Complete Chloroplast Genomes of *Tulipa turkestanica* and *T. biflora* (Liliaceae): Structural Conservation and Repeat Variation in Central Asian Tulips

Umida Tojiboeva[a,b], Ziyoviddin Yusupov[c], Deng Tao[a,*]

a CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China
b University of Chinese Academy of Sciences, Beijing 100049, China
c Institute of Botany, Academy of Sciences of Uzbekistan, Tashkent, Uzbekistan

Article type	Dates		Accession
Research Article	Published online: 09.02.2026		TASH00160734 · 2003202104

DOI 10.66273/3134-6359.2026.1.1.006

Abstract

The genus *Tulipa* is taxonomically complex and highly diverse in Central Asia, yet genomic resources for many wild species remain limited. We sequenced and assembled the complete chloroplast genomes of *Tulipa turkestanica* and *T. biflora*, two morphologically variable species from Uzbekistan, to investigate plastome structure, repeat composition, and phylogenetic relationships. Both genomes exhibited conserved quadripartite organization with nearly identical sizes (152,022–152,025 bp) and each encoding 84 protein-coding genes, reflecting high structural stability within the genus. Comparative analyses revealed strong A/T bias in simple sequence repeats (SSRs) in both species. However, *T. turkestanica* displayed a higher number and greater diversity of SSR motifs, including species-specific dinucleotide repeats (particularly AT and AG/CT motifs), suggesting differential microsatellite accumulation during evolutionary divergence. Phylogenomic reconstruction based on complete chloroplast genomes provided robust resolution within subgenus *Eriostemones*, placing *T. biflora* with *T. sogdiana* and *T. turkestanica* with *T. buhseana*, both with strong bootstrap support (BS = 100). The *petD-rpoA* intergenic region exhibited the highest nucleotide diversity (π = 0.01625), representing a promising marker for population genetic studies. These genomic resources enhance our understanding of tulip systematics and provide valuable tools for conservation genetics in these threatened wild species.

Keywords: chloroplast genome; *Tulipa*; Central Asia; SSR; phylogenomics; plastome; conservation genetics; Liliaceae

Introduction

Tulipa L. (Liliaceae) is a species-rich and taxonomically complex genus comprising approximately 97 described species worldwide (POWO, 2025). Mountainous Central Asia represents the primary center of origin and diversification of the genus and is recognized as a global biodiversity hotspot (Myers et al., 2000; Botschantzeva, 1962). Early floristic surveys documented 63 wild *Tulipa* species in this region (Vvedensky and Kovalevskaya, 1971), while subsequent taxonomic revisions have substantially expanded this number (Zonneveld, 2009; de Groot and Zonneveld, 2024; Asatulloev et al., 2023).

Persistent taxonomic uncertainty in *Tulipa* arises from pronounced morphological variation, frequent interspecific hybridization, and historical classification inconsistencies (Hall, 1940; Christenhusz et al., 2013; Dekhkonov et al., 2022). The complex geological history of Central Asia, including mountain uplift and climatic

oscillations, has further promoted species diversification and endemism (Miao et al., 2012), supporting the hypothesis that the region represents the evolutionary cradle of tulips (Botschantzeva, 1962).

Despite their ecological and ornamental value, many *Tulipa* species are experiencing population declines, with 62 species currently listed in the IUCN Red List (2022). Among them, *Tulipa biflora* and *T. turkestanica* are widespread yet taxonomically challenging species characterized by high morphological variability and unresolved phylogenetic relationships. The scarcity of genomic resources has limited detailed evolutionary and comparative investigations for these taxa.

Chloroplast genomes provide powerful tools for resolving phylogenetic relationships and conducting comparative genomic analyses in plants. Their conserved structure, moderate evolutionary rate, and predominantly uniparental inheritance make them particularly valuable for systematic and evolutionary studies. Here, we present and compare the complete chloroplast genomes of *T. biflora* and *T. turkestanica* to elucidate their genomic features, clarify evolutionary relationships, and advance the systematics of *Tulipa* in Central Asia.

Materials and Methods

Plant Material and DNA Sequencing

Fresh leaf material of *T. biflora* and *T. turkestanica* was collected from natural populations in Uzbekistan. Voucher specimens were deposited in the TASH herbarium, and leaves were preserved in silica gel prior to DNA extraction. Genomic DNA was isolated using the DP305 Plant Genomic DNA Kit (Tiangen, China) following the manufacturer's protocol.

Sequencing libraries were prepared using the NEBNext Ultra™ DNA Library Prep Kit for Illumina (NEB, USA). DNA was fragmented to approximately 350 bp, followed by end repair, adapter ligation, PCR amplification, and purification using AMPure XP beads. Library quality was assessed using the Agilent 5400 Fragment Analyzer, and quantification was performed by qPCR. Paired-end sequencing (2 × 150 bp) was conducted on Illumina platforms at Novogene (Beijing, China).

Genome Assembly and Annotation

Clean reads were assembled into complete chloroplast genomes using the GetOrganelle pipeline v1.7.5 (Jin et al., 2020). Chloroplast genome annotation was performed in Geneious Prime v2023.1.2 using *Tulipa buhseana* (GenBank accession NC_052014) as a reference. All gene annotations were manually curated to verify gene boundaries and correct potential errors. Chloroplast genome maps were generated using OGDRAW v1.1 (Lohse et al., 2007).

Comparative and Phylogenetic Analyses

Simple sequence repeats (SSRs) in the chloroplast genomes were identified using the MISA web tool (Beier et al., 2017), with thresholds set to detect mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide motifs with minimum repeat units of 10, 5, 4, 3, 3, and 3, respectively. Nucleotide diversity (π) across complete chloroplast genomes and protein-coding regions was calculated using DnaSP v6.12.03 (Rozas et al., 2017). Sliding-window analyses were conducted with a window size of 600 bp and step size of 200 bp for individual genes, while whole-genome variability was assessed using default batch settings.

For phylogenetic reconstruction, the newly sequenced chloroplast genomes of *T. biflora* and *T. turkestanica* were aligned with available *Tulipa* and related genera chloroplast genomes retrieved from GenBank. Maximum likelihood (ML) phylogenetic analysis was performed using RAxML v8.2.12 (Stamatakis, 2014) with the GTR + I + G4 substitution model. Branch support was assessed using 1,000 bootstrap replicates.

Results

Chloroplast Genome Organization and Gene Content

The complete chloroplast genomes of *T. turkestanica* and *T. biflora* were 152,022 bp and 152,025 bp in length, respectively. Both genomes exhibited the conserved quadripartite structure characteristic of angiosperms, comprising a large single-copy (LSC) region, a small single-copy (SSC) region, and two inverted repeat (IR) regions (Fig. 1). Each genome encoded 84 protein-coding genes with identical gene composition. The high degree of structural conservation reflects the overall stability of plastome architecture within *Tulipa*.

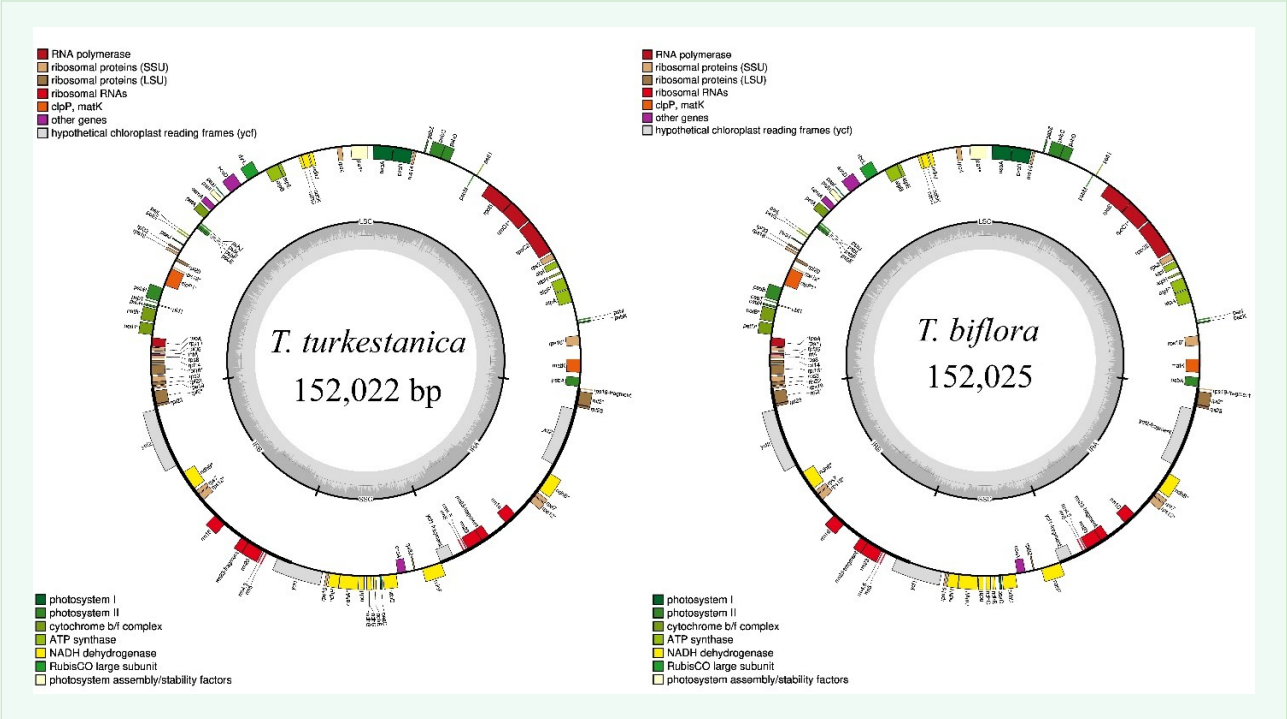

Figure 1. Chloroplast genome maps of *T. turkestanica* (left, 152,022 bp) and *T. biflora* (right, 152,025 bp). Genes shown on the outside of the circle are transcribed clockwise; those inside are transcribed counterclockwise. Genes are colour-coded by functional category as indicated in the legend. LSC — large single-copy region; SSC — small single-copy region; IR — inverted repeat.

Nucleotide Diversity

Nucleotide diversity analysis revealed that the *petD-rpoA* intergenic spacer region exhibited the highest variability (π = 0.01625), suggesting this region as a potential molecular marker for population genetic studies (Fig. 2). Protein-coding regions showed generally lower diversity, consistent with functional constraints on these sequences. The genome-wide sliding-window profile highlighted several variable hotspots distributed across the LSC region, while the IR regions displayed near-zero diversity reflecting their role in genome stabilisation through recombination.

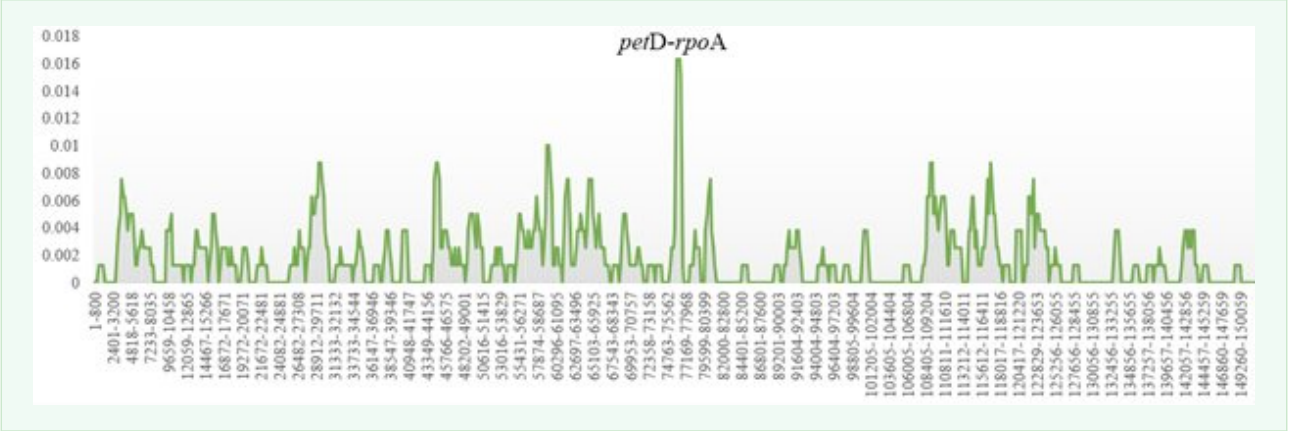

Figure 2. Nucleotide diversity (π) sliding-window analysis across the aligned chloroplast genomes of *T. turkestanica* and *T. biflora*. The *petD-rpoA* intergenic spacer region (π = 0.01625, annotated above the peak) shows the highest variability and represents a promising molecular marker. Window size: 600 bp; step size: 200 bp.

Simple Sequence Repeat Analysis

A total of 88 SSR motifs representing 23 distinct types were identified in *T. turkestanica*, whereas 71 SSR motifs representing 17 types were detected in *T. biflora* (Fig. 3). In both species, mononucleotide repeats were the most abundant, with adenine/thymine (A/T) motifs predominating. Specifically, A motifs occurred 29 times in *T. turkestanica* and 31 times in *T. biflora*, while T motifs were found 23 times in *T. turkestanica* and 24 times in *T. biflora*.

Dinucleotide repeats showed marked differences between the two species. In *T. turkestanica*, the AT motif was most common (10 occurrences), whereas only one AT repeat was observed in *T. biflora*. Furthermore, other dinucleotide motifs (AG, CT, GA, TA, and TC) were present at low frequencies in *T. turkestanica* but were absent or rare in *T. biflora*. AG/CT repeats were unique to *T. turkestanica* with seven occurrences, representing a species-specific marker.

Trinucleotide and higher-order repeats (tetranucleotide and pentanucleotide) were detected in both species but at comparatively low frequencies. Common motifs included ATA, TAT, TTA, AATA, AGAA, ATAA, ATCA, ATTT, GAAT, TAAT, TAGA, and TATT. Both plastomes exhibited a strong bias toward A/T-rich repeats, consistent with the high A/T content of chloroplast genomes. The greater number and diversity of SSRs in *T. turkestanica* suggests differential microsatellite accumulation dynamics between the two species.

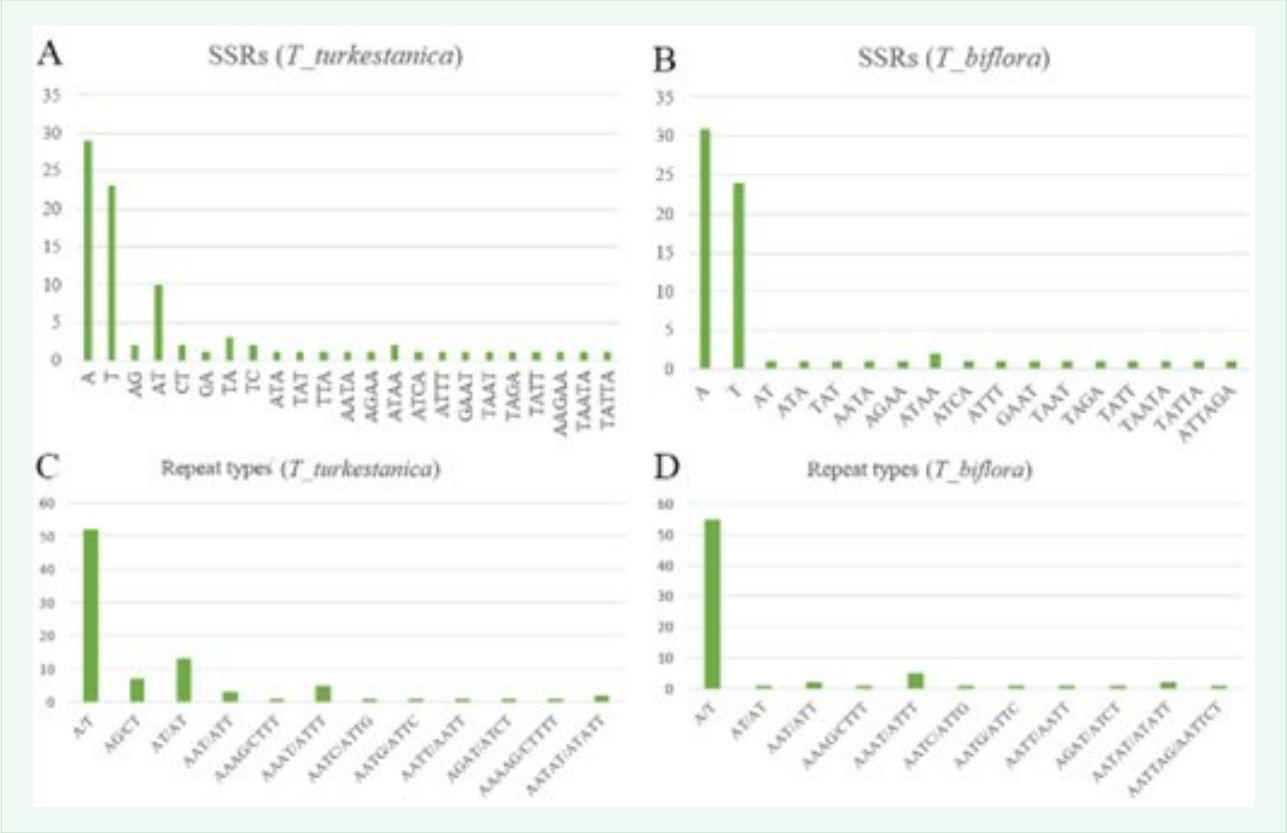

Figure 3. Simple sequence repeat (SSR) analysis of *T. turkestanica* and *T. biflora*. (A) SSR motif frequencies in *T. turkestanica* (88 SSRs, 23 types). (B) SSR motif frequencies in *T. biflora* (71 SSRs, 17 types). (C) Expanded repeat-type distribution in *T. turkestanica*. (D) Expanded repeat-type distribution in *T. biflora*. Note the species-specific AT and AG/CT dinucleotide repeats in *T. turkestanica*.

Phylogenetic Relationships

Maximum likelihood phylogenetic analysis based on complete chloroplast genomes resolved relationships within subgenus *Eriostemones* with strong bootstrap support (Fig. 4). *Tulipa biflora* formed a well-supported clade with *T. sogdiana* (BS = 100), while *T. turkestanica* clustered with two accessions of *T. buhseana*, also with maximum support (BS = 100). These results indicate close evolutionary relationships among these taxa and confirm their placement within distinct lineages of subgenus *Eriostemones*.

The phylogenetic topology demonstrated clear separation of additional species within the subgenus, while *T. sinkiangensis* (subgenus *Orithyia*) was clearly distinguished as the outgroup. Overall, the chloroplast genome-based phylogeny provides robust resolution of relationships within *Eriostemones* and supports the systematic placement of *T. biflora* and *T. turkestanica* within well-defined evolutionary lineages.

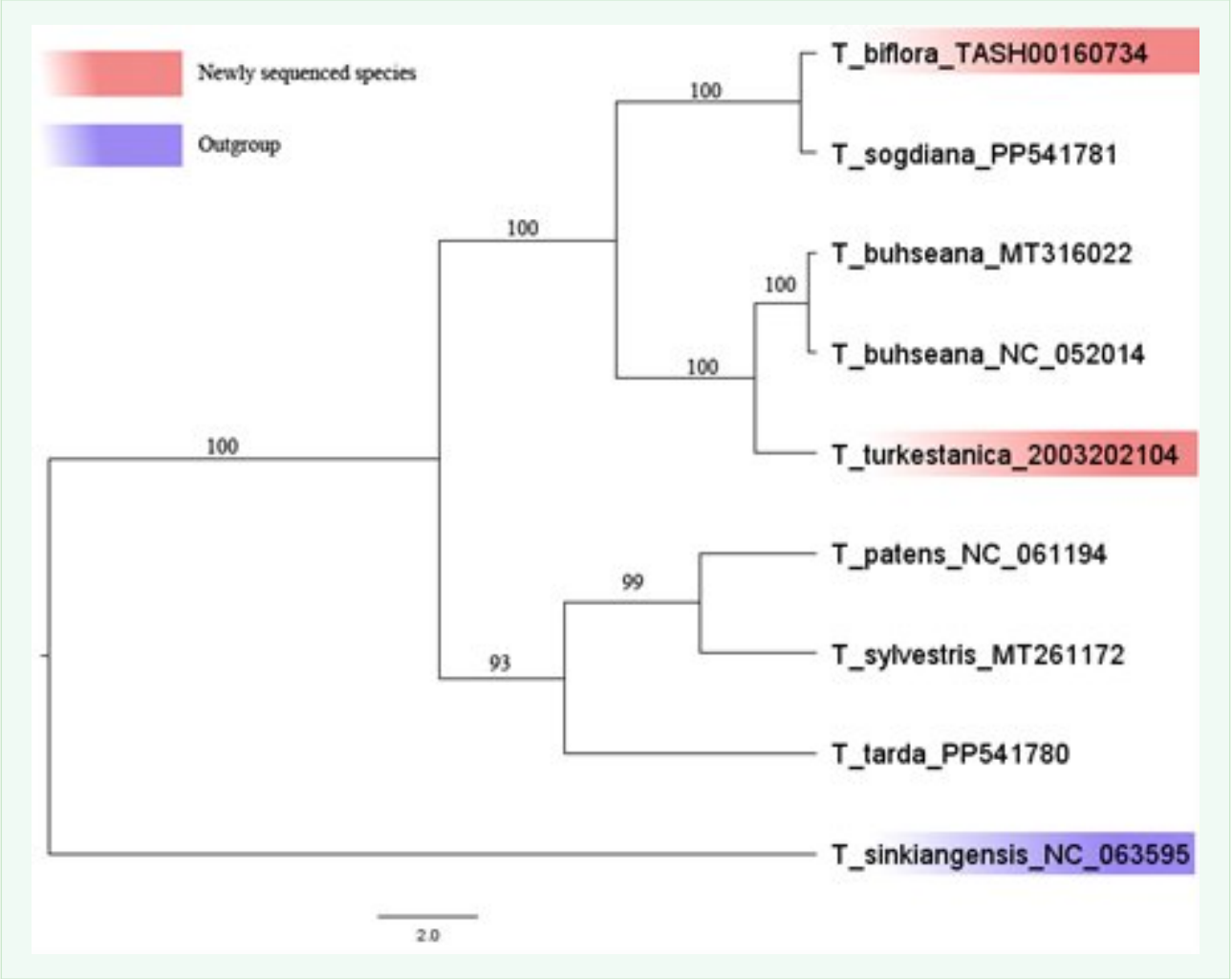

Figure 4. Maximum likelihood phylogenetic tree of subgenus *Eriostemones* based on complete chloroplast genomes. Newly sequenced species are highlighted in red; the outgroup (*T. sinkiangensis*, subgenus *Orithyia*) is highlighted in blue/purple. Bootstrap support values (1,000 replicates; ≥ 50%) are shown at internal nodes. Scale bar = 2.0 substitutions per site.

Discussion

Chloroplast Genome Conservation in *Tulipa*

The complete chloroplast genomes of *T. turkestanica* and *T. biflora* exhibited highly conserved structural organization, including the typical quadripartite architecture (LSC, SSC, and two IR regions) and nearly identical genome sizes. The presence of 84 protein-coding genes in both species, with complete gene content conservation, reflects the structural stability of chloroplast genomes across the genus *Tulipa*. This pattern is consistent with previous comparative studies in Liliaceae, which have documented limited genome rearrangements and gene losses within the family (She et al., 2020). Such structural uniformity indicates that major genomic reorganization events have played a minimal role in plastome evolution within this lineage.

Microsatellite Variation and Species Differentiation

Despite overall structural conservation, notable variation was observed in microsatellite composition between the two species. Both plastomes were strongly enriched in A/T-rich SSRs, a pattern widely documented in plant chloroplast genomes and generally attributed to replication slippage and the inherent AT bias of plastid DNA (Oliveira et al., 2006). However, *T. turkestanica* exhibited a higher total number (88 vs. 71) and greater diversity (23 vs. 17 types) of SSR motifs compared to *T. biflora*, suggesting differential accumulation of microsatellites

during evolutionary divergence.

The most striking difference was observed in dinucleotide repeats. *T. turkestanica* displayed significantly higher frequencies of AT motifs (10 occurrences) and possessed unique AG/CT repeats (7 occurrences), which were absent in *T. biflora*. These species-specific differences in repeat composition indicate distinct mutation dynamics and may reflect different evolutionary histories or population genetic processes. Such SSR polymorphisms represent valuable molecular markers that could be developed for population genetic studies, phylogeographic analyses, and species authentication within *Tulipa* (Raskina, 2017). The *petD-rpoA* intergenic region, showing the highest nucleotide diversity (π = 0.01625), provides an additional target for developing informative markers for evolutionary and conservation studies.

Phylogenetic Insights and Evolutionary Implications

Phylogenetic reconstruction based on complete chloroplast genomes provided robust resolution within subgenus *Eriostemones*. The close clustering of *T. biflora* with *T. sogdiana* and of *T. turkestanica* with *T. buhseana* corroborates morphological affinities previously proposed for these taxa (Dekhkonov et al., 2022) and confirms their placement within distinct evolutionary lineages. However, this chloroplast genome-based topology contrasts with previous phylogenies based on nuclear ITS markers (Kubentayev et al., 2024), suggesting the influence of introgression and incomplete lineage sorting.

Cytonuclear discordance is a common phenomenon in Liliaceae and has been extensively documented in recent phylogenomic studies of tribe Tulipeae (Zhang et al., 2025). This discordance typically arises from processes such as ancient hybridization events, asymmetric gene flow, and incomplete sorting of ancestral polymorphisms. In the case of Central Asian tulips, the complex geological history of the region — including repeated climatic oscillations and geographic fragmentation — likely promoted multiple opportunities for interspecific gene flow while simultaneously facilitating lineage diversification. The clear separation of *T. sinkiangensis* (subgenus *Orithyia*) from subgenus *Eriostemones* in our phylogeny supports the major subgeneric divisions within *Tulipa* and validates the use of chloroplast genomes for resolving higher-level systematic relationships.

Conservation and Future Directions

The comparative chloroplast genome analysis of *T. biflora* and *T. turkestanica* demonstrates that while plastome structure and gene content are highly conserved, repeat sequences and intergenic spacer regions provide informative variation for species differentiation and evolutionary studies. Given the ongoing taxonomic challenges and conservation concerns surrounding wild tulips — with 62 species currently listed on the IUCN Red List — the genomic resources generated in this study represent a valuable foundation for future research. These resources can support species delimitation efforts, facilitate the development of molecular markers for population genetic monitoring, enable phylogeographic studies to identify evolutionarily significant units, and inform evidence-based conservation strategies. Further integration of nuclear genome data with chloroplast genomic information will be essential for fully resolving the complex evolutionary history and systematic relationships within this ecologically and culturally important genus.

Acknowledgments

This research was supported by the CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany. We thank the staff of the TASH herbarium for assistance with specimen deposition and the field workers who helped with sample collection in Uzbekistan.

Data Availability

The complete chloroplast genome sequences of *T. biflora* and *T. turkestanica* have been deposited in GenBank under accession numbers TASH00160734 and 2003202104, respectively. All data supporting the conclusions are available from the corresponding author upon reasonable request.

References

Asatulloev B, Yusupov Z, Tojiboeva U, et al. (2023) Taxonomic revision of *Tulipa* (Liliaceae) in Uzbekistan. *Plant Diversity*.

Beier S, Thiel T, Münch T, et al. (2017) MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33:2583–2585.

Botschantzeva ZP (1962) Tulips: Taxonomy, Morphology, Cytology, Phytogeography and Physiology. Nauka, Leningrad.

Christenhusz MJ, Govaerts R, David JC, et al. (2013) Tiptoe through the tulips — cultural history, molecular phylogenetics and classification of *Tulipa*. *Bot J Linn Soc* 172:280–328.

de Groot YA, Zonneveld BJ (2024) Species recognition and classification in *Tulipa* (Liliaceae). *Phytotaxa*.

Dekhkonov T, Yusupov Z, Tojiboeva U, et al. (2022) New records and notes on *Tulipa* in Central Asia. *Nord J Bot*.

Hall AV (1940) A revision of *Tulipa*. *J S Afr Bot* 6:33–55.

Jin J-J, Yu W-B, Yang J-B, et al. (2020) GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol* 21:241.

Kubentayev SA, Yusupov Z, Tojiboeva U, et al. (2024) Phylogeny of Central Asian *Tulipa* based on ITS and plastid markers. *Plant Syst Evol*.

Lohse M, Drechsel O, Bock R (2007) OrganellarGenomeDRAW — OGDRAW. *Nucleic Acids Res* 35:W575–W580.

Miao B-G, Zhang M-L, Zhang Z-Y (2012) Diversification of *Tulipa* and its biogeographic implications. *Plant Syst Evol* 298:981–992.

Myers N, Mittermeier RA, Mittermeier CG, et al. (2000) Biodiversity hotspots for conservation priorities. *Nature* 403:853–858.

Oliveira EJ, Pádua JG, Zucchi MI, et al. (2006) Origin, evolution and genome distribution of microsatellites. *Genet Mol Biol* 29:294–307.

POWO (2025) Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. https://powo.science.kew.org/

Raskina O (2017) Microsatellite evolution in plant chromosomes. *Int J Mol Sci*.

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, et al. (2017) DnaSP 6: DNA sequence polymorphism analysis. *Mol Biol Evol* 34:3299–3302.

She C-W, Jiang X-H (2020) Chloroplast genome evolution in Liliaceae. *Front Plant Sci*.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.

Vvedensky AI, Kovalevskaya SS (1971) Flora of Uzbekistan, Vol. 6. Fan, Tashkent.

Zhang F, Li X, Yusupov Z, et al. (2025) Phylogenomics of tribe Tulipeae resolves cytonuclear discordance. *Mol Phylogenet Evol*.

Zonneveld BJM (2009) New record holders for maximum genome size in eudicots and monocots. *J Bot*.

Citation: Tojiboeva U, Yusupov Z, Deng T (2026) Complete Chloroplast Genomes of *Tulipa turkestanica* and *T. biflora* (Liliaceae): Structural Conservation and Repeat Variation in Central Asian Tulips. *Contig.* 1:202606. DOI: 10.66273/3134-6359.2026.1.1.006

Correspondence: Deng Tao, Kunming Institute of Botany, CAS · **Reviewed by:** Ruslan Kalendar