

The complete chloroplast genome, DNA barcoding, and taxonomic notes of *Swertiabanzragczii* and *Swertia marginata* (Gentianaceae) in Mongolia

Dashzeveg Oyuntsetseg

National University of Mongolia

Nudkhuu Nyamgerel

Changwon National University

Shukherdorj Baasanmunkh

Changwon National University

Batlai Oyuntsetseg

National University of Mongolia

Magsar Urgamal

Mongolian Academy of Sciences

Jung Won Yoon

Korea National Arboretum

Gun-Ajav Bayarmaa

National University of Mongolia

Hyeok Jae CHOI (✉ hjchoi1975@changwon.ac.kr)

Changwon National University <https://orcid.org/0000-0001-6315-0071>

Research Article

Keywords: *Swertia*, endemic, chloroplast genome, taxonomy, flora of Mongolia

Posted Date: December 25th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3748879/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Swertia banzragczii and *S. marginata* are important medicinal species in Mongolia. However, their taxonomic positions and genetic backgrounds remain unknown. In this study, we explored the complete chloroplast genomes and morphological features of these species and compared them with those of closely related species within the subgenus to determine their taxonomic positions and phylogenetic relationships.

Result

The chloroplast genomes of *S. banzragczii* and *S. marginata* encoded 114 genes, including 80 protein-coding genes, 30 tRNA genes, and 4 rRNA genes. Among them, 16 genes contained a single intron, and 2 genes had two introns. Closely related species had a conserved genome structure and gene content. Only differences in genome length were noticed, which were caused by the expansion and contraction of the inverted repeat (IR) region and loss of exons in some genes. The *trnH*-UUU–*psbA* and *trnD*-GUC–*trnY*-GUA intergenic regions had high genetic diversity within *Swertia* plastomes and could be used as barcoding markers. Overall, *S. banzragczii* and *S. marginata* are true species and belong to the subgenus *Swertia*.

Conclusions

These results provide valuable genetic and morphological information on rare and subendemic *Swertia* species in Mongolia, which can be used for further advanced studies on the *Swertia* genus.

Background

Swertia L. is one of the largest genera in the Gentianaceae Juss. and comprises 160 species (Ho and Liu 2015; Nampy et al. 2015; POWO, 2023). Members of the *Swertia* genus are mainly distributed in the temperate and alpine regions of Asia, Europe, and the Americas (Ho and Liu 2015). This genus has great medicinal importance and is the most taxonomically challenging taxon in the family, largely because of the morphological similarities of species within this genus and closely related genera (Chassot et al. 2001; Ho and Liu 2015; Cao et al. 2022). It is distinguished by a rotated corolla and the presence of coralline nectariferous glands in related genera within the family (Ho and Liu 2015). Furthermore, the nature and shape of these glands are two of the main diagnostic characteristics for species delimitation in this genus (Favre et al. 2015).

The development of barcoding markers has greatly advanced our understanding of the relationships between morphologically similar species; however, it has been revealed that *Swertia* is strongly paraphyletic to other genera within the Gentianaceae family (Chassot et al. 2001; Xi et al. 2014). Recently,

the chloroplast genome has been widely used in plant system evolution, related species identification, genetic diversity analysis, chloroplast genetic engineering, and other applications due to the development of high-throughput sequencing technology (Yang et al. 2016; Li et al. 2021; Fang et al. 2023; Nyamgerel et al. 2023). The chloroplast genome (plastome) has a conserved structure, encodes 120–130 genes, most of which encode part of the photosynthetic apparatus of the organelle (Jensen and Leister, 2014), and is maternally inherited (Wicke et al. 2011). Therefore, plastomes have stronger species-discriminating abilities and provide more information on genetic variation than nuclear DNA barcodes and ultra-barcodes (Hollingsworth et al. 2016). Several studies have explored the complete chloroplast genome of *Swertia* species, mainly from China (Bi et al. 2020; Du et al. 2022; Yang et al. 2020; Cao et al. 2022; Yang et al. 2022; Yang et al. 2023). These studies determined the plastome structure, genetic variety, and phylogenetic relationships across 26 species of *Swertia* and compared them with those of the subtribe Swertiinae species. However, previous studies only explored 20% of *Swertia* species, and Mongolian *Swertia* species are yet to be explored. Further plastome studies are needed to gain a better understanding of the adaptation, evolution, and phylogenetic relationships of the *Swertia* genus.

Four to five species of *Swertia* have been recorded in Mongolia, i.e., *Swertia banzragczii* Sanchir, *S. marginata* Schrenk (or *S. komarovii* Pissajuk.), *S. obtusa* Ledeb. and *S. dichotoma* L. (Baasanmunkh et al. 2022a). The current taxonomic status of *S. banzragczii* and *S. marginata* (or *S. komarovii* Pissajuk.) is still doubtful (Ho and Liu, 2015). However, these species are currently accepted taxa on websites such as the Global Biodiversity Information Facility (GBIF) and Plants of the World Online (POWO, 2023). The two *Swertia* species have not been fully revised based on morphological and molecular evidence, including DNA barcoding and plastomes.

Therefore, the goal of the present study was to explore the plastomes of two rare and endemic *Swertia* species from Mongolia, identify genetically variable regions by comparing species within the *Swertia* genus, and reconstruct their taxonomic position based on complete chloroplast and internal transcribed spacer (ITS) sequences and morphological features.

Results

Taxonomic notes on *S. banzragczii* and *S. marginata*

Swertia banzragczii Sanchir, Novosti Sist. Vyssh. Rast. 21: 136 (1984). (Fig. 1). TYPE: Bayan-Ulgii Province, Sagsai Soum, Dayan Lake, 26 July 1977, Ch. Sanchir, Z.V. Karamisheva, I.Yu. Sumerina, D., Baatarsukh 1273 (Holotype LE).

Taxonomic notes. *Swertia banzragczii* was first described in Western Mongolia (Sanchir 1984). Later, this species was found in Xinjiang, China (Chen, 2011); however, the taxonomic status of *S. banzragczii* is uncertain (Ho and Liu 2015). Our results showed that *S. banzragczii* is a true species based on morphological and molecular evidence and that its current Altai endemic species has a narrow distribution in the Altai Mountain range (Erst et al. 2022).

Penology. Flowering and fruiting in July to September.

General distribution and habitat. China (Xinjiang) and Mongolia (the Khovd and Mongolian Altai regions) (Baasanmunkh et al. 2022a). This species grows in alpine meadows, dwarf birch groves, sparse larch groves, and the borders of the alpine belt (Baasanmunkh et al. 2022b).

Conservation status. *Swertia banzragczii* was assessed as near-threatened at the global level, according to our previous study (Erst et al. 2023).

Swertia marginata Schrenk, Bull. Sc. Acad. Petersb. x. 353 (1842). (Fig. 2)

Synonym. *Swertia komarovii* Pissajuk. TYPE: Montes Sajanenses Orientalis, prope trajectum Carganskyi, 29–31 July 1902, fl. V.L. Komarov. (Holotype LE01057204; <http://re.herbariumle.ru/01057204>)

Taxonomic notes. The taxonomic status of *S. marginata* has been thoroughly addressed by Ho and Liu (2015). According to Ho and Liu (2015), *S. komarovii* Pissajuk. was treated as a synonym of *S. marginata* based on its morphological characteristics, but it is currently accepted as a species in GBIF (<https://www.gbif.org/species/5596000>). *Swertia komarovii* was first described in Russia and may have occurred in the Khusgul region of Mongolia (Pissajukova 1961). Based on our phylogenetic results, the Mongolian samples clustered with *S. marginata*; therefore, we confirmed that *S. komarovii* is a synonym of *S. marginata*.

Penology. Flowering and fruiting in July to September.

General distribution and habitat. Afghanistan, China, Kazakhstan, Kyrgyzstan, Mongolia (Khuvsgul and Mongolian Altai regions), Russia, and Tajikistan (Ho and Liu 2015). In Mongolia, this species grows in wet meadows, syrts (elevated watersheds), along the banks of brooks, moraines, shrubs, rubbly sites, slopes, taluses, and rocks in spruce forests.

Conservation status. The conservation status is currently not assessed at the global level.

Chloroplast genome features of *S. banzragczii* and *S. marginata*

The chloroplast genomes of *S. banzragczii* and *S. marginata* were sequenced for the first time. In total, 54,848,640 and 66,783,272 reads were produced for *S. banzragczii* and *S. marginata*, respectively (Table 1). After *de novo* assembly, a single contig was generated for each plastome. The two plastomes had typical quadripartite natural structures comprising a large single copy (LSC), a pair of inverted repeats (IR), and a small single copy (SSC) (Fig. 3). The total length of *S. marginata* plastome was shorter than the *S. banzragczii* plastome in all regions (Table 1). The overall guanine-cytosine (GC) content of the two plastomes was similar, differing by 0.1% in total length. The two plastomes encoded 114 genes, including 80 protein-coding genes, 30 tRNA genes, and 4 rRNA genes. Among them, 18 genes (7 protein-coding, 7 tRNA, and 4 rRNA genes) were duplicated in the IR regions. Sixteen genes contained a single intron, and two genes had two introns (Table 2). Of the 18 genes, 12 intron-containing genes (*trnK*-UUU,

rps16, *trnG-UCC*, *atpF*, *rpoC1*, *trnL-UAA*, *trnV-UAC*, *petB*, *petD*, and *rp16*) were located in the LSC and 1 gene (*ndhA*) was located in the SSC. Five genes (*rp12*, *ndhB*, *rps12*, *trnI-GAU*, and *trnA-UGC*) were duplicated in the IRs. The *rps12* gene was a trans-spliced gene with 5' end located in the LSC and 3' end located in the IR.

Table 1
Raw reads and genome assembly information for *S. banzragczii* and *S. marginata* plastomes.

	<i>S. banzragczii</i>	<i>S. marginata</i>
Input reads (bp)	54,848,640	66,783,272
Trimmed reads (bp)	7,153,975,259	7,538,583,714
Total raw base (bp)	8,282,144,640	10,084,274,072
Trimmed bases (bp)	47,559,948	50,265,678
Q20 (%)	99.23	98.95
Q30 (%)	93.54	95.89
Coverage (%)	100	100
Depth (X)	46.493	49,282
CG content (%)	38.0	38.1
Cp genome length (bp)	153,872	152,968
LSC/GC content	83,841/36.1	83,216/36.1
IR/GC content	25,855/43.3	25,755/43.4
SSC/GC content	18,321/31.8	18,242/31.9

Table 2
Gene category of *S. banzragczii* and *S. marginata* plastomes

Category	Group of genes	Name of genes
RNA genes	ribosomal RNA genes (rRNA)	<i>rrn5^c, rrn4.5^c, rrn16^c, rrn23^{a,c}</i>
	Transfer RNA genes (tRNA)	<i>trnA-UGC^{a,c}, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnM-CAU, trnG-GCC, trnG-UCC^a, trnH-GUG, trnI-CAU^c, trnI-GAU^a, trnK-UUU^a, trnL-CAA^c, trnL-UAA^a, trnL-UAG, trnM-CAU, trnN-GUU^c, trnP-UGG, trnQ-UUG, trnR-ACG^c, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC^c, trnV-UAC^a, trnW-CCA, trnY-GUA</i>
Ribosomal proteins	Small subunit of ribosome	<i>rps2, rps3, rps4, rps7^c, rps8, rps11, rps12^{c,b,d}, rps14, rps15, rps16^a, rps18 rps19^c</i>
Transcription	Large subunit of ribosome	<i>rpI2^{a,c}, rpI16^a, rpI14, rpI20, rpI22, rpI23^c, rpI33, rpI36, rpI32</i>
	RNA polymerase	<i>rpoA, rpoB, rpoC1^a, rpoC2</i>
Protein genes	Photosystem I	<i>psaA, psaB, psaC, psal, psaJ, ycf3^b, ycf4</i>
	Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbl, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	Cytochrome b6/f	<i>petA, petB^a, petD^{a,d}, petG, petL, petN</i>
	ATP synthase	<i>atpA, atpB, atpE, atpF^a, atpH, atpI</i>
	Rubisco	<i>rbcL</i>
	NADH dehydrogenase	<i>ndhA^a, ndhB^{a,c}, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
	ATP-dependent protease subunit P	<i>clpP^b</i>
	Chloroplast envelope membrane protein	<i>cemA</i>
	Transitional initiation factor	<i>infA</i>
Other genes	Maturase	<i>matK</i>
^a Gene with one intron, ^b Gene with two introns, ^c Gene with copies, ^d Trans-splicing gene		

Category	Group of genes	Name of genes
	Subunit acetyl-coA carboxylase	<i>accD</i>
	C-type cytochrome synthesis	<i>ccsA</i>
	Hypothetical proteins	<i>ycf1^c, ycf2^c, ycf15^{e,c}</i>
	Component of TIC complex	<i>ycf3^b</i>
^a Gene with one intron, ^b Gene with two introns, ^c Gene with copies, ^d Trans-splicing gene		

Chloroplast genome comparison of subgenus *Swertia* species

The cp genome size, structure, and gene positions of *S. banzragczii* and *S. marginata* were compared with those of five closely related species within the subgenus *Swertia*: *S. bifolia* Batalin, *S. erythrosticta* Maxim., *S. souliei* Burkill, *S. przewalskii* Pissjauk, and *S. wolfgangiana* Grüning. The gene content and location in all regions of closely related species were similar, and the plastomes did not display any species-specific inversion or translocation (Fig. 4). The IR of *S. banzragczii* and *S. marginata* were 25,855 and 25,755 bp, respectively. The borders between the IR regions and two single-copy regions (LSC and SSC) were similar to closely related species (Fig. 5). The *rps19*, *ndhF*, and *ycf1* genes spanned the LSC/IRb, IRb/SSC, and SSC/IRa junctions, respectively. However, the *ndhF* gene of *S. marginata* was 935 bp shorter than that of the others and was entirely located in the SSC region. Relative synonymous codon usage (RSCU) was computed for protein-coding genes. These genes revealed 26,770 and 26,596 codons in *S. banzragczii* and *S. marginata* plastomes, respectively, and they had similar patterns in both plastomes, with the highest RSCU values for the AGA (1.9) codon of arginine (Arg) and the highest coding rate for leucine (Leu; 5,216 codons), isoleucine (Ile; 4,138 codons), and phenylalanine (Phe; 3,603 codons) (Fig. 6). Codons with A or T at the third position had a strong codon bias. The two species studied exhibited similar patterns. All coding sequences (CDS) contained a standard ATG start codon. Among the three stop codons, TAA was the most common. The sequence identities of the seven *Swertia* species were analyzed using mVISTA, with the *S. banzragczii* plastome serving as a reference. As expected, the genic regions were more conserved than the intergenic regions, and genes in the IR regions were highly conserved, followed by those in the LSC and SSC regions (Fig. 7).

We analyzed the genetic divergence of genes and intergenic regions in the *Swertia* genus. Overall, *Swertia* plastomes had an average genetic diversity (Pi) value of 0.17. The highest Pi values were observed in the *trnH*-UUU-*psbA* (0.082) and *trnD*-GUC-*trnY*-GUA (0.016) intergenic regions of the LSC regions (Fig. 8). Among these genes, the highest value (0.67) was found for *rps16* in the LSC region. The genic and intergenic regions in the IR region were relatively constant compared to those in the LSC and SSC regions.

Common chloroplast barcoding markers such as *matK*, *rbcl*, *trnL-F*, and *trnH-psbA* showed relatively low genetic diversity within the *Swertia* genus.

Phylogenetic position of *S. banzragczii* and *S. marginata*

We constructed plastome datasets for 26 *Swertia* species using *Gentianopsis paludosa* (Hook.f.) Ma as an outgroup. The ITS dataset was constructed using 25 ITS sequences from *Gentiana straminea* Maxim. as an outgroup. Additional sequences of the compared species were downloaded from GenBank, and the accession numbers are listed in Table 3. The alignment of *Swertia* plastomes consisted of 152,075 bp, of which 82,024 bp (94.2%) were similar, and 5,051 bp (5.8%) were parsimony-informative. The alignment of the ITS sequences consisted of 622 bp, 153 of which were parsimony-informative. The topologies of maximum parsimony (MP) and Bayesian interference (BI) were highly congruent with the datasets and supported by strong bootstrap values and posterior probabilities. *Swertia* species were divided into two clades based on the plastome and the ITS trees. These clades consisted of subbranches that contained similar species in both trees. *S. banzragczii* and *S. marginata* formed a monophyletic group with *S. bifolia*, *S. erythrosticta*, *S. souliei*, *S. przewalskii* and *S. wolfgangiana* within Clade II (Fig. 9). In the ITS tree, *S. perennis* was closest to *S. banzragczii*. *S. marginata* was closely related to *S. bifolia*, *S. przewalskii*, and *S. wolfgangiana* in both trees.

Table 3
The accession numbers and the sequence lengths of the specimens.

	Species	Plastome		ITS	
		Sequence ID	Length (bp)	Sequence ID	Length (bp)
1	<i>Swertia banzragczii</i>	OR643855	153,872	OR034100	740
2	<i>S. marginata</i>	OR786442	152,968	OR019670	746
3	<i>S. bifolia</i>	MZ261897	153,293	KC861298	634
4	<i>S. bimaculata</i>	MW344296	153,751	LM644018	697
5	<i>S. cincta</i>	MZ261898	149,089	MN561222	687
6	<i>S. cordata</i>	NC_054359	153,429	-	-
7	<i>S. dichotoma</i>	MZ261899	152,977	DQ317488	630
8	<i>S. dilatata</i>	MW344298	150,057	JX569819	786
9	<i>S. diluta</i>	NC_057681	153,691	-	-
10	<i>S. erythrosticta</i>	MW344299	153,039	KF563976	625
11	<i>S. franchetiana</i>	NC_056357	153,428	KC935872	625
12	<i>S. hispidicalyx</i>	NC_044474	149,488	KP419964	627
13	<i>S. japonica</i>	LC744566	153,208	LC645204	812
14	<i>S. kouitchensis</i>	MZ261902	153,475	-	-
15	<i>S. leducii</i>	NC_045301	153,015	MN561242	708
16	<i>S. macrosperma</i>	MZ261903	152,737	MN561226	714
17	<i>S. multicaulis</i>	NC_050660	152,190	FJ010799	724
18	<i>S. mussotii</i>	KU641021	153,499	MN561214	695
19	<i>S. nervosa</i>	NC_057596	153,690	FJ010794	757
20	<i>S. przewalskii</i>	MW344305	153,160	AF255913	623
21	<i>S. pubescens</i>	MZ261905	149,036	KC861341	625
22	<i>S. punicea</i>	MZ261896	153,448	MN561219	688
23	<i>S. souliei</i>	ON164641	152,804	-	-
24	<i>S. tetraptera</i>	MW044653	152,787	JF978834	674
25	<i>S. verticillifolia</i>	MF795137	151,682	KX549987	624

Species		Plastome		ITS	
		Sequence ID	Length (bp)	Sequence ID	Length (bp)
26	<i>S. wolfgangiana</i>	MZ261906	153,090	AF255914	622
27	<i>S. perennis</i>	-	-	AJ580550	824
28	<i>S. marginata</i>	-	-	KC861304	624
29	<i>Gentianopsis paludosa</i>	NC_050656	151,308	-	
30	<i>Gentiana straminea</i>	-	-	MF579735	627

Discussion

The complete plastome of *Swertia* has recently been studied (Bi et al. 2020; Yang et al. 2020; Cao et al. 2022; Qing et al. 2022; Yang et al. 2022; Yang et al. 2023). However, previous studies have not explored Mongolian *Swertia* species. In this study, we sequenced and characterized the plastomes and determined their taxonomic positions based on the genetic and morphological features of two *Swertia* species in Mongolia.

The complete plastomes of *S. banzragczii* and *S. marginata* have a typical quadripartite structure with one LSC, one SSC, and two IR regions. They have relatively conserved features, genes, and GC content of sequences, similar to most species in Gentianaceae, and no significant structural rearrangements (Fig. 4) (Zhang et al. 2020; Yang et al. 2023). The *Swertia* plastome ranges from 149,089 bp (*S. cincta*) to 153,872 bp (*S. banzragczii*), and the difference in length is caused by expansion and contraction of the IR region, as previously reported in the *Swertia* genus (Cao et al. 2022; Yang et al. 2022). The IR lengths of *Swertia* ranged from 25,331 to 25,890 bp, and *rps19* and *ycf1* spanned the LSC/IRb and SSC/IRa junctions, respectively, within all species. The IRb/SSC junction varied greatly owing to the extension of 5–100 bp of *ndhF* from SSC to IRb. In the *S. banzragczii* plastome, the IRb/SSC junction includes *ndhF*, which consists of 54 and 2,175 bp in the IRb and SSC regions, respectively (Fig. 5). In contrast, the IRb/SSC junction of *S. marginata* plastome did not contain *ndhF*, which was located entirely in the SSC region. Synonymous codons are genetic codons encoding the same amino acids. Codon usage bias describes the frequency of synonymous codons that vary among different species. Codon usage patterns can be used to evaluate genetic evolution, environmental adaptation, and evolutionary relationships between species (Athey et al. 2017). Previous studies have shown that analysis of codon bias in the *Swertia* plastome revealed that leucine, isoleucine, and phenylalanine had the highest coding rates (Yang et al. 2022), and similar results have been reported in *S. banzragczii* and *S. marginata* plastomes (Fig. 6). In addition, the plastomes of the subtribe Swertiinae had a weak codon bias, which might be related to the conserved of plastome (Yang et al. 2023).

The plastome has a slower rate of nucleotide substitution in the coding region and a more conservative gene structure and content than the nuclear genome. It can act as a molecular marker for the

interpretation of evolutionary relationships and improve phylogenetic resolution at lower taxonomic levels (Wicke et al. 2011). The sequence divergence of 23 *Swertia* species was discovered by Yang et al. (2022); five (*rpoC1*, *ccsA*, *ndhI*, *ndhA*, and *rps15*) protein-coding genes and five (*trnH-GUG-psbA*, *psaA-ycf3*, *cemA-petA*, *ycf15-trnL-CAA*, and *ccsA-ndhD*) non-coding regions had the highest variability in the plastome. The plastomes of 26 *Swertia* species were used in this study to calculate genetic diversity; the *rps16* gene and certain intergenic regions (*trnH-UUU-psbA*, *trnD-GUC-trnY-GUA*, *ndhF-rpl32*, *trnK-rps16*, *ccsA* and *ndhD*) showed the highest values (Fig. 7). However, Yang et al. (2022) reported that some species of *Swertia* have lost an exon of *rps16*, which has been found in several Gentianaceae species, and that its structure was similar in all species. Among the plastid barcoding markers, *trnH-psbA* and *trnK-rps16* regions can be used in phylogenetic analyses or as potential DNA molecular barcodes for *Swertia* species.

The phylogenetics of *Swertia* have been the subject of extensive research. Previous phylogenetic studies of *Swertia* species have used nrDNA (ITS) and plastid (*matK*, *trnL*, *trnL-trnF*, and *trnH-psbA*) barcoding markers (Chassot et al. 2001; Xi et al. 2014; Groff et al. 2015). These findings indicate that *Swertia* is paraphyletic, with certain species sharing a foundational polytomy with other closely related genera. Moreover, pollen morphology analysis revealed that the *Swertia* genus has 14 different lineages at the respective phylogenetic positions within the Swertiinae subtribe (Chassot and Hagen 2008). Recently, the taxonomic relationships of *Swertia* have been revised using the complete chloroplast genome (Cao et al. 2022; Yang et al. 2022; Yang et al. 2023). Some studies have investigated the connections between *Swertia* and other genera in Gentianaceae, such as *Gentianopsis*, *Halenia*, *Lomatogonium*, *Lomatogoniopsis*, and *Sinoswertia* (Park et al. 2019; Li et al. 2021; Yang et al. 2023). These findings support the polyphyly of *Swertia*, which is consistent with previous molecular investigations but provides significantly better resolution. For example, the polyphyly of *Swertia* appears to be a consequence of recent rapid radiation (Cao et al. 2022). *Swertia* formed in the early Miocene (29.60 Ma), and speciation occurred in the middle Miocene as the global temperature declined from the optimal temperature. However, the molecular data only supported the separation of sections *Ophelia*, *Platynema*, *Poephila*, *Swertia*, and *Macranthos* (Yang et al. 2022). In this study, we first determined the phylogenetic position of *S. banzragczii* and *S. marginata*, which were separated from other species with high support values of BI and MP analyses, forming a monophyletic clade with section *Swertia* species based on the plastome and nrDNA (ITS) phylogenetic trees (Fig. 12). In previous studies, the section *Swertia* always formed a well-supported monophyletic clade, whereas the genus was polyphyletic in phylogenetic trees (Xi et al. 2014; Yang et al. 2022). The species included in this study belonged to the section *Swertia* and are truly separate species.

Conclusion

In this study, the plastomes of *S. banzragczii* and *S. marginata* from Mongolia were sequenced and assembled for the first time. The genome size, organization, and gene content were similar to those of other *Swertia* species. However, they varied slightly with respect to genome length and sequence of intergenic regions. Based on the phylogenetic analyses, they belonged to the subgenus *Swertia*. In

general, our results demonstrated that *S. banzragczii* and *S. marginata* are true species and contribute more information towards resolving taxonomic ambiguities around *Swertia* species in Mongolia. Furthermore, this study used molecular evolution and phylogenetic analyses to determine the genetic connections within *Swertia*.

Material and Methods

Plant material, DNA extraction, and sequencing

Fresh leaves of *S. banzragczii* and *S. marginata* were collected from Bayan-Ulgii and Khuvsgul provinces, Mongolia, and voucher specimens (UBU0005243 and UBU0030443) were deposited in the herbarium of the National University of Mongolia, Mongolia. Photographs of each species taken in the wild were obtained from field surveys conducted by the authors. Total genomic DNA was extracted from the silica gel-dried leaf material using the CTAB method (Doyle and Doyle 1987). DNA purity and concentration were quantified using agarose gel electrophoresis and a NanoDrop 2000 spectrophotometer, respectively.

The nuclear ITS region (White et al. 1990) was used for amplification and sequencing. The PCR cycling parameters were 5 min at 94°C, followed by 30 cycles of 45 s at 94°C, 45 s at 56°C, and a final 7 min at 72°C. The PCR reaction (20 µL) includes 10X PCR buffer, 1.5 U Taq DNA polymerase, 0.2 mM dNTPs, 1 µM of each primer, and 100 ng genomic DNA. PCR products were subjected to gel electrophoresis and cleaned using a PCR clean-up kit. The purified PCR products were directly sequenced in two directions using the corresponding primers. Sequences were aligned using ClustalW (Thompson et al. 2003), and default settings and manual adjustments were made using BioEdit (Hall et al. 2011). Ambiguous nucleotide bases were corrected using the corresponding bases of sequences obtained using the reverse primer. The DNA sequences generated in this study have been deposited in GenBank (<https://www.ncbi.nlm.nih.gov/>).

The complete cp genome library was constructed using the TrueSeq DNA Nano Kit protocol on the NectSeq 500 platform (Illumina, San Diego, CA, USA), following the extraction of total DNA. Trimmomatic software version 0.36 (Bolger et al. 2014) was used to eliminate adapter sequences and low-quality reads to mitigate biases in the analysis. A FastQC v. 0.11 (Antil et al. 2023). was used to assess the overall data quality by generating a base quality plot. The plot illustrates the distribution of the quality values across each cycle.

Genome assembly and annotation

De novo assembly was conducted using various k-mers with NOVOplasty (Nicolas et al. 2017), using the default settings. A BLAST analysis was conducted following the assembly of the complete plastome or draft plastome to determine the species with which each contig (or scaffold) exhibited similarity. *S. banzragczii* and *S. marginata* plastomes were annotated using the online software GeSeqAnnotation of Organellar Genomes (Tillich et al. 2017). After initial annotation, the tRNA genes were further checked

using tRNAscan-SE v1.21 (Lowe and Chan 2016). We manually calibrated the start and stop codons of coding sequences in the software Geneious v9.0.2 (Kearse et al. 2012). The gene map was drawn using OrganellarGenomeDRAW (OGDRAW) (Lohse et al. 2007). The newly sequenced plastomes were deposited in GenBank under accession numbers OR643855 (*S. banzragczii*) and OR786442 (*S. marginata*).

Comparative analysis

Gene rearrangements of *S. banzragczii* and *S. marginata* plastomes were detected based on collinear blocks using Mauve v2.4.0 (Darling et al. 2010). The SC/IR boundary shifts at the four junctions (LSC/IRa, IRa/SSC, SSC/IRb, and IRb/LSC) were compared using IRScope (<https://irscope.shinyapps.io/irapp/>) (Amiryousefi et al. 2018). The RSCU of the plastomes was analyzed using MEGA11 software (Tamura et al. 2021). RSCU < 1 indicated a codon used less frequently than expected, whereas RSCU > 1 indicated a codon used more frequently than expected. Genome divergence within the five *Swertia* species was identified using mVISTA (Frazer et al. 2004) in Shuffle-LAGAN mode. The plastome sequences were aligned using MAFFT ver. 7.388 (Kato 2002). The protein-coding genes (coding regions), intergenic spacers, and introns (non-coding regions) of the *Swertia* plastomes were separately extracted to screen for polymorphic hotspots. Sliding window analysis was used to calculate the nucleotide variability (Pi) using DnaSP v.6.11 (Rozas et al. 2017). All the compared plastomes were downloaded from the GenBank database (Table 3). The step size was 200 bp with a 600 bp window length.

Phylogenetic analysis

The nrDNA (ITS) and plastome sequences were used to determine the phylogenetic positions of *S. banzragczii* and *S. marginata* within the *Swertia* genus. Plastome alignment datasets were filtered to remove ambiguously aligned regions using GBlock ver. 0.91.1 (Talavera and Castresana 2007). The best-fitting model for nucleotide substitutions was determined using the Akaike information criterion in jModelTest v2.1.10 (Darriba et al. 2012). A phylogenetic tree was constructed using the MP and BI methods. The MP analysis was performed using PAUP*v4.3.99.169.0 (Swofford and Documentation 1989). The branch-swapping algorithm for MP analysis was performed using tree bisection reconnection. The robustness of the tree was evaluated using 1000 bootstrap replication indices, and the consistency, retention, and composite indices were calculated. The reconstructed trees were visualized using FigTree v.1.4.4 (Rambaut 2012). BI analysis was carried out using the MrBayes 3.2.6 (Ronquist et al. 2012) plugin in Geneious v.23.2. (<https://www.geneious.com>; Biomatters Ltd.).

Abbreviations

LSC Large single-copy

SSC Small single-copy

IR Inverted repeat

GC Guanine-Cytosine

CDS Coding sequences

tRNA transfer RNA

rRNA ribosomal RNA

RSCU Relative synonymous codon usage

MP Maximum parsimony

BI Bayesian interference

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Not applicable.

Competing interests

The authors declare no competing interests.

Funding

This research was supported by National University of Mongolia, Mongolia (Grant No. P2023-4591) and the Korea National Arboretum (Grant Number: KNA1-2-42, 22-2) and Korea Basic Science Institute (Nantional Research Facilities and Equipment Center) grant funded by the Ministry of Education (Grant No. 2023R1A6C101B022).

Author' contribution

SB conceived and designed the study; DO and NN conducted the experiment; SB, BO, GB, and UM collected the samples; DO, NN, and SB prepared the original draft; GB and HJC revised and finalized the manuscript. All authors have read and approved the final version of the manuscript.

Acknowledgments

This study is a part of the Ph.D. thesis of the first author.

References

1. Amiryousefi A, Hyvönen J, Poczai P (2018) IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34:3030–3031. <https://doi.org/10.1093/bioinformatics/bty220>
2. Antil S, Abraham JS, Sripoorna S, Maurya S, Dagar J, Makhija S, Bhagat P, Gupta R, Sood U, Lal R, Toteja R (2023) DNA barcoding, an effective tool for species identification: a review. *Mol Biol Rep* 50:761–775. <https://doi.org/10.1007/s11033-022-08015-7>
3. Athey J, Alexaki A, Osipova E, Rostovtsev A, Santana-Quintero LV, Katneni U, Simonyan V, Kimchi-Sarfaty C (2017) A new and updated resource for codon usage tables. *BMC Bioinform* 18:391. <https://doi.org/10.1186/s12859-017-1793-7>
4. Baasanmunkh S, Urgamal M, Oyuntsetseg B, Sukhorukov AP, Tsegmed Z, Son D C, Erst AS, Oyundelger K, Kechaykin AA, Norris J et al (2022a) Flora of Mongolia: annotated checklist of native vascular plants. *PhytoKeys* 192:63–169. <https://doi.org/10.3897/phytokeys.192.79702>
5. Baasanmunkh S, Oyuntsetseg B, Tsegmed Z, Oyundelger K, Urgamal M, et al (2022b) Distribution of vascular plants in Mongolia– I Part. *Mong J Biol Sci* 20(1):3–28. DOI:10.22353/mjbs.2022.20.01
6. Bi H, Yang Y, Hu Q, Ma Y (2020) Characterization of the complete chloroplast genome of *Swertia souliei* (Gentianaceae). *Mitochondrial DNA Part B: Resour* 5:1901–1902. <https://doi.org/10.1080/23802359.2020.1754944>
7. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
8. Cao Q, Gao Q, Ma X, Zhang F, Xing R, Chi X, Chen S (2022) Plastome structure, phylogenomics and evolution of plastid genes in *Swertia* (Gentianaceae) in the Qing-Tibetan Plateau. *BMC Plant Biol* 22:195. <https://doi.org/10.1186/s12870-022-03577-x>
9. Chassot P, Nemomissa S, Yuan YM, Küpfer P (2001) High paraphyly of *Swertia* L. (Gentianaceae) in the Gentianella-lineage as revealed by nuclear and chloroplast DNA sequence variation. *Plant Syst Evol* 229:1–21. <https://doi.org/10.1007/s006060170015>
10. Chassot P, von Hagen KB (2008) Pollen morphology of the Swertiinae (Gentianaceae): phylogenetic implications. *Bot J Linn* 157(2):323–341. <https://doi.org/10.1111/j.1095-8339.2008.00798.x>
11. Chen WL, Zhang SR, Wang J, Liu JQ, Ivanovich SA (2011) Some new or noteworthy plant species for China found in North West Xinjiang. *Turczaninowia* 14(1):75–80.
12. Darling AE, Mau B, Perna NT (2010) progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS One* 5:e11147. <https://doi.org/10.1371/journal.pone.0011147>

13. Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772–772. <https://doi.org/10.1038/nmeth.2109>
14. Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11–15.
15. Du Q, Li J, Wang L, Chen H, Jiang M, Chen Z, Jiang C, Gao H, Wang B, Liu C (2022) Complete chloroplast genomes of two medicinal *Swertia* species: the comparative evolutionary analysis of *Swertia* genus in the Gentianaceae family. *Planta* 256(4):73. <https://doi.org/10.1007/s00425-022-03987-z>
16. Erst AS, Nikulin AY, Nikulin VY, Ebel AL, Zibzeev EV, Sharples MT, Baasanmunkh S, Choi HJ, Olonova MV, Pyak AI, et al (2022) Distribution analysis, updated checklist, and DNA barcodes of the endemic vascular flora of the Altai mountains, a Siberian biodiversity hotspot. *Syst Biodivers* 20:1–30. <https://doi.org/10.1080/14772000.2022.2049391>
17. Erst AS, Baasanmunkh S, Tsegmed Z, Oyundelger K, Sharples MT, Oyuntsetseg B, Krivenko DA et al (2023) Hotspot and conservation gap analysis of endemic vascular plants in the Altai Mountain Country based on a new global conservation assessment. *Glo Ecol Conser* 47:e02647. <https://doi.org/10.1016/j.gecco.2023.e02647>
18. Fang H, Dai G, Liao B, Zhou P, Liu Y (2023) Application of chloroplast genome in the identification of *Phyllanthus urinaria* and its common adulterants. *Front Plant Sci* 13. <https://doi.org/10.3389/fpls.2022.1099856>
19. Favre A, Päckert M, Pauls SU, Jähnig SC, Uhl D, Michalak I, Muellner-Riehl AN (2015) The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biol Rev* 90:236–253. <https://doi.org/10.1111/brv.12107>
20. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: Computational tools for comparative genomics. *Nucleic Acids Res* 32:273–279. <https://doi.org/10.1093/nar/gkh458>
21. Groff PA, Hale AM, Whitlock BA (2015). Chloroplast lineages in disjunct western north American populations of *Swertia perennis* (Gentianaceae). *Syst Bot* 40(1):220–228. <http://www.jstor.org/stable/24546439>
22. Hall T, Biosciences I, Carlsbad C (2011) BioEdit: an important software for molecular biology. *GERF Bull Biosci* 2:60–61.
23. Ho TN, Liu SW (2015) A Worldwide Monograph of *Swertia* and its allies. Science Press, Beijing
24. Hollingsworth PM, Li, DZ, van der Bank M, Twyford AD (2016) Telling plant species apart with DNA: from barcodes to genomes. *Philos Trans R Soc Lond B Biol Sci* 371(1702):20150338. <https://doi.org/10.1098/rstb.2015.0338>
25. Jensen PE, Leister D (2014) Chloroplast evolution, structure and functions. *F1000Prime Rep* 6:40. <https://doi.org/10.12703/P6-40>
- Katoh K (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059–3066. <https://doi.org/10.1093/nar/gkf436>

26. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
27. Li L, Hu Y, He M, Zhang B, Wu W, Cai PM, Huo D, Hong Y (2021) Comparative chloroplast genomes: insights into the evolution of the chloroplast genome of *Camellia sinensis* and the phylogeny of *Camellia*. *BMC Genom* 22:138. <https://doi.org/10.1186/s12864-021-07427-2>
28. 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. *Curr Genet* 52:267–274. <https://doi.org/10.1007/s00294-007-0161-y>
29. Lowe TM, Chan PP (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* 44:W54–W57. <https://doi.org/10.1093/nar/gkw413>
30. Nampy S, Shahina PM, Haseena T, Ashwini HS (2015) A taxonomic revision of *Swertia* L. (Gentianaceae) in South India, with one new species and seven lectotypifications. *Phytotaxa* 195:31–52. <https://doi.org/10.11646/phytotaxa.195.1.2>
31. Nicolas D, Patrick M, Guillaume S (2017) NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res* 45(4):e18. <https://doi.org/10.1093/nar/gkw955>
32. Nyamgerel N, Baasanmunkh S, Oyuntsetseg B, Bayarmaa GA, Erst A, Park I, Choi HJ (2023) Insight into chloroplast genome structural variation of the Mongolian endemic species *Adonis mongolica* (Ranunculaceae) in the Adonideae tribe. *Sci Rep* 13:22014 <https://doi.org/10.1038/s41598-023-49381-x>
33. Park J, Yun N, Oh S (2019) The complete chloroplast genome of an endangered species in Korea, *Halenia corniculata* (L.) Cornaz (Gentianaceae). *Mitochondrial DNA Part B: Resour* 4:1539–1540. <https://doi.org/10.1080/23802359.2019.1601532>
34. Pissjaukova V (1961) Notulae de genera *Swertia* L. *Botanicheskie Materialy Gerbariya Botanicheskogo Instituta Imeni V. L. Komarova Akademii Nauk S S S R. Leningrad.* 21:292–313.
35. POWO (2023) Plants of the World Online. Royal Botanic Gardens, Kew. Retrieved 27 November 2023. <http://www.plantsoftheworldonline.org/>
36. Rambaut A. 2012. FigTree v1. 4. Molecular evolution, phylogenetics and epidemiology. University of Edinburgh, Institute of Evolutionary Biology, Edinburgh, UK.
37. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenback JP (2012) MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst Biol* 61:539–542. <https://doi.org/10.1093/sysbio/sys029>
38. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sanchez-Gracia A (2017) DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol Biol Evol* 34:3299–3302. <https://doi.org/10.1093/molbev/msx248>
39. Sanchir C (1984) New species of genus *Swertia* L. (Gentianaceae) from the Mongolian People's Republic [*Swertia banzragczii*]. *Novosti Sistematiki Vysshikh Rastenii* 21: 136.

40. Swofford DL, Documentation B (1989) Phylogenetic analysis using parsimony. Illinois Nat Hist Surv Champaign.
41. Talavera G, Castresana J (2007) Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. *Syst Biol* 56:564–577. <https://doi.org/10.1080/10635150701472164>
42. Tamura K, Stecher G, Kumar S (2021) MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol* 38:3022–3027. <https://doi.org/10.1093/molbev/msab120>
43. Thompson JD, Gibson TJ, Higgins DG (2003) Multiple Sequence Alignment Using ClustalW and ClustalX. *Curr Protoc Bioinforma* 1:2–3. <https://doi.org/10.1002/0471250953.bi0203s00>
44. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S (2017) GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res* 45:W6–W11. <https://doi.org/10.1093/nar/gkx391>
45. Wicke S, Schneeweiss GM, Depamphilis CW, Müller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol* 76:273–297. <https://doi.org/10.1007/s11103-011-9762-4>
46. White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protoc* 18(1):315–322. <https://doi.org/10.1016/b978-0-12-372180-8.50042-1>
47. Xi HC, Sun Y, Xue CY (2014) Molecular phylogeny of Swertiinae (Gentianaceae-Gentianeae) based on sequence data of the ITS and *matK*. *Plant Divers Resour* 36(2):145–56. <https://journal.kib.ac.cn/EN/Y2014/V36/I02/145>
48. Yang L, Xiong F, Xiao Y, Li J, Chen C, Zhou G (2020) The complete chloroplast genome of *Swertia tetraptera* and phylogenetic analysis. *Mitochondrial DNA Part B: Resour* 5: 164–165. <https://doi.org/10.1080/23802359.2019.1698368>
49. Yang L, Li J, Zhou G (2022) Comparative chloroplast genome analyses of 23 species in *Swertia* L. (Gentianaceae) with implications for its phylogeny. *Front Genet* 13:895146. <https://doi.org/10.3389/fgene.2022.895146>
50. Yang L, Deng S, Zhu Y, Da Q (2023) Comparative chloroplast genomics of 34 species in subtribe Swertiinae (Gentianaceae) with implications for its phylogeny. *BMC Plant Biol* 23(1):1-20. <https://doi.org/10.1186/s12870-023-04183-1>
51. Yang Y, Tao Z, Dong D, Yang J, Feng L, Zhao GF (2016) Comparative analysis of the complete chloroplast genomes of five quercus species. *Front Plant Sci* 7:959. <https://doi.org/10.3389/fpls.2016.00959>
52. Zhang X, Sun Y, Landis JB, Lv Z, Shen J, Zhang H, Lin N, Li L, Sun J, Deng T, et al (2020) Plastome phylogenomic study of Gentianeae (Gentianaceae): widespread gene tree discordance and its association with evolutionary rate heterogeneity of plastid genes. *BMC Plant Biol* 20:340. <https://doi.org/10.1186/s12870-020-02518-w>

Figures

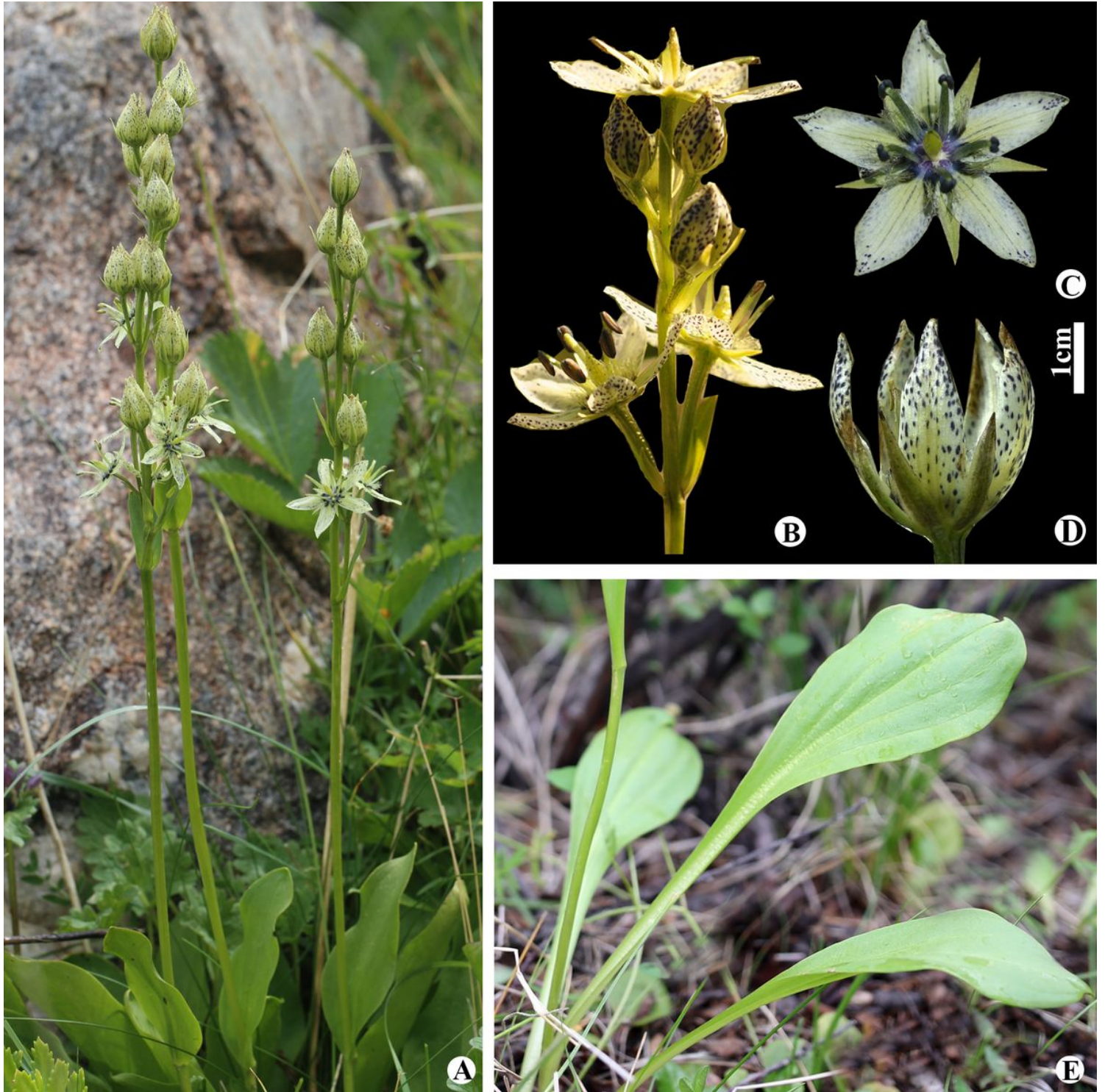


Figure 1

Swertia banzragczii in Mongolia: **A** general habitat; **B** inflorescence; **C** flower front view; **D** flower; **E** cauline leaves.



Figure 2

Swertia marginata in Mongolia: **A** general habitat; **B** flowers front view; **C, D** calyx; **E** cauline leaves.

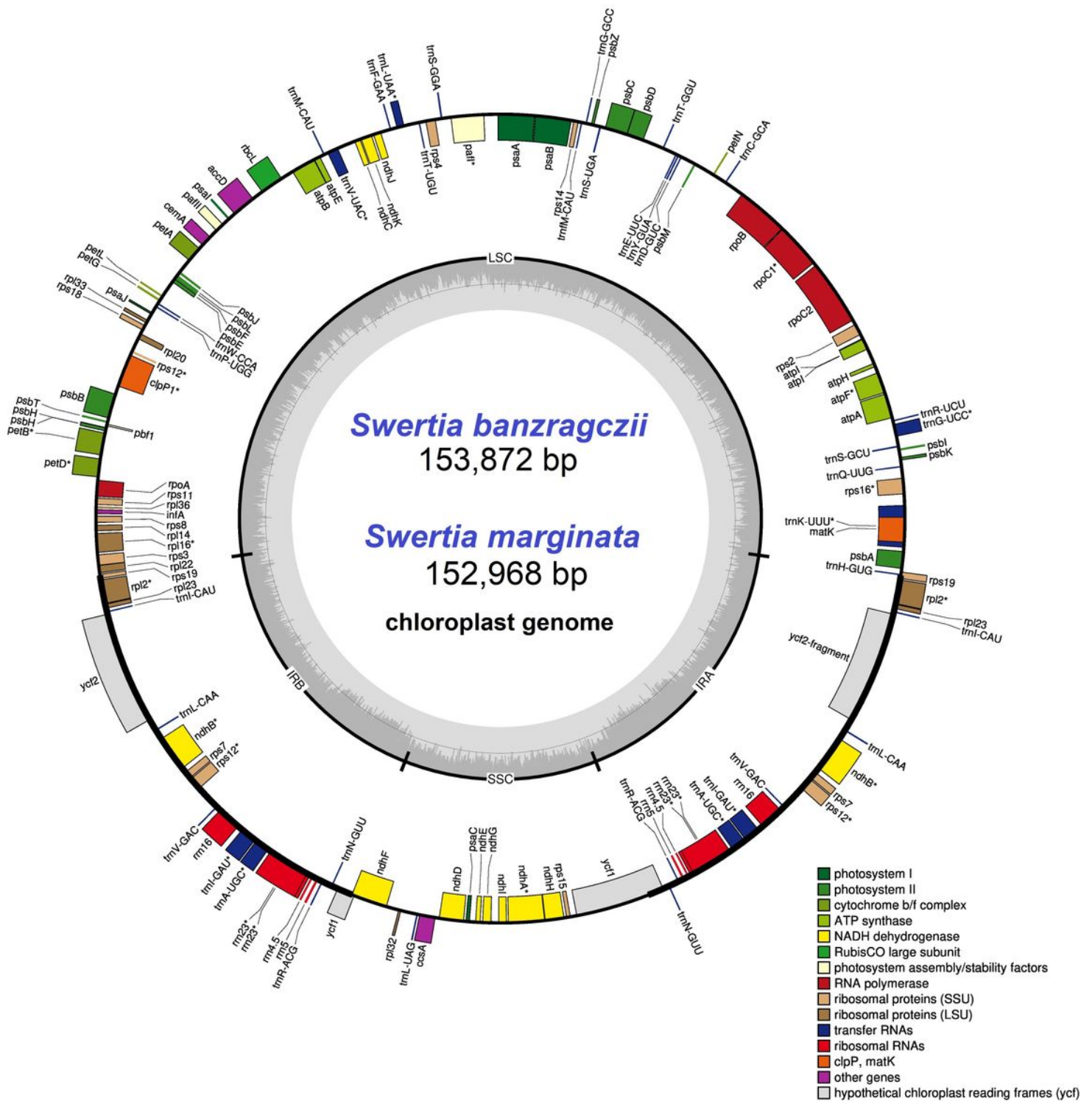


Figure 3

Circular map of the *S. banzragczii* and *S. marginata* complete chloroplast genomes. The inner ring is divided into four areas: SSC, IRb, LSC, and IRa. The inner ring indicates the GC content (dark gray or green) or the AT content (light gray or green). The genes in the outer ring region are transcribed clockwise, while those in the inner ring are transcribed counterclockwise. The legend classifies the cp genes based on their functions.

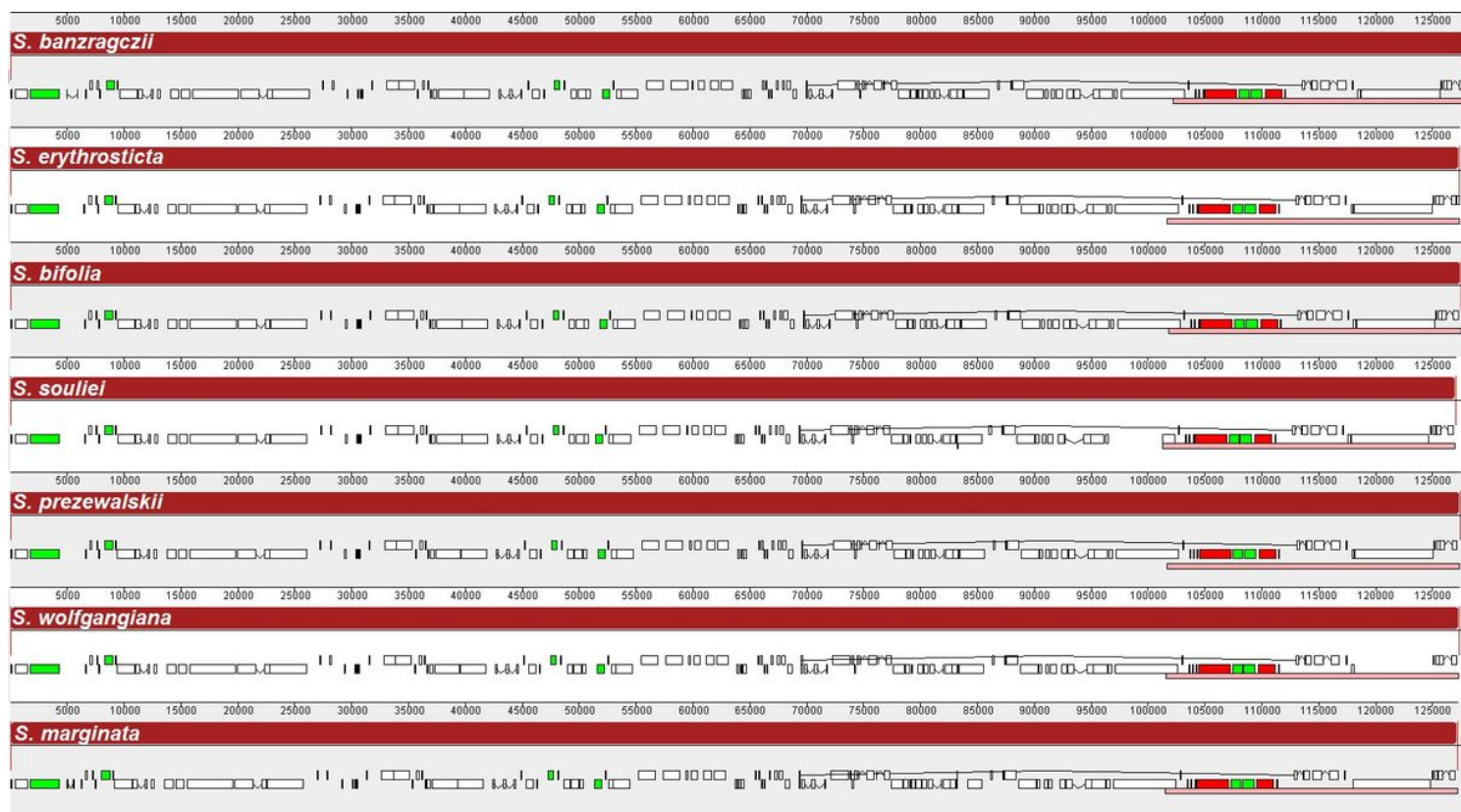


Figure 4

Comparison of complete plastomes from seven *Swertia* species using the MAUVE alignment. Blocks in the top of each row are in the same orientation, while blocks on the bottom of each row are in the inverse orientation. Boxes under each genome map represent protein-coding genes (white), rRNAs (red), and tRNAs (green). Red blocks in the lowest bottom row are inverted repeat (IR) regions.

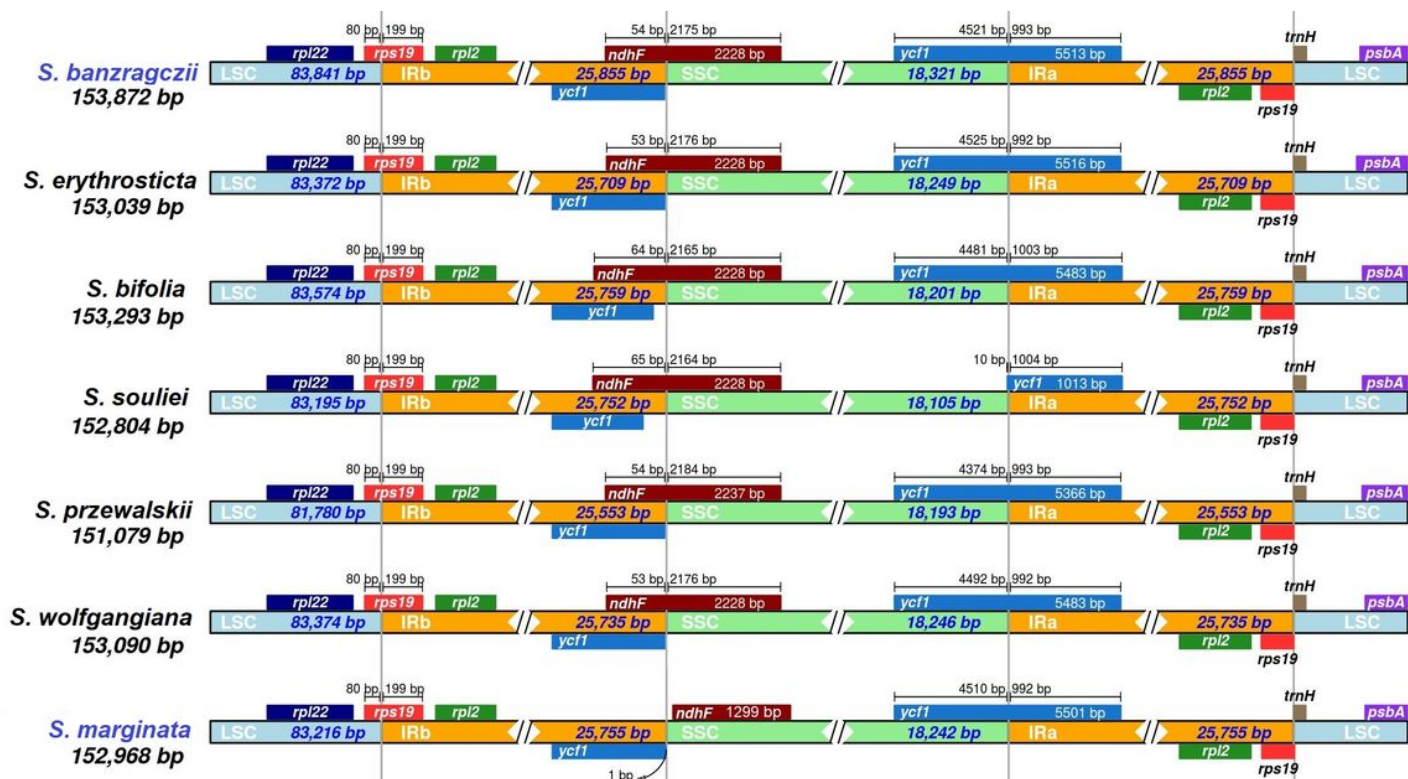


Figure 5

Comparison of LSC, IR, and SSC junction positions of *S. banzragczii* and *S. marginata* plastomes. LSC, large single-copy region; SSC, small single-copy region; IR, inverted repeat.

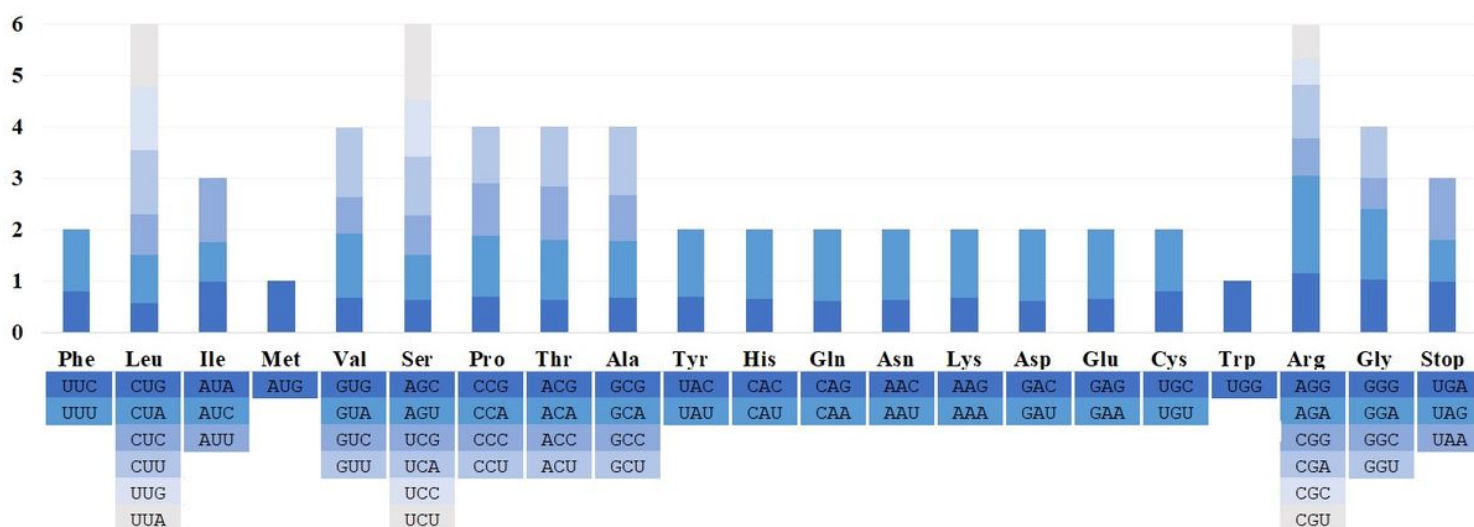


Figure 6

Codon usage for 20 amino acids in the protein-coding genes of *S. banzragczii* and *S. marginata*. The colors in the bars are the same as that of the codons below the figure. RSCU: relative synonymous codon usage.

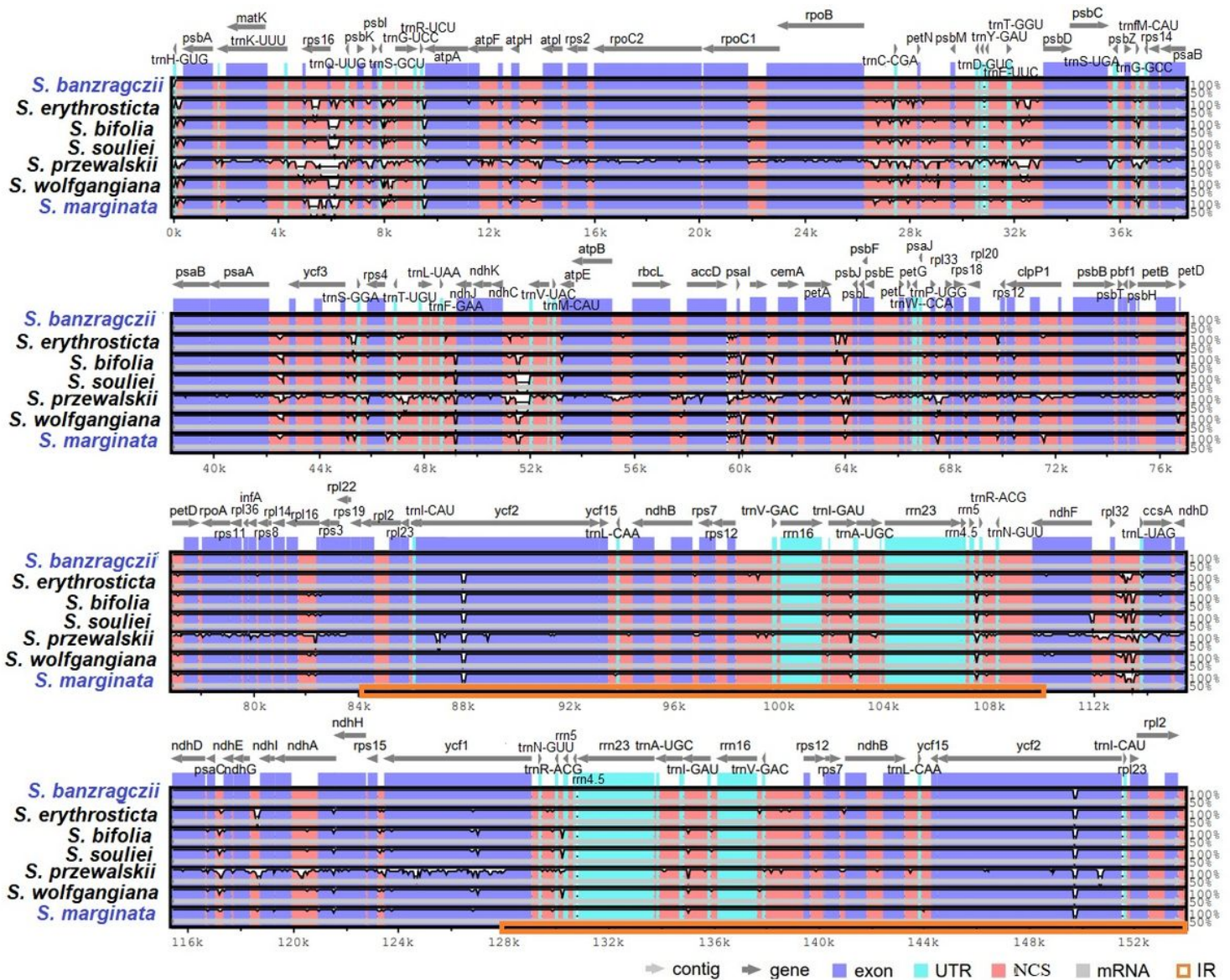


Figure 7

Comparison of complete plastomes from seven *Swertia* species, including *S. banzragczii* as a reference, using the mVISTA software. The gray arrows above the alignment indicate genes. Different colors represent different regions (coding and non-coding). The horizontal axis indicates the coordinates within the plastomes. The vertical scale represents the percentage of identity, ranging from 50 to 100%.

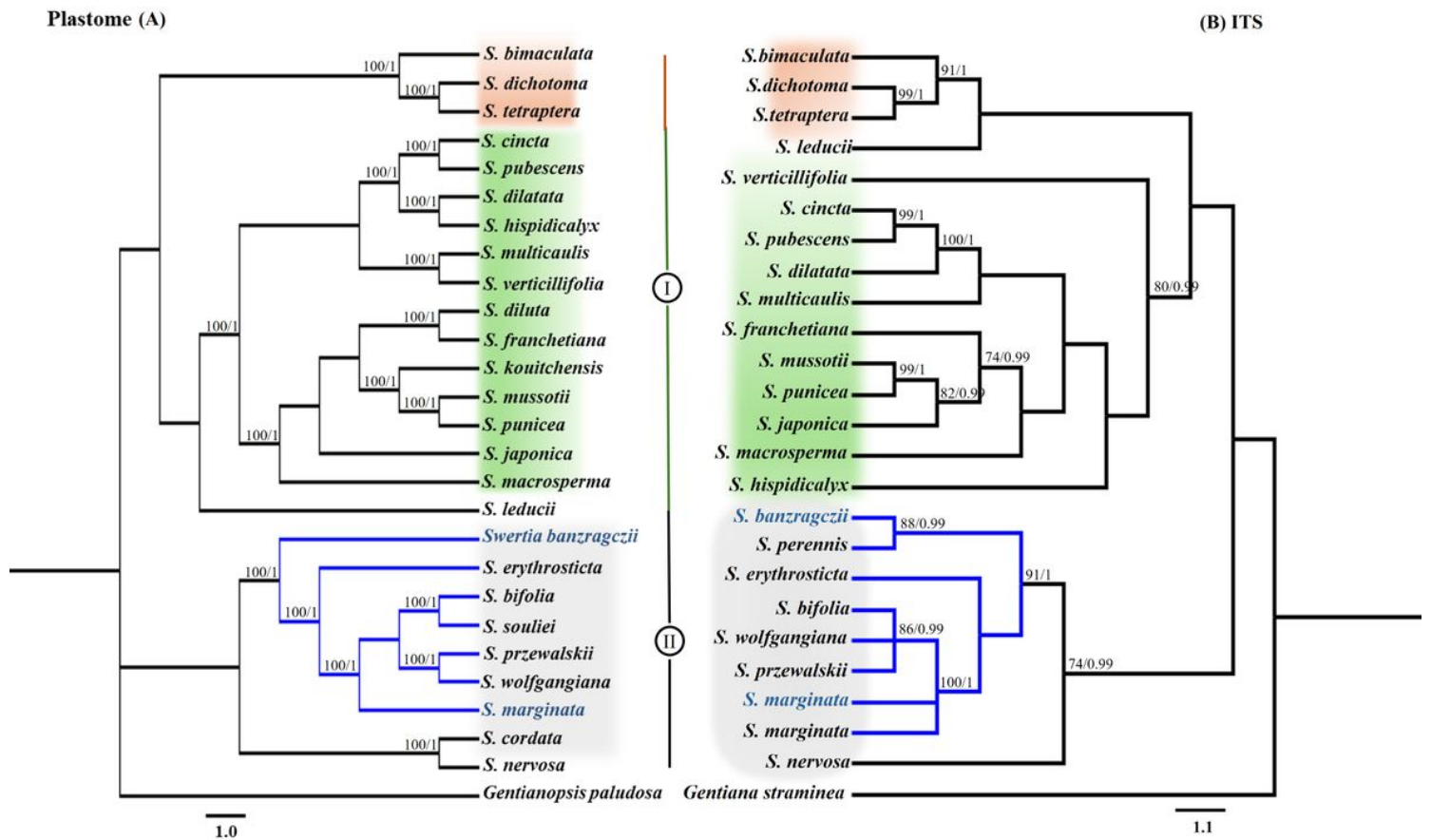


Figure 9

Phylogenetic tree of *Swertia banzragczii* and *Swertia marginata* with related taxa based on complete chloroplast genome (A) and nuclear ITS marker (B) Bayesian inference (BI) and maximum parsimony (MP) support values indicated at each branch, following the order BI/MP. The newly sequenced plastome samples are indicated in blue in both trees.