichik-Yangoklik-say and Ertash-say	18 Jun 1952	Rodina A.	Abdullaeva
S. cristata Popov	*Euscutellaria / Nevskianthe*	TASH066899	Foothills near Bishkent	29 April 1929	M. Kultiasov	–
S. fedtschenkoi Bornm.		TASH131433	Hissar Range, Sangardak River basin	30 May 2021	Gaziev A. et al.	Gaziev A
S. villosissima Gontsch. ex Juz.	*Anaspis*	TASH069953	western Pamir-Alai, Hissar Range, Sangardak River, h-2000	Jun 1984	Malsev I.I.	Xasanov F.O.
S. holosericea Gontsch. ex Juz.		TASH069598	Baysuntau, near Khoja-Gurgur-ata	06 Jul 2011	Turginov O.T.	Malsev I.I.
S. schachristanica Juz.	*Cystaspis*	TASH069802	Turkestan range, Chulpar-tau Mt., upper Koksu River, dry stony slopes	11 Aug 1932	Titov V.S. et al.	Abdullaeva (1981)
S. kuramensis M.N. Abdull. & I.I. Malzev	–	TASH002754 Type	SW Tian Shan, Kuramin Range, upper Lashkerek River	20 Aug 1991	Malsev I.I.	Malsev I.I.

Received: 26 March 2025; Accepted: 13 April 2025; Editor: David E. Boufford

removal of low-quality reads, and filtering of short sequences. More than 99% of reads were retained after filtering, with high base quality (Q30 > 95%) and stable GC content (~35%), indicating high-quality datasets suitable for downstream analyses.

Chloroplast genome assembly. Clean reads were assembled de novo using GetOrganelle (Jin et al., 2020) within the CGAS pipeline (v1.0.1) (Abdullah et al., 2025), with parameters optimized for embryophyte plastomes. Assembly quality was evaluated by mapping reads back to the assembled plastomes, resulting in high average coverage.

Genome annotation. Chloroplast genomes were annotated using PGA (Qu et al., 2019) with a reference-guided approach. Protein-coding genes, transfer RNAs (tRNAs), and ribosomal RNAs (rRNAs) were identified and manually curated where necessary. Gene annotations were standardized across all taxa, and final annotated genomes were exported in FASTA and GenBank formats. The annotation results were manually checked and adjusted in Geneious v.9.0.2 (Kearse et al. 2012). Chloroplast genome structure and gene maps were visualized using Chloroplot (Zheng et al., 2020; https://irscope.shinyapps.io/Chloroplot/), an online tool for graphical representation of plastome organization.

Comparative plastome analysis. Comparative analyses were performed within the CGAS framework (Abdullah et al., 2025). Genome structure (LSC, SSC, IR regions), gene content, genome size, and GC content were analyzed across all species. Codon usage bias (RSCU), amino acid composition, intron variation, single-nucleotide polymorphisms (SNPs), and simple sequence repeats (SSRs) were also examined. Nucleotide diversity (π) was calculated to identify highly variable regions.

Phylogenetic analyses. Phylogenetic relationships were reconstructed based on complete chloroplast genome sequences. Sequence alignment was performed using MAFFT (Katoh & Standley, 2013). Maximum likelihood (ML) analysis was conducted in IQ-TREE (Nguyen et al., 2015), with automatic selection of the best-fit substitution model and branch support assessed using bootstrap analysis. The resulting phylogenetic tree was visualized and annotated using iTOL v6 (Interactive Tree Of Life) (Letunic & Bork, 2021). The dataset included 14 newly sequenced species of *Scutellaria*, 184 taxa of *Scutellaria* retrieved from NCBI, and 70 representative species from major subfamilies of Lamiaceae, including Callicarpoideae (2), Prostantheroideae (1), Nepetoideae (35), Ajugoideae (5), Lamioideae (14), Scutellarioideae (3), Viticoideae (3), Premnoideae (2), Peronematoideae (1), Cymarioideae (1), Tectonoideae (1), and Symphorematoideae (2). Three taxa—*Paulownia tomentosa*, *Paulownia kawakamii*, and *Mazus pumilus*—retrieved from NCBI were used as outgroups to root the phylogenetic tree (Figure S1).

Biogeographic analysis. Ancestral area reconstruction was performed using Statistical Dispersal–Vicariance Analysis (S-DIVA) implemented in the software RASP (Reconstruct Ancestral State in Phylogenies) (Yu et al., 2015). The phylogenetic tree inferred from complete chloroplast genome sequences was used as the input topology. The geographic distribution data for all taxa of *Scutellaria* were compiled primarily from Plants of the World Online (POWO, 2025). Based on the overall distribution patterns of the genus, the species were assigned to discrete biogeographic regions: southern Asia (A), Eastern Asia (B), Central Asia (C), western Asia (D), Southeastern Asia (E), Australia (F), Eastern Europe (G), Southwestern Europe (H), Africa (I), North America (J), Central America (K), and South America (L).

Software and workflow integration. Analyses were conducted using the CGAS pipeline (Abdullah et al., 2025), which integrates multiple bioinformatics tools, including GetOrganelle (Jin et al., 2020), PGA (Qu et al., 2019), MAFFT (Katoh & Standley, 2013), IQ-TREE (Nguyen et al., 2015), Biopython (Cock et al., 2009), and R (R Core Team, 2023),

Received: 26 March 2025; Accepted: 13 April 2025; Editor: David E. Boufford
Plant Diversity of Central Asia. 2 (2025) 1–22

ensuring a standardized and reproducible workflow.

Results

Sequencing and assembly of chloroplast genomes of Scutellaria. A total of 14 species of *Scutellaria* were sequenced and assembled, generating high-quality chloroplast genome datasets (Table 2). The number of clean reads after quality control ranged from 13,519,606 (*S. urticifolia*) to 23,935,534 (*S. schachristanica*), corresponding to 2.01–3.56 Gb of clean bases.

Table 2. Sequencing and assembly statistics of *Scutellaria* chloroplast genomes

Species	Reads After QC	Bases (Gb)	Q20 (%)	Q30 (%)	Mapped Reads	Mapping Rate (%)	Mean Coverage (×)
S. urticifolia	13,519,606	2.01	99.13	95.72	892,348	6.60	874.21
S. pycnoclada	17,066,768	2.53	99.17	95.79	1,172,813	6.87	1,149.99
S. villosissima	18,845,604	2.81	99.23	95.90	4,875,590	25.86	4,774.86
S. kuramensis	15,336,474	2.28	99.19	95.84	1,921,052	12.52	1,877.30
S. schachristanica	23,935,534	3.56	99.16	95.71	8,494,142	35.46	8,298.63
S. colpodea	15,619,190	2.32	99.12	95.59	3,155,851	20.20	3,089.36
S. striatella	20,009,864	2.98	99.19	95.80	11,126,451	55.57	10,901.81
S. holosericea	22,033,526	3.30	99.29	96.16	3,951,662	17.93	3,883.72
S. glabrata	15,499,832	2.31	99.18	95.82	1,464,427	9.45	1,437.68
S. fedtschenkoi	16,576,146	2.48	99.28	96.12	1,958,742	11.81	1,917.04
S. tschimganica	19,884,282	2.96	99.18	95.90	3,797,485	19.09	3,721.58
S. intermedia	23,652,518	3.53	99.12	95.61	2,142,173	9.06	2,101.10
S. phyllostachya	19,803,884	2.95	99.26	96.22	4,453,400	22.48	4,369.65
S. cristata	22,275,254	3.32	99.17	95.83	5,827,958	26.15	5,707.10

The sequencing quality was consistently high across all samples, with Q20 values ranging from 99.12% to 99.29% and Q30 values from 95.59% to 96.22%, indicating excellent base-calling accuracy. The number of reads mapped to chloroplast genomes varied substantially among species, from 892,348 reads (*S. urticifolia*) to 11,126,451 reads (*S. striatella*), with corresponding mapping rates ranging from 6.60% to 55.57%. The highest mapping efficiency was observed in *S. striatella* (55.57%), followed by *S. schachristanica* (35.46%) and *S. cristata* (26.15%), whereas lower mapping rates were detected in *S. urticifolia* (6.60%) and *S. pycnoclada* (6.87%). Mean sequencing coverage of chloroplast genomes was high in all species, ranging from 874.21× (*S. urticifolia*) to 10,901.81× (*S. striatella*). Most species exhibited coverage exceeding 1,000×, ensuring reliable genome assembly and annotation. Notably, *S. striatella*, *S. schachristanica*, and *S. cristata* showed exceptionally high coverage levels (>5,000×).

Genome structure and features of chloroplast genomes of Scutellaria. The complete chloroplast genomes of 14 *Scutellaria* species exhibited a typical quadripartite structure, consisting of a LSC region, a SSC region, and a pair of IR regions (Figure 1).

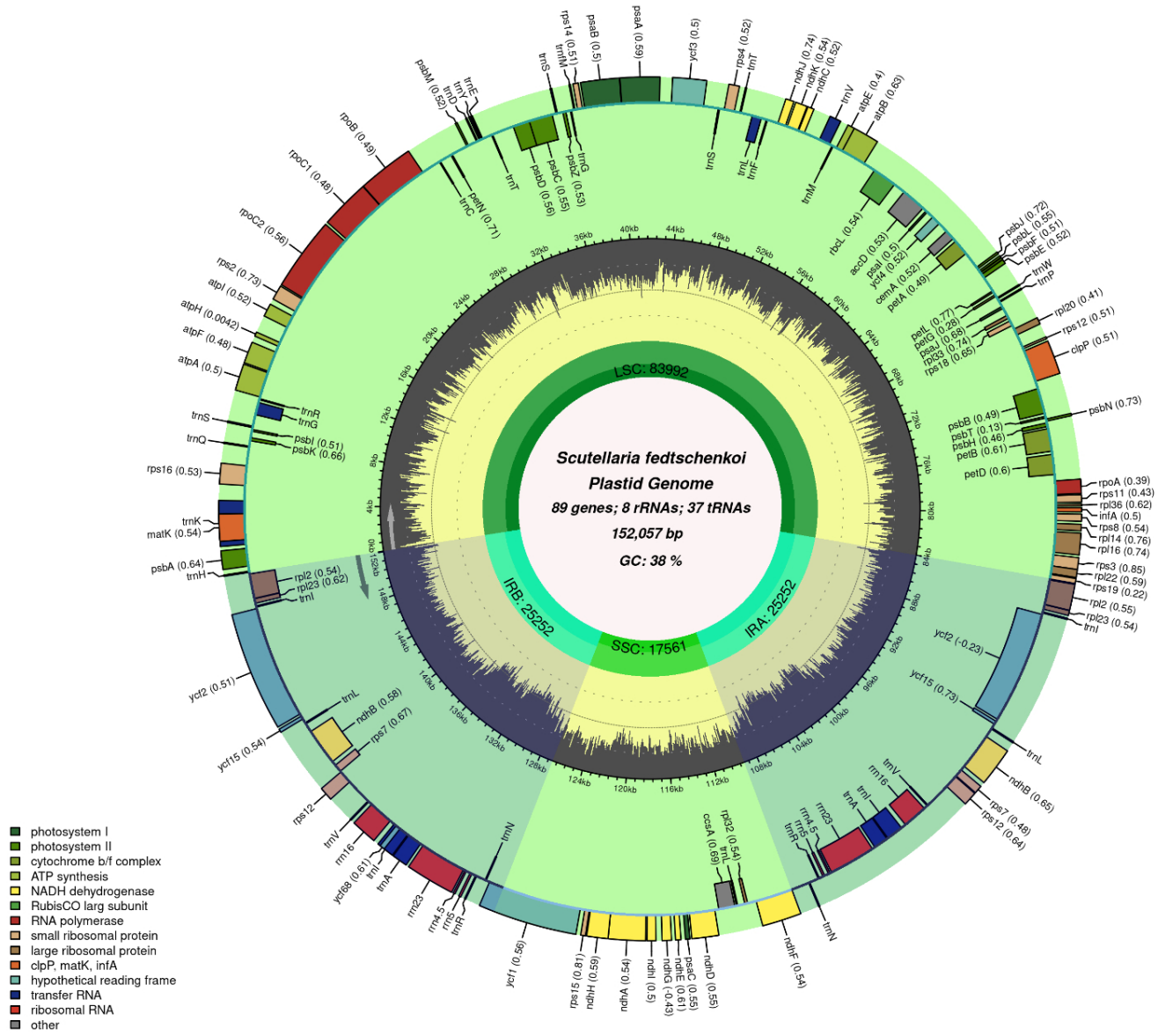

Figure 1. Circular map of the complete plastid genome of *Scutellaria fedschenkoi*. Genes shown outside the circle are transcribed clockwise, while those inside are transcribed counterclockwise. Different colors indicate functional gene categories. The inner circle represents GC content, with darker shading corresponding to higher GC levels. The plastome exhibits a typical quadripartite structure consisting of a large single-copy (LSC) region, a small single-copy (SSC) region, and two inverted repeat regions (IRa and IRb).

Received: 26 March 2025; Accepted: 13 April 2025; Editor: David E. Boufford
Plant Diversity of Central Asia. 2 (2025) 1–22

The total genome size ranged from 151,638 bp in *S. cristata* to 152,273 bp in *S. kuramensis*, indicating a relatively conserved genome structure across the genus (Table 3; Table S1). The LSC region varied from 83,839 bp to 84,453 bp, while the SSC region ranged between 17,319 bp and 17,561 bp. The length of the IR regions was highly conserved, ranging from 25,220 bp to 25,251 bp. Overall, only minor variations were observed in the size of the four genomic regions among the studied species. The complete chloroplast genome length of 189 taxa of *Scutellaria* showed relatively low variation, ranging from 150,867 bp (*S. incana*) to 154,253 bp (*S. dumetorum*). Most species exhibited a genome size clustered between 151,500–152,600 bp.

Table 3. Chloroplast genome characteristics of species of *Scutellaria* sequenced in this study.

Species	Genome Length (bp)				GC (%)				GC (%)			Accession Number
	Complete	LSC	SSC	IR	Complete	LSC	SSC	IR	tRNA	rRNA	CDS	
S. colpodea	151,891	84,092	17,327	25,236	38.33	36.34	32.61	43.62	53.28	55.31	38.36	PZ130064
S. cristata	151,638	83,839	17,319	25,240	38.33	36.33	32.63	43.60	53.21	55.31	38.42	PZ130072
S. fedschenkoi	152,057	83,993	17,561	25,251	38.23	36.25	32.39	43.57	53.14	55.33	38.34	PZ130068
S. glabrata	151,771	83,966	17,335	25,235	38.34	36.36	32.60	43.61	53.29	55.31	38.35	PZ130067
S. holosericea	151,891	83,847	17,551	25,246	38.26	36.29	32.30	43.60	53.17	55.33	38.32	PZ130066
S. intermedia	151,765	83,961	17,342	25,231	38.33	36.35	32.57	43.61	53.28	55.31	38.35	PZ130070
S. kuramensis	152,273	84,453	17,338	25,241	38.29	36.28	32.58	43.61	53.21	55.31	38.41	PZ130062
S. phyllostachya	151,814	83,989	17,355	25,235	38.34	36.36	32.58	43.61	53.21	55.31	38.44	PZ130071
S. pycnoclada	151,781	83,962	17,357	25,231	38.34	36.35	32.58	43.61	53.21	55.31	38.45	PZ130060
S. schachristanica	151,772	84,002	17,330	25,220	38.32	36.34	32.59	43.60	53.28	55.29	38.36	PZ130063
S. striatella	152,168	84,358	17,322	25,244	38.29	36.28	32.62	43.61	53.21	55.31	38.41	PZ130065
S. tschimganica	151,789	83,962	17,355	25,236	38.34	36.35	32.59	43.62	53.21	55.31	38.45	PZ130069
S. urticifolia	151,777	83,977	17,342	25,229	38.34	36.35	32.58	43.62	53.21	55.31	38.42	PZ130059
S. villosissima	151,937	83,886	17,552	25,246	38.25	36.29	32.37	43.56	53.14	55.29	38.34	PZ130061

GC: guanine-cytosine content; LSC: large single-copy region; SSC: small single-copy region; IR: inverted repeat region; Complete: complete chloroplast genome; tRNA: transfer RNA; rRNA: ribosomal RNA; CDS: protein-coding gene.

Received: 26 March 2025; Accepted: 13 April 2025; Editor: David E. Boufford
Plant Diversity of Central Asia. 2 (2025) 1–22

The overall GC content of the chloroplast genome of 14 species of *Scutellaria* ranged from 38.23% to 38.34%, showing high conservation among species. As expected, the GC content was unevenly distributed across regions, with the highest values observed in the IR regions (43.56–43.62%), followed by the LSC region (36.25–36.36%), and the lowest in the SSC region (32.30–32.63%). The elevated GC content in the IR regions is associated with the presence of ribosomal RNA genes. All chloroplast genomes encoded a consistent set of genes, including protein-coding genes (CDS), transfer RNA (tRNA), and ribosomal RNA (rRNA) genes. The GC content of different gene categories was similar across species, with tRNA genes showing relatively higher GC content (53.14–53.29%) and rRNA genes exhibiting the highest values (55.29–55.33%), while protein-coding genes had comparatively lower GC content (38.32–38.45%).

All sequenced chloroplast genomes of *Scutellaria* contained a total of 132 genes (Table 4). Several genes were duplicated in the IR regions, including all rRNA genes (*rrn16, rrn23, rrn4.5,* and *rrn5*), as well as selected tRNA and protein-coding genes such as *rpl2, rpl23, rps7, ndhB, ycf2* and *ycf15*. A number of genes contained introns. Most intron-containing genes (e.g., *rpl16, rps16, rpoC1, atpF, petB, petD, ndhA,* and *ndhB*) possessed a single intron, whereas *clpP* and *ycf3* contained two introns. The *rps12* gene was trans-spliced. Intron lengths were highly conserved among species, with only slight variation. The intron of *rps16* ranged from 890 to 911 bp, *atpF* from 674 to 676 bp, and *rpoC1* from 768 to 791 bp. The two introns of *ycf3* were approximately 709–716 bp and 701–705 bp, respectively, while those of *clpP* ranged from 608–617 bp and 728–733 bp. The intron of *ndhA* showed relatively larger variation (1016–1030 bp), whereas *ndhB* introns were highly conserved (~679 bp). In some species (e.g., *S. colpodea, S. glabrata, S. intermedia,* and *S. schachristanica*), a pseudogene of *ycf1* was detected, and the gene *pabJ* was excluded, whereas other species retained a complete set of conserved open reading frames including *ycf68*.

Table 4. Gene content of the chloroplast genome of *Scutellaria fedschenkoi*

Category for genes	Group of genes	Name of genes	Total
Self-replication	Large subunit of ribosome	*rpl14*, *rpl16**, *rpl2**[a], *rpl20*, *rpl22*, *rpl23*[a], *rpl32*, *rpl33*, *rpl36*	11
	Small subunit of ribosome	*rps11*, *rps12***, *rps14*, *rps15*, *rps16**, *rps18*, *rps19*, *rps2*, *rps3*, *rps4*, *rps7*[a], *rps8*	13
	DNA dependent RNA polymerase	*rpoA*, *rpoB*, *rpoC1**, *rpoC2*	4
	rRNA genes	*rrn16*[a], *rrn23*[a], *rrn4.5*[a], *rrn5*[a]	8
	tRNA genes	*trnA-UGC**[a], *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnG-GCC*, *trnG-UCC**, *trnH-GUG*, *trnI-CAU*[a], *trnI-GAU**[a], *trnK-UUU**, *trnL-CAA*[a], *trnL-UAA**, *trnL-UAG*, *trnM-CAU*, *trnN-GUU*[a], *trnP-UGG*, *trnQ-UUG*, *trnR-ACG*[a], *trnR-UCU*, *trnS-GCU*, *trnS-GGA*, *trnS-UGA*, *trnT-GGU*, *trnT-UGU*, *trnV-GAC*[a], *trnV-UAC**, *trnW-CCA*, *trnY-GUA*, *trnfM-CAU*	37
Photosynthesis	Photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*	5

Received: 26 March 2025; Accepted: 13 April 2025; Editor: David E. Boufford

	Photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN (pbf1)*, *psbT*, *psbZ (lhbA)*	15
	NADPH dehydrogenase	*ndhA**, *ndhB*[a]*, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*	12
	Cytochrome b/f complex	*petA*, *petB**, *petD**, *petG*, *petL*, *petN*	6
	Subunits of ATP synthase	*atpA*, *atpB*, *atpE*, *atpF**, *atpH*, *atpI*	6
	Large subunit of Rubisco	*rbcL*	1
	Photosynthesis assembly genes	*ycf3 (pafI)***, *ycf4 (pafII)*	2
Other genes	Protease	*clpP***	1
	Maturase	*matK*	1
	Envelop membrane protein	*cemA*	1
	Subunit of Acetyl-CoA-carboxylase	*accD*	1
	C-type cytochrome synthesis gene	*ccsA*	1
	Translation initiation factor	*infA*	1
	Conserved open reading frames	*ycf1*, *ycf15[a]*, *ycf2[a]*, *ycf68*	6
		Total number of genes	132

Note: * and ** indicate genes containing one and two introns, respectively; [a] indicates duplicated genes in inverted repeat (IR) regions. The *rps12* gene is a trans-spliced gene and is not marked as duplicated despite appearing in multiple locations.

Codon usage bias. Codon usage patterns were highly conserved across all examined chloroplast genomes of *Scutellaria* (Figure 2; Table S2). Most amino acids exhibited a clear preference for synonymous codons ending in A or T, whereas codons with G or C at the third position generally showed lower relative synonymous codon usage (RSCU) values. For instance, alanine preferentially used GCT and GCA, while GCC and GCG were less frequently utilized. Arginine showed a strong bias toward AGA, whereas CGC and CGG were underrepresented. A similar trend was observed in leucine, with TTA being more frequently used than CTC and CTG. This bias toward A/T-ending codons was consistent across multiple amino acids, including phenylalanine (TTT > TTC), asparagine (AAT > AAC), and lysine (AAA > AAG). Stop codon usage showed minor variation among species, with TAA being the most frequently used, followed by TAG and TGA.

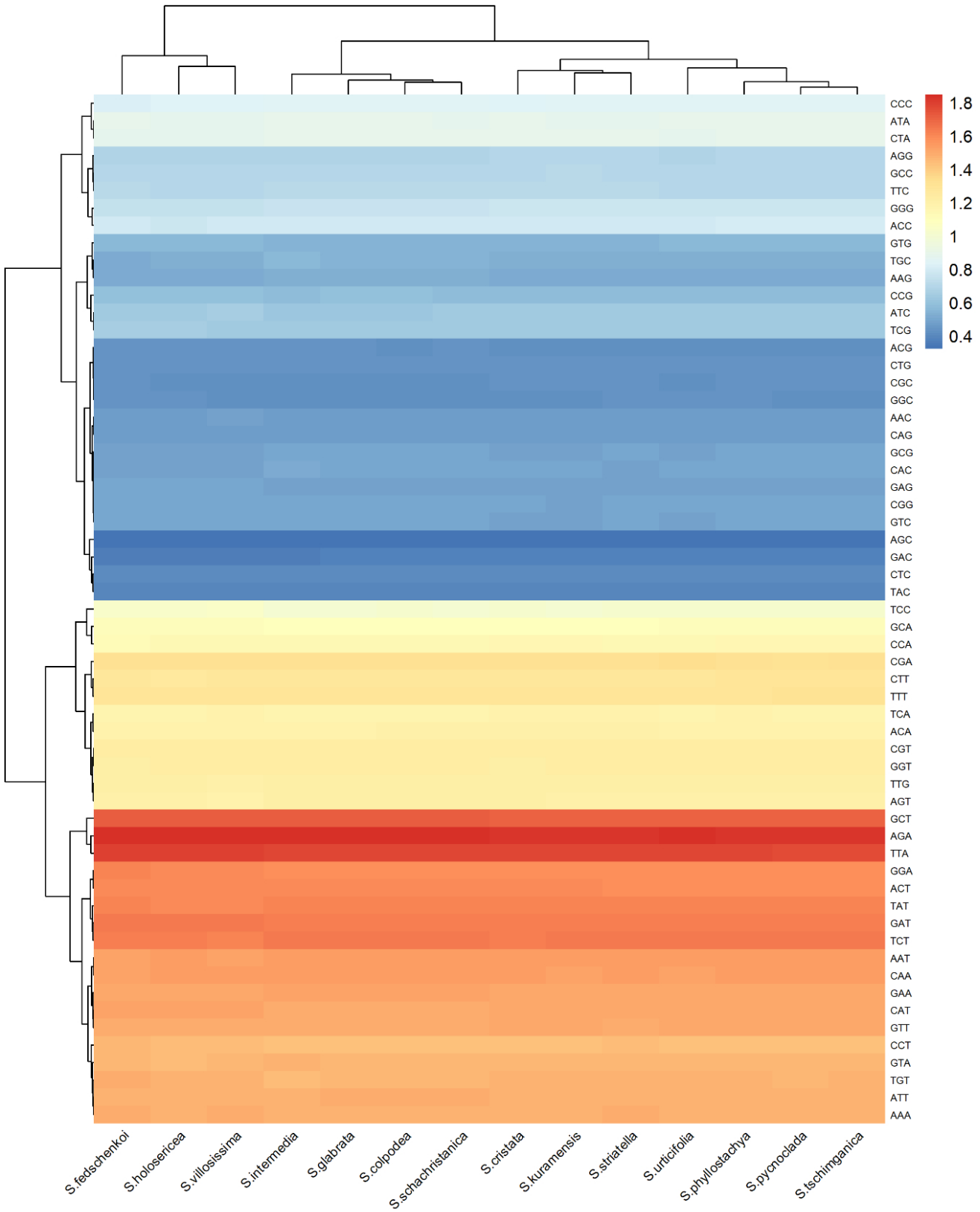

Figure 2. Codon usage bias in the protein-coding genes of plastid genomes of *Scutellaria*

Amino acid usage. The analysis of amino acid composition across all examined taxa of *Scutellaria* revealed a highly conserved pattern with minimal interspecific variation

Received: 26 March 2025; Accepted: 13 April 2025; Editor: David E. Boufford

(Figure 3; Table S3) Leucine (L) was consistently the most abundant amino acid (approximately 10.57–10.72%), followed by isoleucine (I) (~8.35–8.52%) and serine (S) (~7.78–7.92%), together constituting a major proportion of the total amino acid content. Moderately abundant amino acids included glycine (G) (~6.80–6.89%), arginine (R) (~6.18–6.36%), and phenylalanine (F) (~5.66–5.79%), while alanine (A), valine (V), threonine (T), glutamate (E), and lysine (K) showed relatively stable intermediate frequencies around ~5.0–5.4%. Lower abundance amino acids included asparagine (N) (~4.65–4.76%), aspartate (D) (~3.98–4.13%), proline (P) (~4.23–4.32%), and tyrosine (Y) (~3.59–3.68%), whereas cysteine (C) (~1.13–1.16%) and tryptophan (W) (~1.71–1.74%) were consistently the least represented. The observed variation among species was extremely limited, generally within 0.2–0.3%, and even subspecies exhibited nearly identical profiles, suggesting strong compositional stability across taxonomic levels.

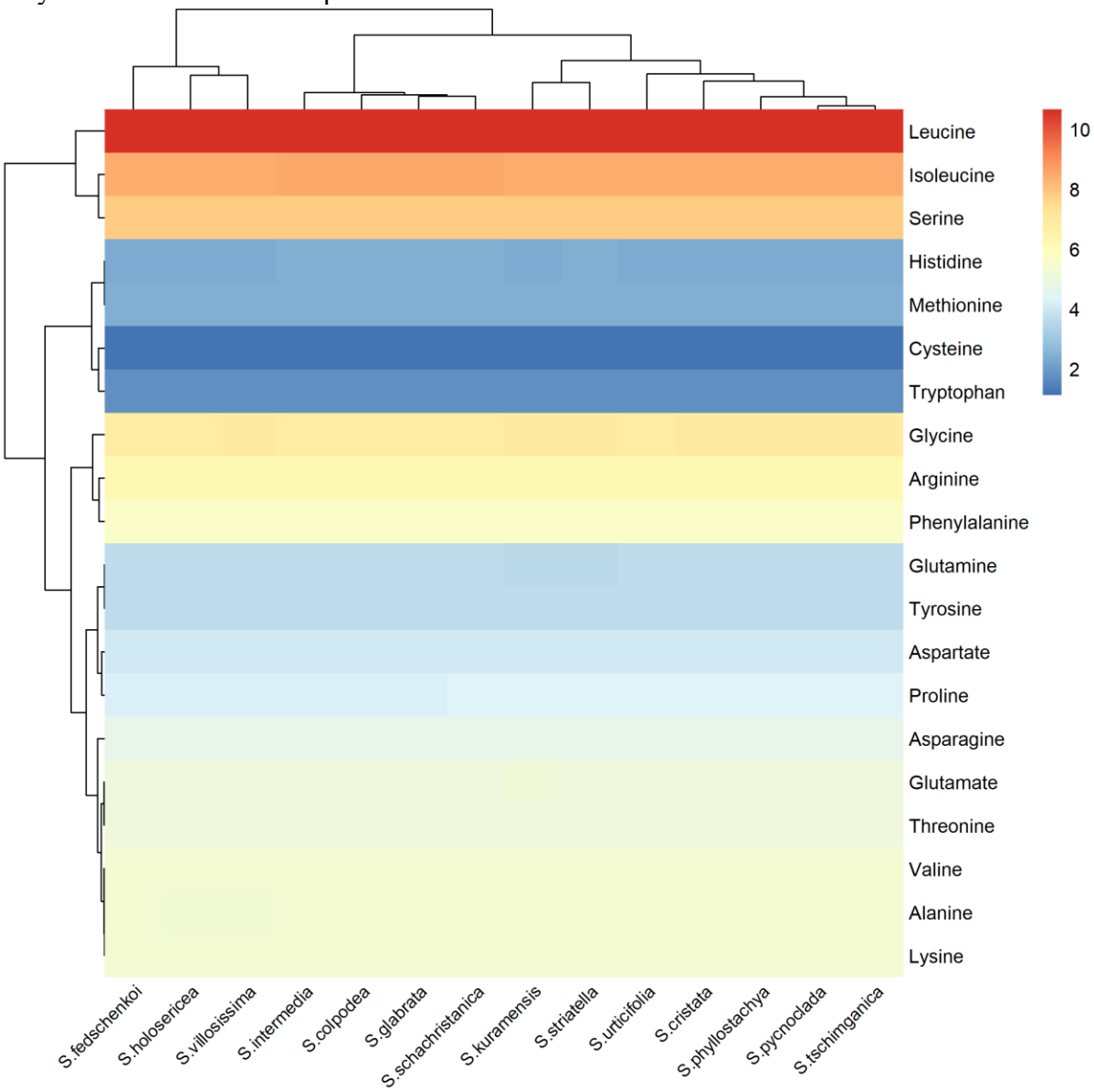

Figure 3. Comparative amino acid composition of plastid-encoded proteins in *Scutellaria*

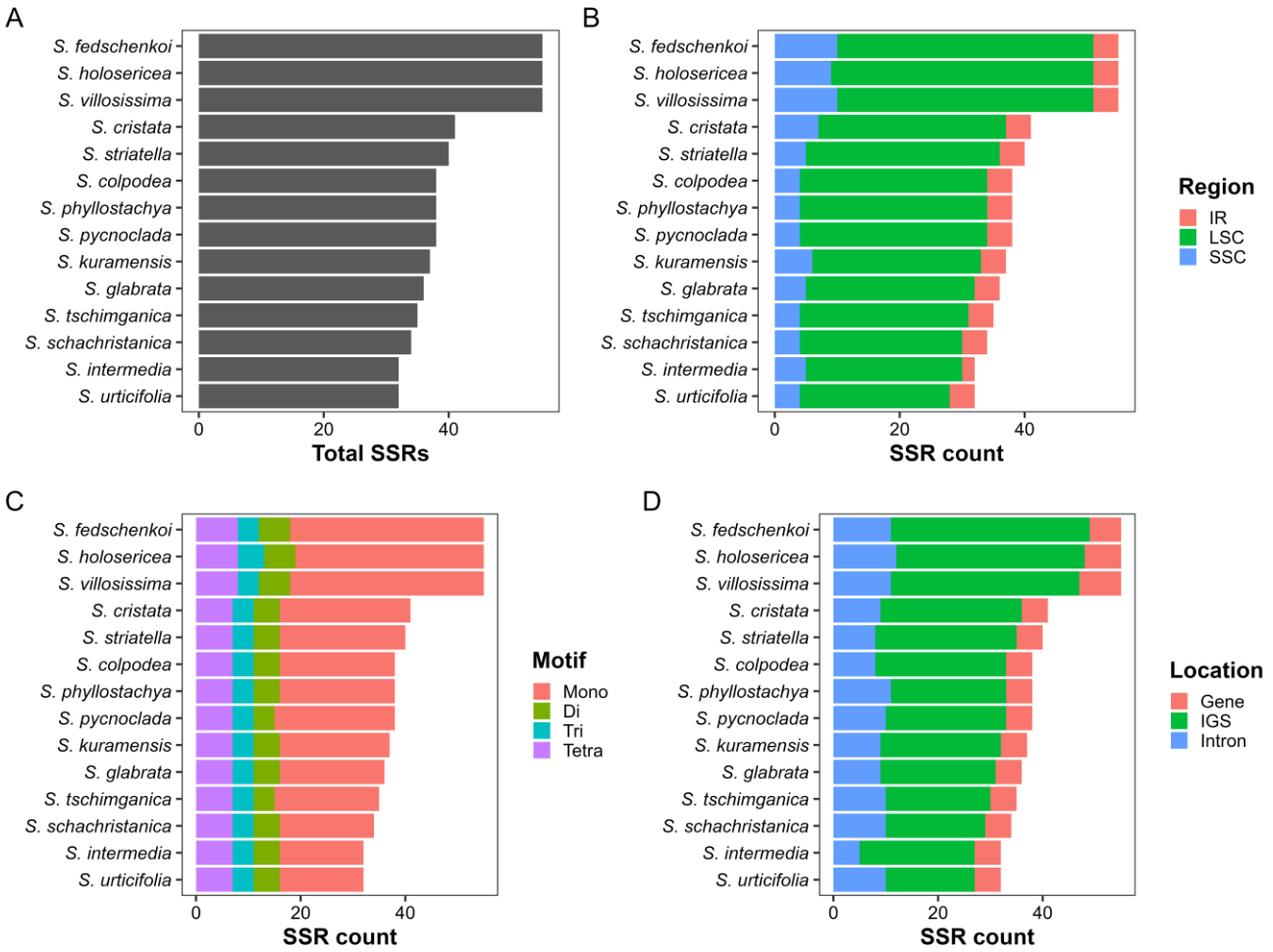

Figure 4. Comparative distribution of simple sequence repeats (SSRs) among plastomes of *Scutellaria*. (A) Total number of SSR loci detected in each species. (B) Distribution of SSRs across plastome regions. (C) Composition of SSR motif types. (D) Genomic distribution of SSRs among intergenic spacer (IGS), intron, and gene regions.

Simple sequence repeats. A total of 32 to 55 simple sequence repeats (SSRs) were identified across the 189 analyzed chloroplast genomes of *Scutellaria* (Figure 4; Table S4). Most species contained 37–45 SSR loci, indicating relatively conserved SSR abundance within the genus. However, several species exhibited higher SSR counts, including *S. fedschenkoi*, *S. holosericea*, *S. villosissima*, and *S. petiolata* (55 SSRs), whereas lower values were observed in *S. adsurgens* and *S. urticifolia* (32 SSRs). SSRs were unevenly distributed across the plastome regions. The majority, ranging from 24 to 44 loci, were located in the LSC region, followed by the SSC region with 3 to 13 loci. In contrast, the inverted repeat IR regions contained the fewest SSRs (0 to 6 loci) and were completely devoid of SSRs in several species (e.g., *S. tuberifera*, *S. purpurascens*, *S. calcarata*). Mononucleotide repeats were the most abundant SSR type across all species, ranging from 15 to 37 loci, followed by dinucleotide repeats (3–8 loci), trinucleotide repeats (3–6 loci), and tetranucleotide repeats (5–10 loci). Pentanucleotide and hexanucleotide repeats were rare, typically occurring in 0–2 loci, and were restricted to a limited number of taxa. Hexanucleotide repeats were extremely rare and found only in a few taxa (*S. flocculosa*, *S. minor*, *S. racemosa*, *S. rehderiana*, *S. tenax*, and *S. tienchuanensis*), each typically represented by a

Received: 26 March 2025; Accepted: 13 April 2025; Editor: David E. Boufford

single locus Most SSRs were located in intergenic spacer (IGS) regions, with counts ranging from 17 to 39 loci. Fewer SSRs were found in introns (3–13 loci) and protein-coding genes (4–10 loci), while SSRs were nearly absent in tRNA and rRNA genes.

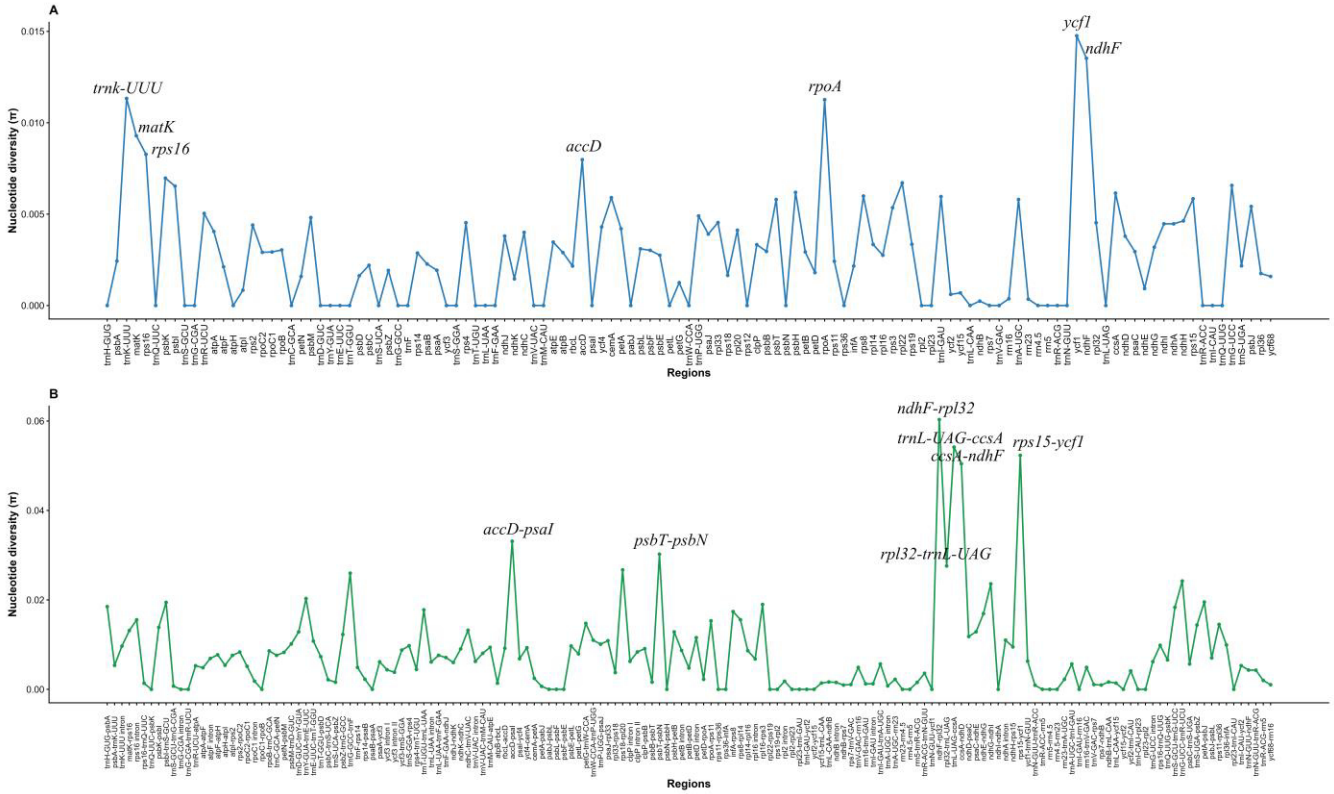

Figure 5. Comparative nucleotide diversity (π) across coding and non-coding regions of plastid genomes in 14 newly sequenced Central Asian species of *Scutellaria*.

Nucleotide diversity patterns across coding and non-coding regions of Scutellaria plastomes. A comprehensive nucleotide diversity (π) analysis across 14 plastomes of *Scutellaria*, covering 283 regions (121 genes, 22 introns, and 140 intergenic spacers), revealed pronounced heterogeneity in sequence variation (Figure 5). Protein-coding regions were generally conserved, exhibiting low nucleotide diversity. Among them, *ycf1* (π = 0.01476) and *ndhF* (π = 0.01353) showed comparatively higher variability, followed by *trnK-UUU* (π = 0.01133) and *rpoA* (π = 0.01126), whereas *matK*, *rps16*, and *accD* displayed moderate but still low diversity (π = 0.00798–0.00929). In contrast, intergenic spacers exhibited substantially higher nucleotide diversity, with several regions identified as mutational hotspots. The most variable regions included *ndhF–rpl32* (π = 0.06032), *trnL-UAG–ccsA* (π = 0.05416), *rps15–ycf1* (π = 0.05232), and *ccsA–ndhD* (π = 0.05045), followed by *accD–psaI* (π = 0.03311) and *psbT–psbN* (π = 0.03022).

Phylogenetic relationships and biogeographic patterns of Scutellaria. The phylogenetic analysis based on complete chloroplast genomes resolves 12 subfamilies of Lamiaceae, including *Scutellarioideae*, as well-supported monophyletic lineages (Figure 6; Figure S1). Following the outgroups (*Mazaceae* and *Paulowniaceae*), Prostantheroideae and Callicarpoideae occupy basal positions. This was followed by a sequential radiation of subfamilies Nepetoideae, Symphorematoideae, Viticoideae, Tectonoideae, Ajugoideae, Premnoideae, Peronematoideae, and Cymarioideae. Lamioideae and Scutellarioideae were also recovered as distinct and strongly supported clades, with Scutellarioideae showing pronounced internal structuring into multiple well-defined clades, reflecting substantial diversification within *Scutellaria*.

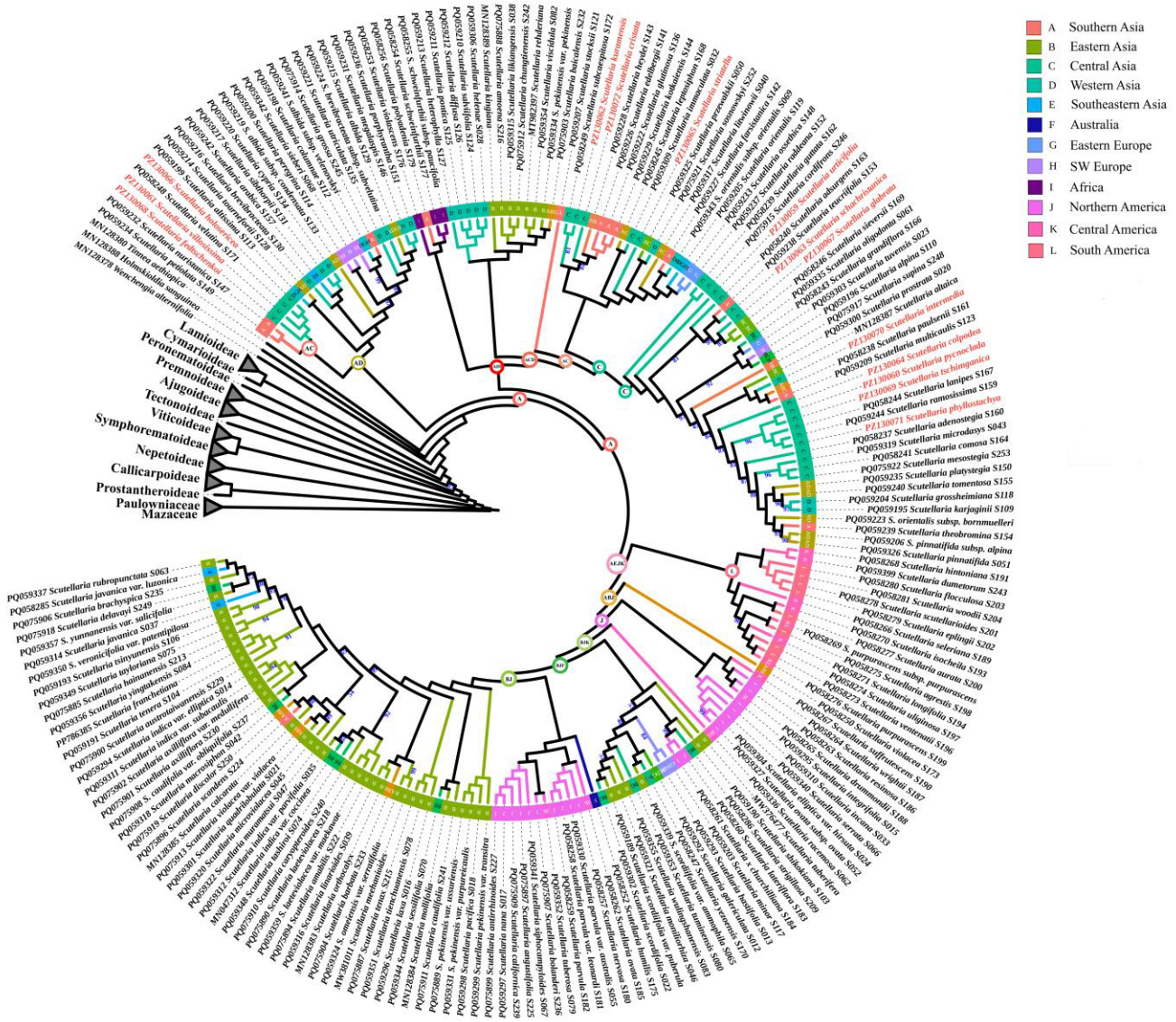

Figure 6. Circular phylogenetic tree and biogeographic reconstruction of *Scutellaria* inferred from plastome data. Newly sequenced species are given in red.

Newly sequenced taxa (*S. holosericea*, *S. villosissima*, and *S. fedtschenkoi*) were recovered within a well-supported clade together with *S. velutina*, *S. nuristanica*, and *S. petiolata*, supporting their placement within subgenus *Anaspis*. In contrast, *S. schachristanica*, previously assigned to subgenus *Cystaspis*, was phylogenetically nested within members of sect. *Lupulinaria*, indicating inconsistency with its current taxonomic position. Furthermore, species assigned to sect. *Nevskianthe* (*S. colpodea*, *S. striatella*, and *S. cristata*) did not form a monophyletic group, but were instead embedded within sect. *Lupulinaria*.

Ancestral area reconstruction based on 198 taxa strongly supported southern Asia (A) as the primary center of origin (Figure 6; Figure S2). The basal nodes were consistently inferred as southern Asian, with the earliest diverging taxa including *S. nuristanica* and *S. petiolata*, both confined to this region. These early lineages were followed by species distributed in Central Asia (C), such as *S. fedtschenkoi*, *S. villosissima*, *S. velutina*, and *S. holosericea*, indicating an initial phase of regional diversification. Subsequent ancestral reconstructions (AC, AD, DH, DI) reflected early dispersal events from southern Asia into

Received: 26 March 2025; Accepted: 13 April 2025; Editor: David E. Boufford

Central and western Asia, and further expansion from western Asia into Europe and Africa. The phylogenetic topology supports a southern Asian origin of *Scutellaria*, followed by stepwise geographic expansion, with diversification occurring along a western dispersal route.

The second line of evolution showed a clear basal-to-derived sequence beginning with an ADI ancestral state, indicating an early connection among southern Asia (A), western Asia (D), and Africa (I). This was followed by a transition to ACD and then AC, reflecting movement toward Central Asia (C) while maintaining links with southern and western Asia. The lineage subsequently stabilized in a predominantly Central Asian (C) state, where the majority of taxa diversified, including *S. comosa*, *S. microdasy*, *S. adenostegia*, *S. ramossissima*, and *S. lanipes*. From this Central Asian core, further diversification involved more complex combinations such as BCH, BC, and CGH, indicating expansion into eastern Asia (B), eastern Europe (G), and southwestern Europe (H). The topology suggests an initial widespread ancestral range involving southwestern Asia and Africa, followed by consolidation in Central Asia and subsequent expansion into Eurasian regions.

The third line begins with a basal AEJK ancestral state, indicating an early connection among southern Asia (A), southeastern Asia (E), North America (J), and Central America (K). This was followed by transitions to ABJ, then J, and subsequently BJK, BFJ, and BJ, reflecting a shift toward eastern Asia (B) and North America (J) with additional links to Central America (K) and Australia (F). A large portion of the clade is dominated by eastern Asia (B), where extensive diversification occured, including species such as *S. scordifolia*, *S. moniliorhiza*, *S. pekinensis*, and *S. tuberifera*. From this eastern Asian core, the lineage expanded into North America (J), forming well-supported sublineages with taxa such as *S. californica*, *S. angustifolia*, *S. bolanderi*, and *S. nana*. Further diversification occured in Central America (K) and South America (L), including species such as *S. purpurascens*, *S. uliginosa*, *S. aurata*, and *S. dumetorum*. Collectively, these patterns highlight eastern Asia as a principal center of diversification and a major source of subsequent intercontinental dispersal within the lineage

Discussion

In this study, we sequenced and assembled the complete chloroplast genomes of 14 Central Asian species of *Scutellaria*, revealing a high degree of similarity to previously published plastomes (Zhao et al. 2020; Shan et al. 2021; Wang et al. 2024). The slight size variation observed among the taxa of *Scutellaria* (~150.8–154.2 kb across 189 taxa) was consistent with previous findings, with the variation mainly caused by large indels (insertions/deletions) in the noncoding regions (Zhao et al. 2020). The narrow GC content range (38.19–38.54%) further confirmed high conservation among the plastomes of *Scutellaria*. The strong preference for A/T-ending codons in the plastomes of *Scutellaria* was consistent with patterns observed across Lamiaceae and other angiosperms (Niu et al. 2023).

SSRs are widely applied as molecular markers for species identification, genetic diversity assessment, and population genetic analyses (Karimov et al., 2024). In this study, 198 taxa ranged 32-55 SSRs. *S. fedschenkoi*, *S. holosericea*, *S. villosissima* (Central Asian taxa), and *S. petiolata* (southern Asian taxon) exhibited the highest number of SSR loci, indicating relatively greater plastome variability compared to other species within the genus. The newly identified SSRs therefore represent valuable genetic resources for evaluating diversity in medicinally important species of *Scutellaria* and provide useful markers for phylogenetic analyses at both species and genus levels.

DNA barcoding is a key method for species identification and detecting new or cryptic taxa, supporting food safety and

exemplar selection (Vassou et al. 2015). Despite the overall conservation of plastid genomes, several loci in 14 Central Asian species of *Scutellaria* showed relatively elevated nucleotide diversity. Loci with high nucleotide diversity ($\pi > 0.025$) include *ndhF–rpl32* ($\pi = 0.06032$), *trnL-UAG–ccsA* ($\pi = 0.05416$), *rps15–ycf1* ($\pi = 0.05232$), and *ccsA–ndhD* ($\pi = 0.05045$), as well as moderately variable regions such as *accD–psaI* ($\pi = 0.03311$) and *psbT–psbN* ($\pi = 0.03022$). These intergenic spacers exhibited substantially higher variation than coding regions. Such hypervariable regions represent promising candidates for developing high-resolution DNA barcodes in Central Asian *Scutellaria*, particularly for resolving closely related or morphologically similar taxa. In contrast, more conserved coding loci (e.g., *matK*, *rbcL*) remain useful for higher-level taxonomic resolution but may lack sufficient variability for species-level discrimination within the genus. Consistent with these findings, Wang et al. (2024) identified 10 hypervariable regions ($\pi > 0.025$), including *trnK–rps16*, *trnC–petN*, *petN–psbM*, *accD–psaI*, *petA–psbJ*, *ndhF*, *rpl32–trnL*, *ccsA–ndhD*, *rps15–ycf1*, and *ycf1*, with only two located in coding regions, supporting previous results (Zhao et al. 2020).

Phylogenetic relationships and biogeographic patterns in Scutellaria. The plastome-based phylogenetic reconstruction presented here provides a robust and highly resolved framework for understanding evolutionary relationships within *Scutellaria*, and is broadly consistent with previous studies supporting the monophyly of the genus (Zhao et al. 2017; Zhao et al. 2020; Safikhani 2018; Salimov et al., 2021; Wang et al. 2024). The improved resolution and strong support across nodes highlight the utility of complete chloroplast genomes in resolving complex phylogenetic relationships and reveal pronounced internal structuring, indicating extensive diversification within the genus.

Moreover, the lack of congruence between traditional morphology-based classifications and molecular phylogeny, particularly at subgeneric and sectional levels, was consistent with earlier reports and suggests the need for taxonomic reassessment (Wang et al. 2024). Recent taxonomic treatment by Wang et al. (2024) proposed elevating sect. *Anaspis* (excluding *S. kingiana*) to subgeneric rank, in agreement with its earlier recognition as a subgenus in *Flora USSR*. However, their conclusions were constrained by the lack of molecular data for several critical taxa, limiting the robustness of phylogenetic inference. In the present study, the inclusion of newly sequenced species (*S. holosericea*, *S. villosissima*, and *S. fedtschenkoi*) provided additional resolution. Those taxa form a strongly supported clade together with *S. velutina*, *S. nuristanica*, and *S. petiolata*, thereby offering clear molecular evidence for the cohesion of this lineage. This finding reinforces the recognition of sect. *Anaspis* at the subgeneric level and supports its phylogenetic distinctiveness. Conversely, our findings reveal notable inconsistencies with existing classifications. *S. schachristanica*, previously assigned to subgenus *Cystaspis*, was recovered within sect. *Lupulinaria*, indicating that its current placement is not phylogenetically justified. Similarly, species attributed to sect. *Nevskianthe* (*S. colpodea*, *S. striatella*, and *S. cristata*) failed to form a monophyletic group and were instead nested within sect. *Lupulinaria*.

According to POWO (2025), *S. tschimganica* Juz. is treated as a synonym of *S. supina*. However, our phylogenomic results did not support this taxonomic treatment. According to current taxonomic treatments, *S. nuristanica* is considered a synonym of *S. petiolata.* However, the observed chloroplast genome similarity between these taxa is relatively low (99.16%) compared to other closely related species pairs, suggesting that they may warrant further taxonomic reassessment rather than being strictly treated as conspecific.

Within *Scutellaria*, the recovery of three major evolutionary lineages suggests a complex history of diversification. Our results strongly support southern Asia as the center of origin, consistent with the presence of early-diverging taxa restricted to this region. From this ancestral

area, diversification proceeded through multiple dispersal routes. The first lineage reflects a stepwise westward expansion through Central and western Asia toward Europe and Africa, aligning with previously proposed Irano-Turanian diversification patterns and rapid radiation (Salimov et al. 2021)

The second lineage indicates an initially widespread ancestral range spanning southern and western Asia and Africa, followed by subsequent consolidation and extensive diversification in Central Asia, with further expansion into eastern Asia, eastern Europe, and Southwestern Europe, supporting its role as a secondary center of radiation.

The third lineage exhibits a distinct biogeographic pattern characterized by early connections among South Asia, Southeast Asia, and the Americas. This was followed by substantial diversification in East Asia and subsequent dispersal events into North, Central, and South America. The inferred East Asia–North America disjunction suggests a likely migration route via the Beringian land bridge (Paton, 1989), consistent with patterns reported for the tribe Elsholtzieae (Lamiaceae) (Li et al. 2017; Wang et al. 2024).

Conclusion

This study provides a comprehensive plastome-based framework for understanding genome evolution, phylogenetic relationships, and biogeographic history within *Scutellaria*. The newly sequenced Central Asian taxa confirmed a high level of structural conservation across plastomes, with limited variation in genome size, gene content, and GC composition. Nevertheless, the identification of SSRs and highly variable loci, particularly in intronic and intergenic regions, highlighted their utility as effective molecular markers for species identification, phylogenetic reconstruction, and future population-level studies. Phylogenomic analyses strongly supported the monophyly of *Scutellaria* and significantly improved the resolution of infrageneric relationships. The inclusion of previously unsampled Central Asian taxa provided robust evidence for the cohesion of the *Anaspis* lineage, supporting its recognition at the subgeneric level, while also revealing inconsistencies in the current classification of several sections and species. The findings emphasize the need for taxonomic reassessment within the genus. Biogeographic reconstruction indicates a southern Asian origin of *Scutellaria*, followed by complex diversification and multiple dispersal events across Eurasia, Africa, and the Americas. Central Asia emerges as a major secondary center of diversification. Overall, this work has provided a foundational genomic resource for *Scutellaria* and contributes to a more comprehensive understanding of its evolutionary history. Future studies should expand taxon sampling and incorporate nuclear genomic data to further resolve phylogenetic relationships, assess hybridization, and refine species delimitation within the genus.

Data Availability Statement

All data supporting the findings of this study are included in this article and its Supplementary materials (https://doi.org/10.5281/zenodo.19494253)

Tables S1. Comparative plastome genome size, regional structure, GC content, and GenBank accession numbers of species of *Scutellaria*

Tables S2. Relative synonymous codon usage (RSCU) values of protein-coding genes in the plastomes of species of *Scutellaria*.

Tables S3. Amino acid composition of plastid protein-coding sequences in species of *Scutellaria*.

Tables S4. Characteristics and genomic distribution of simple sequence repeats (SSRs) identified in the plastome of *Scutellaria*

Author Contributions

Bobur Karimov: Conceptualization, methodology, formal analysis, data curation, visualization, and writing—original draft. Elena Nikitina: Investigation, data curation, and writing—review and editing. Kobiljon Bobokalonov: Sample collection and investigation. Vera Alekseevna Cheryomushkina: Sample collection and investigation. Alexey Yu. Astashenkov: Sample collection and investigation. Ziyoviddin

Yusupov: Conceptualization, supervision, project administration, and writing—review and editing. Komiljon Sh. Tojibaev: Supervision, funding acquisition, and writing—review and editing.

Declaration of competing interests

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

Acknowledgments

The authors sincerely thank Dr. David E. Boufford (Senior Researcher at Harvard University, USA) for his valuable assistance in improving the English of this manuscript.

Funding

This study was conducted within the framework of the project "Digital Nature: Development of a Digital Platform for the Flora of Central Uzbekistan" (2025–2029), implemented by the Institute of Botany, Academy of Sciences of the Republic of Uzbekistan

References

Abdullah, Yan, R., & Tian, X. (2025). CGAS (Chloroplast genome analysis Suite): an automated python pipeline for comprehensive comparative Chloroplast genomics. *iMetaOmics*, e70093. https://doi.org/10.1002/imo2.70093

Bobur, K., Jorabek, T., Sarviniso, M., & Ziyoviddin, Y. (2025). First complete chloroplast genome of *Thymus seravschanicus*: Insights into genome structure and phylogenetic relationships within Lamiaceae. *Journal of Asia-Pacific Biodiversity*, *18*(1), 185-189. https://doi.org/10.1016/j.japb.2024.09.001

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Cock, P. J. A., et al. (2009). Biopython: Python tools for computational biology. *Bioinformatics*, 25(11), 1422–1423. https://doi.org/10.1093/bioinformatics/btp163

Georgieva, Y., Katsarova, M., Stoyanov, P., Mladenov, R., Denev, P., Teneva, D., ... & Dimitrova, S. (2020). Metabolite profile and antioxidant activity of some species of Genus *Scutellaria* growing in Bulgaria. *Plants*, *10*(1), 45. https://doi.org/10.3390/plants10010045

Jiang, D., Zhao, Z., Zhang, T., Zhong, W., Liu, C., Yuan, Q., & Huang, L. (2017). The chloroplast genome sequence of Scutellaria baicalensis provides insight into intraspecific and interspecific chloroplast genome diversity in *Scutellaria*. *Genes*, *8*(9), 227. https://doi.org/10.3390/genes8090227

Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., dePamphilis, C. W., Yi, T. S., & Li, D. Z. (2020). GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology*, 21, 241. https://doi.org/10.1186/s13059-020-02154-5

Karimov, B., Asatulloev, T., Buxorov, G. I., Turginov, O., Naralieva, N., Azimova, D., ... & Yusupov, Z. (2024). Analysis of chloroplast genomes of ten central Asian *Fritillaria* species and their phylogenetic relationships. *Nordic Journal of Botany*, *2024*(12), e04413. https://doi.org/10.1111/njb.04413

Karimov, B. (2025) «Supplementary Materials for: Phylogenomics and biogeographic history of Scutellaria reveal a southern Asian origin and global dispersal patterns», *Plant Diversity of Central Asia*. *Zenodo*. https://doi.org/10.5281/zenodo.19494253

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7. *Molecular Biology and Evolution*, 30(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., *et al.* (2012).

Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v6: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, 49(W1), W293–W296. https://doi.org/10.1093/nar/gkab301

Li, B., Cantino, P. D., Olmstead, R. G., Bramley, G. L., Xiang, C. L., Ma, Z. 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. https://doi.org/10.1038/srep34343

Li, P., Qi, Z. C., Liu, L. X., Ohi-Toma, T., Lee, J., Hsieh, T. H., ... & Qiu, Y. X. (2017). Molecular phylogenetics and biogeography of the mint tribe Elsholtzieae (Nepetoideae, Lamiaceae), with an emphasis on its diversification in East Asia. *Scientific Reports*, *7*(1), 2057. https://doi.org/10.1038/s41598-017-02157-6

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm. *Molecular Biology and Evolution*, 32(1), 268–274. https://doi.org/10.1093/molbev/msu300

Nikitina, E., Karimov, B., & Tojiboeva, U. (2025). Complete chloroplast genome of the medicinally important plant *Ajuga turkestanica* (Regel) Briq.(Lamiaceae) from Uzbekistan. *Journal of Asia-Pacific Biodiversity*. https://doi.org/10.1016/j.japb.2025.07.003

Niu, Y., Qin, Q., Dong, Y., Wang, X., Zhang, S., & Mu, Z. (2023). Chloroplast genome structure and phylogenetic analysis of 13 Lamiaceae plants in Tibet. *Frontiers in Bioscience-Landmark*, *28*(6), 110. https://doi.org/10.31083/j.fbl2806110

Paton A (1990a) A global taxonomic investigation of *Scutellaria* L. Kew Bull 45:399–450. https ://doi.org/10.2307/4110512

Paton A (1990b) The phylogeography of *Scutellaria* L. Notes Roy Bot Gard Edinburgh 46:345–359

Paton A, Suddee S, Bongcheewin B (2016) Two new species of *Scutellaria* (Lamiaceae) from Thailand and Burma. Kew Bull 71:1–6. https://doi.org/10.1007/s12225-016-9620-2

Paton AJ (1989). A global taxonomic investigation of *Scutellaria* L. and its allies (Labiatae). Phd thesis, University of Edinburgh

Qu, X. J., Moore, M. J., Li, D. Z., & Yi, T. S. (2019). PGA: A software package for rapid, accurate, and flexible batch annotation of plastomes. *Bioinformatics*, 35(21), 4510–4512. https://doi.org/10.1093/bioinformatics/btz326

Rodideal, T., Frunzete, M. E., Olaru, Ș. M., GRIGORE, M. N., Zamfirache, M. M., & Ivănescu, L. C. (2024). Comparative analysis of seed morphology in genus *Scutellaria* L. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, *52*(3). https://doi.org/10.15835/nbha52314088

Safikhani, K., Jamzad, Z., & Saeidi, H. (2018). Phylogenetic relationships in Iranian *Scutellaria* (Lamiaceae) based on nuclear ribosomal ITS and chloroplast trn LF DNA data. *Plant Systematics and Evolution*, *304*(9), 1077-1089. https://doi.org/10.1007/s00606-018-1533-0

Salimov, R. A., Parolly, G., & Borsch, T. (2021). Overall phylogenetic relationships of Scutellaria (Lamiaceae) shed light on the origin of the predominantly Caucasian and Irano-Turanian *S. orientalis* group. *Willdenowia*, *51*(3), 395-427. https://doi.org/10.3372/wi.51.51307

Shan, Y., Pei, X., Yong, S., Li, J., Qin, Q., Zeng, S., & Yu, J. (2021). Analysis of the complete chloroplast genomes of *Scutellaria tsinyunensis* and *Scutellaria tuberifera* (Lamiaceae). *Mitochondrial DNA Part B*, *6*(9), 2672-2680. https://doi.org/10.1080/23802359.2021.1920491

Shen, J., Li, P., Liu, S., Liu, Q., Li, Y., Sun, Y., ... & Xiao, P. (2021). Traditional uses, ten-years research progress on phytochemistry and pharmacology, and clinical studies of the genus *Scutellaria*. *Journal of Ethnopharmacology*, *265*, 113198. https://doi.org/10.1016/j.jep.2020.113198

Vassou, S. L., Kusuma, G., & Parani, M. (2015). DNA barcoding for species identification from dried and powdered plant parts: a case study with authentication of the raw drug market samples of Sida cordifolia. *Gene*, *559*(1), 86-93. https://doi.org/10.1016/j.gene.2015.01.025

Wang, Y., Xu, C., Guo, X., Wang, Y., Chen, Y., Shen, J., ... & Wang, Q. (2024). Phylogenomics analysis of *Scutellaria* (Lamiaceae) of the world. *BMC biology*, *22*(1), 185. https://doi.org/10.1186/s12915-024-01982-2

Yu, Y., Harris, A. J., Blair, C., & He, X. (2015). RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. *Molecular phylogenetics and evolution*, *87*, 46-49. https://doi.org/10.1016/j.ympev.2015.03.008

Zhao, F., Li, B., Drew, B. T., Chen, Y. P., Wang, Q., Yu, W. B., ... & Xiang, C. L. (2020). Leveraging plastomes for comparative analysis and phylogenomic inference within Scutellarioideae (Lamiaceae). *PloS one*, *15*(5), e0232602. https://doi.org/10.1371/journal.pone.0232602

Zhao, F., Liu, E. D., Peng, H., & Xiang, C. L. (2017). A new species of *Scutellaria* (Scutellarioideae, Lamiaceae) from Sichuan Province in southwest China. *PeerJ*, *5*, e3624. https://doi.org/10.7717/peerj.3624

Zheng S, Poczai P, Hyvönen J, Tang J and Amiryousefi A (2020) Chloroplot: An Online Program for the Versatile Plotting of Organelle Genomes. *Front. Genet.* 11:576124 https://doi.org/10.3389/fgene.2020.576124

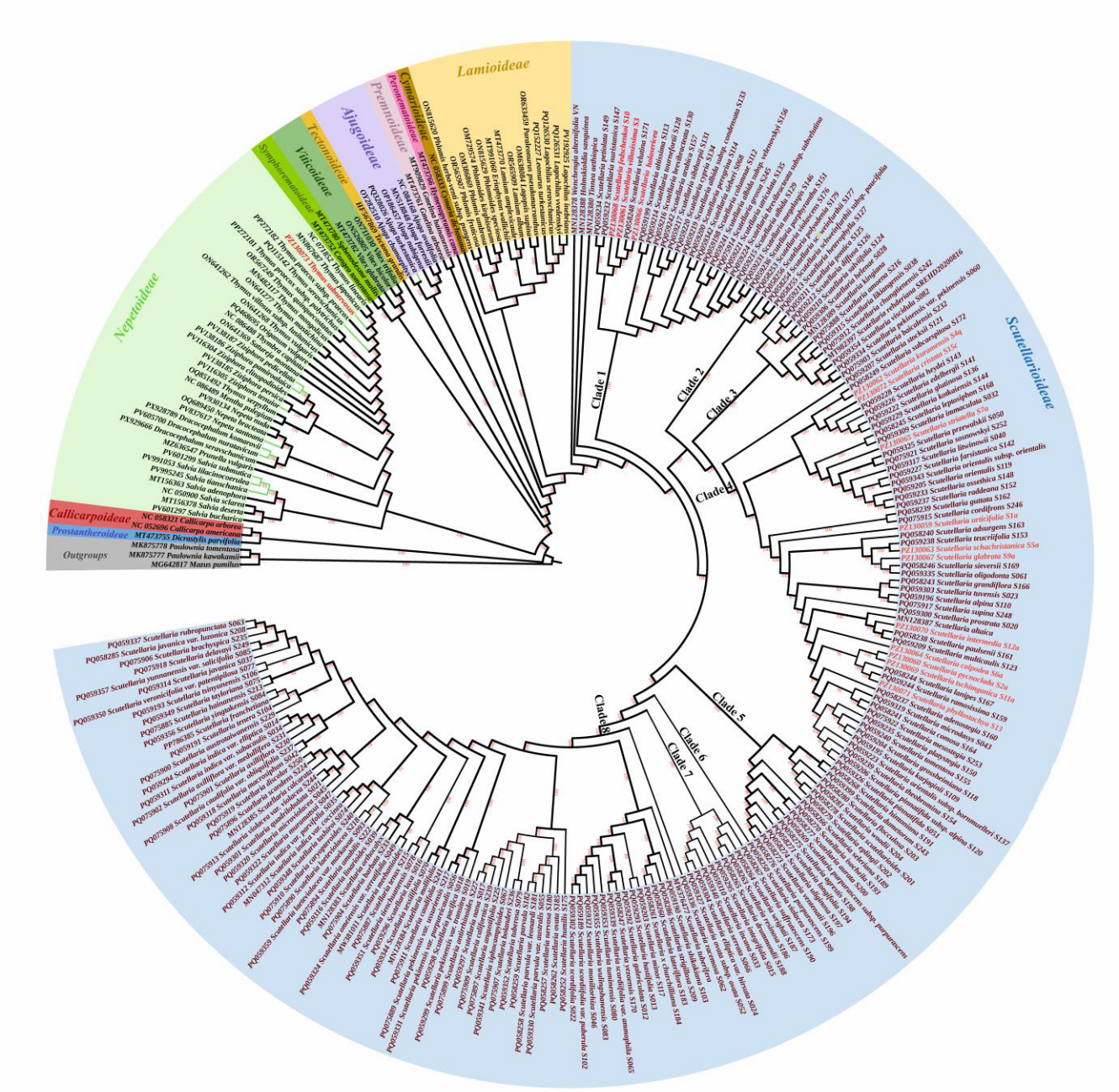

Figure S1. Maximum likelihood phylogenetic tree of *Scutellaria* and related taxa of Lamiaceae based on complete plastome sequences and NCBI GenBank accession numbers. Newly sequenced species are given in red.

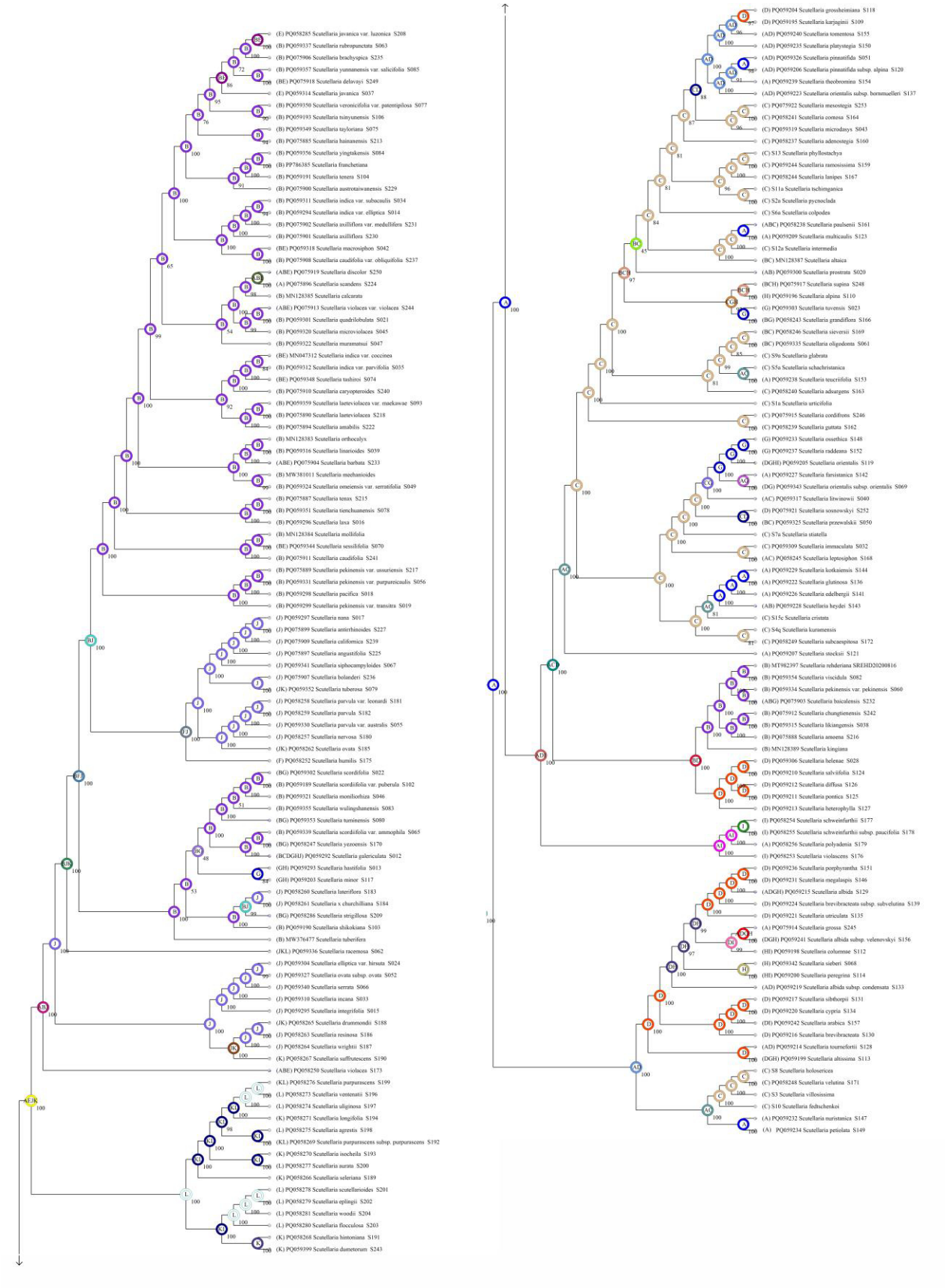

Figure S2. RASP ancestral area reconstruction of *Scutellaria* based on plastome phylogeny, showing inferred ancestral distributions,