



## OPEN ACCESS

## EDITED BY

Khurram Shahzad,  
University of Nebraska-Lincoln, United States

## REVIEWED BY

Lu Gong,  
Guangzhou University of Traditional Chinese  
Medicine, China  
Xiaoyun Wang,  
Jiangxi University of Traditional Chinese  
Medicine, China  
Rongpeng Liu,  
Beijing University of Chinese Medicine, China  
Minh Trong Quang,  
Ho Chi Minh City Medicine and Pharmacy  
University, Vietnam

## \*CORRESPONDENCE

Qin-Qin Li

✉ liqq@imnu.edu.cn

<sup>†</sup>These authors have contributed equally to  
this work

RECEIVED 26 September 2025

REVISED 09 November 2025

ACCEPTED 13 November 2025

PUBLISHED 04 December 2025

CORRECTED 09 December 2025

## CITATION

Fu W-T, Zhang Z-P, Guo J-J, Wen J and  
Li Q-Q (2025) Comparative analyses  
of chloroplast genomes in *Geum*  
species: insights into genome  
characteristics, phylogenomic  
implications, and adaptive evolution.  
*Front. Plant Sci.* 16:1713809.  
doi: 10.3389/fpls.2025.1713809

## COPYRIGHT

© 2025 Fu, Zhang, Guo, Wen and Li. This is an  
open-access article distributed under the terms  
of the [Creative Commons Attribution License](#)  
(CC BY). The use, distribution or reproduction  
in other forums is permitted, provided the  
original author(s) and the copyright owner(s)  
are credited and that the original publication  
in this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Comparative analyses of chloroplast genomes in *Geum* species: insights into genome characteristics, phylogenomic implications, and adaptive evolution

Wen-Tao Fu<sup>1,2†</sup>, Zhi-Ping Zhang<sup>3†</sup>, Jia-Jie Guo<sup>1</sup>, Jun Wen<sup>4</sup>  
and Qin-Qin Li<sup>1,2\*</sup>

<sup>1</sup>College of Life Science and Technology, Inner Mongolia Normal University, Hohhot, China, <sup>2</sup>Key  
Laboratory of Biodiversity Conservation and Sustainable Utilization in Mongolian Plateau for College  
and University of Inner Mongolia Autonomous Region, Hohhot, China, <sup>3</sup>College of Computer Science  
and Technology, Inner Mongolia Normal University, Hohhot, China, <sup>4</sup>Department of Botany, National  
Museum of Natural History, Smithsonian Institution, Washington, DC, United States

The genus *Geum*, comprises about 72 species, most frequently distributed in North America, Asia, and Europe, with a few representatives in South America, South Africa, Australia, and New Zealand. Previous phylogenetic analyses based on several molecular markers have contributed to understanding the delimitation of *Geum*, but the phylogenetic relationships within the genus remain unresolved. Moreover, only a few chloroplast (cp) genomes of *Geum* species have been reported, and no comparative cp genome analyses among *Geum* species have been conducted to date, limiting our understanding of cp genome evolution. This study is the first to conduct comparative genomic analyses on the cp genomes of 32 accessions representing 11 *Geum* taxa. The *Geum* cp genomes showed a typical quadripartite structure, similar to that of most other land plants, with a total of 129 genes, including 84 protein-coding genes (PCGs), 37 transfer RNA (tRNA) genes, and eight ribosomal RNA (rRNA) genes. The *Geum* cp genomes were conserved in structure, size, GC content, gene order, and gene content. Eleven highly variable regions (*3'-trnK-UUU-matK*, *psbZ-trnG-GCC*, *trnR-UCU-atpA*, *petA-psbJ*, *5'-trnK-UUU-rps16*, *rps16-trnQ-UUG*, *rpl32-trnL-UAG*, *ndhF-rpl32*, *trnS-GCU-trnG-UCC*, *ndhC-trnV-UAC*, and *petN-psbM*) were identified as candidate molecular markers for future studies on population genetics and systematic evolution of *Geum* species. Phylogenetic analyses provided new insights into the relationships among *Geum* species and supported Smedmark's recircumscription of *Geum* in a broad sense, corroborating the inclusion of *Acomastylis*, *Coluria*, and *Taihangia* within

*Geum*. Twenty-three genes with sites under positive selection were detected, and the adaptive evolution of these genes may play important roles in the adaptation of *Geum* species to their habitats. Overall, this study enhances our understanding of the cp genome characteristics, phylogeny, and adaptive evolution of *Geum* species.

#### KEYWORDS

*Geum*, chloroplast genome, comparative analyses, phylogeny, adaptive evolution

## 1 Introduction

The chloroplast (cp) is a semiautonomous organelle in green plants that plays key roles in photosynthesis and other aspects of plant physiology and development (Neuhaus and Emes, 2000; Sato et al., 2003; Daniell et al., 2016). The cp genomes of land plants are circular, double-stranded molecules, mostly ranging from 115 to 165 kb in size and containing 120–130 genes (Ravi et al., 2008; Daniell et al., 2016). Land plant cp genomes usually display a typical quadripartite structure, with a large single-copy region (LSC) plus a small single-copy region (SSC) separated by two inverted repeat regions, IRa and IRb (Ravi et al., 2008; Wicke et al., 2011; Daniell et al., 2016). Owing to their small size, moderate substitution rate, conserved structure, and lack of recombination (Palmer, 1985; Drouin et al., 2008; Ravi et al., 2008; Mower and Vickrey, 2018), cp genomes have become important tools for studies on species identification, population genetics, taxonomy, biogeography, and systematic evolution in land plants, particularly with the development of high-throughput sequencing technology (e.g., Li et al., 2021; Wang et al., 2022; Zhang G. J. et al., 2022; Hu et al., 2023; Li et al., 2024a; Zhou et al., 2024; Rana et al., 2025).

As the largest genus, comprising about 72 of the 75 species in its tribe, *Geum* in the sense of Smedmark (2006), together with the other two woody and white-flowered genera *Fallugia* and *Sieversia*, constitutes the tribe Colurieae (Smedmark and Eriksson, 2002; Smedmark et al., 2003; Smedmark, 2006). From a geographical perspective, the monospecific *Fallugia* occurs in the southwestern USA and northern Mexico, *Sieversia* occurs in Alaska and East Asia, and species of *Geum* are most frequently distributed in North America, Asia, and Europe, with a few representatives in South America, South Africa, Australia, and New Zealand (Gajewski, 1959; Smedmark and Eriksson, 2002; Brouillet, 2014; Henrickson and Parfitt, 2014; Phipps, 2014; Rohrer, 2014). In addition to their ornamental value, the primary importance of *Geum* species lies in their medicinal properties. Some *Geum* species have been used in traditional medicine for the treatment of various conditions, including leucorrhea, hemorrhages, gingivitis, muscle pain, gastrointestinal disorders, cardiac disorders, infections, fever, and inflammation of the skin, mucous membranes, and urinary system (Hänsel et al., 1993; Vollmann and Schultze, 1995; Birnesser and Stolt, 2007; Redžić, 2007; Menković et al., 2011; Granica et al., 2016;

Blaschek et al., 2018). The classification of *Geum*, which over the last century was mainly based on morphological evidence, cytogenetic studies, and interspecific crossings, has been ambiguous and conflicting (e.g., Scheutz, 1870; Focke, 1894; Rydberg, 1913; Bolle, 1933; Juzepchuk, 1941; Gajewski, 1957, 1959, 1968; Schulze-Menz, 1964; Hutchinson, 1967; Kalkman, 1988). Later molecular phylogenetic studies based on the cp *trnL-trnF* region and nuclear ribosomal ITS (Smedmark and Eriksson, 2002), as well as the low-copy nuclear gene GBSSI (Smedmark et al., 2003, 2005), did not support the monophyly of any of the previously proposed circumscriptions of *Geum*, and provided good support for delimiting the herbaceous perennials with a rosette of basal leaves in the tribe Colurieae as *Geum* in a broad sense (Smedmark, 2006). *Geum*, with this broad recircumscription (sensu Smedmark, 2006), embraces 12 historically segregated genera, namely *Waldsteinia*, *Stylipus*, *Coluria*, *Acomastylis*, *Erythrocoma*, *Novosieversia*, *Oncostylus*, *Parageum*, *Orthurus*, *Woronowia*, *Taihangia*, and *Oreogeum*. Previous phylogenetic analyses based on several molecular markers have contributed to understanding the delimitation of *Geum*, but the phylogenetic relationships within the genus remain unresolved (Smedmark and Eriksson, 2002; Smedmark et al., 2003, 2005; Faghir et al., 2018). Moreover, only a few cp genomes of *Geum* species have been reported (e.g., Li and Wen, 2021; Feng et al., 2022; Guo et al., 2023), and, to the best of our knowledge, no comparative cp genome analyses among *Geum* species have been conducted to date, limiting our understanding of cp genome evolution in this genus.

This is the first study to conduct comparative genomic analyses of the cp genomes of 32 accessions representing 11 *Geum* taxa. The aims were to (1) analyze the cp genome characteristics of *Geum* species to explore its cp genome evolution, (2) identify mutational hotspot regions across the cp genomes of *Geum* as potential molecular markers for species identification and phylogenetic studies, (3) provide insights into the phylogenetic relationships among *Geum* species to enhance understanding of their classification, and (4) investigate the adaptive evolution of cp genes in *Geum* species to understand their molecular adaptation. This study lays a foundation for future research on molecular identification, phylogenetics, and cp genome evolution of *Geum* species, and also provides an important theoretical basis for the development and utilization of the medicinal plant resources of *Geum*.

## 2 Materials and methods

### 2.1 Taxon sampling, DNA extraction, and Illumina sequencing

A total of 32 cp genome sequences representing 11 *Geum* taxa (17 of which were newly sequenced) were sampled in this study. The 17 new samples of *Geum* were collected during field trips, and species identification of the collected samples was conducted using an optical microscope with reference to relevant literature (Yü and Kuan, 1985; Yü and Li, 1985; Li et al., 2003). Voucher specimen information and GenBank accession numbers for *Geum* samples are presented in Table 1. In addition, cp genome sequences of *Fallugia paradoxa*, *Potentilla suavis*, *Rosa multiflora*, *Agrimonia pilosa*, and *Rubus alceifolius* downloaded from GenBank were included in the phylogenetic analyses, following Zhang et al. (2017). Total genomic DNA was isolated from silica-dried leaves using the CTAB method (Doyle and Doyle, 1987). Sonication was then used to fragment the DNA, and the DNA fragments were used to construct short-insert libraries with an insert size of 300 bp. Finally, the pooled libraries were sequenced on the Illumina NovaSeq platform in Novogene (Beijing, China).

### 2.2 Chloroplast genome assembly and annotation

Illumina paired-end sequencing generated about 5.0 Gb of raw data for each *Geum* sample. Adapters were removed from the raw reads using Trimmomatic (Bolger et al., 2014). NOVOPlasty (Dierckxsens et al., 2017) was employed to assemble the newly sequenced cp genomes from the filtered reads. During assembly, the cp genome of *Geum macrophyllum* (GenBank Accession No. MT774132) was used as the reference sequence, with its *rbcL* gene as seed input, and all other parameters set to default. After successfully assembling the cp genome sequences of some species, these sequences were used as references to assemble cp genomes of other accessions or closely related species. Using the cp genome of *G. macrophyllum* (MT774132) as the reference, cp genome annotations of *Geum* species downloaded from GeneBank were checked, and the cp genome sequence of *Geum elatum* (MT982432) was annotated by transfer annotation in Geneious Prime (Kearse et al., 2012). For newly sequenced *Geum* cp genome sequences, annotated sequences of the same species or closely related species were selected for transfer annotation.

### 2.3 Comparative analyses of chloroplast genomes

Comparative analyses were conducted on the complete cp genomes of 28 accessions representing *Geum* taxa. The whole cp genome size, lengths of the LSC/SSC/IR, Guanine-Cytosine (GC) content, gene composition, and boundary region variation were analyzed in Geneious Prime, and the variation in the LSC/IR/SSC

boundary regions was illustrated. The cp genomes of *Geum* were aligned using MAUVE v. 2.4.0 (Darling et al., 2004, 2010) to identify potential rearrangements and inversions. The level of differentiation among the *Geum* cp genomes was assessed using the Shuffle-LAGAN mode in mVISTA (Frazer et al., 2004) with *Geum aleppicum* 1 as the reference. Coding and noncoding regions of the *Geum* cp genomes were extracted in Geneious Prime, and homologous loci were then aligned by MAFFT v. 7.490 (Katoh and Standley, 2013). Nucleotide variability (Pi) of each region was calculated using DnaSP v. 6.12.03 (Rozas et al., 2017). Both sequence lengths and Pi values were considered to screen candidate molecular markers for *Geum*. A tree-based method was further employed to evaluate the resolution power of the screened candidate molecular markers compared to the core DNA barcodes (*trnH-GUG-psbA*, *rbcL*, and *matK*). MEGA v. 12.0.11 (Kumar et al., 2024) was used to construct neighbor-joining (NJ) trees based on each molecular marker, using the “pairwise deletion” option to treat gaps/missing data and the “d: Transitions + Transversions” option for substitutions under the Kimura 2-parameter model, with 1,000 bootstrap replicates.

### 2.4 Phylogenetic analyses

Maximum likelihood (ML) and Bayesian inference (BI) methods were used to infer the phylogenetic relationships of the 11 *Geum* taxa within the phylogenetic framework of the tribe Colurieae. Based on previous studies (Zhang et al., 2017; Xiang et al., 2017), *Agrimonia pilosa*, *Potentilla suavis*, *Rosa minutifolia*, and *Rubus alceifolius* were selected as outgroups. A total of 37 cp genome sequences, with the IRa removed, were used for phylogenetic analyses (Supplementary Table S1), and these sequences were first aligned using MAFFT v. 7.490 (Katoh and Standley, 2013). The alignment was then trimmed using trimAL v. 1.4 (Capella-Gutiérrez et al., 2009) with a 0.75 gap threshold. RAxML v. 8.2.12 (Stamatakis, 2014) was employed to conduct the ML analysis under the GTRGAMMA model with 1,000 bootstrap replicates. Prior to the BI analysis, the best-fit model was selected using PartitionFinder2 (Lanfear et al., 2017) according to the Corrected Akaike Information Criterion (AICc; Sugiura, 1978) following Posada and Buckley (2004). The BI analysis was then performed using MrBayes v. 3.2.7a (Ronquist et al., 2012) under the best-fit model GTR + I + G. Markov Chain Monte Carlo (MCMC) analyses included four parallel runs with one million generations, sampled every 100 generations, with the initial 25% of trees discarded as burn-in, and the remaining trees used to generate a consensus tree.

### 2.5 Adaptive evolution analyses

Selection pressure analyses were conducted using CodeML (Yang, 2007) implemented in EasyCodeML (Gao et al., 2019), involving 32 complete cp genomes representing nine *Geum* taxa (28 accessions) and four related species (*Potentilla suavis*, *Rosa multiflora*, *Agrimonia pilosa*, *Rubus alceifolius*). First, the 78 common protein-coding genes (PCGs) were extracted from the 32 cp genomes using Geneious Prime (Kearse et al., 2012). Each PCG

TABLE 1 Summary of voucher specimens and chloroplast genome characteristics for *Geum* species.

Species name (synonym)	Voucher	Locality	GenBank accession	Size (bp)				Number of genes				GC content (%)				References
				Total	LSC	SSC	IR	Total	Protein coding	tRNA	rRNA	Total	LSC	SSC	IR	
<i>Geum aleppicum</i> Jacq. 1			OK509085	156,036	85,358	18,410	26,134	129	84 (6)	37 (7)	8 (4)	36.8	34.4	30.7	42.6	Zhang P. et al. (2022)
<i>Geum aleppicum</i> Jacq. 2			OM461318	155,911	85,382	18,289	26,120	129	84 (6)	37 (7)	8 (4)	36.8	34.4	30.7	42.7	Unpublished
<i>Geum aleppicum</i> Jacq. 3	Li QQ 20220722006 (NMTC)	China, Jilin, Antu	PX414088	155,940	85,370	18,320	26,125	129	84 (6)	37 (7)	8 (4)	36.8	34.4	30.7	42.7	This article
<i>Geum aleppicum</i> Jacq. 4	Li QQ 20220716002 (NMTC)	China, Sichuan, Daofu	PX414089	156,038	85,362	18,418	26,129	129	84 (6)	37 (7)	8 (4)	36.8	34.5	30.7	42.6	This article
<i>Geum aleppicum</i> Jacq. 5	Li QQ 20220812064 (NMTC)	China, Inner Mongolia, Hohhot	PX414090	156,038	85,362	18,418	26,129	129	84 (6)	37 (7)	8 (4)	36.8	34.5	30.7	42.6	This article
<i>Geum aleppicum</i> Jacq. 6	Li QQ 20160717001 (NMTC)	China, Xinjiang, Fuyun	PX414091	155,943	85,373	18,320	26,125	129	84 (6)	37 (7)	8 (4)	36.8	34.4	30.8	42.7	This article
<i>Geum elatum</i> Wall. ex G.Don var. <i>elatum</i> 1 [ <i>Acomastylis elata</i> (Wall. ex G.Don) F.Bolle var. <i>elata</i> ]	Li QQ 20150822024 (NMTC)	China, Sichuan, Dege	PX414092	156,114	85,452	18,402	26,130	129	84 (6)	37 (7)	8 (4)	36.7	34.3	30.8	42.6	This article
<i>Geum elatum</i> Wall. ex G.Don var. <i>elatum</i> 2 [ <i>Acomastylis elata</i> (Wall. ex G.Don) F.Bolle var. <i>elata</i> ]			MT982432	156,104	85,507	18,501	26,048	129	84 (6)	37 (7)	8 (4)	36.7	34.3	30.8	42.7	Unpublished
<i>Geum elatum</i> Wall. ex G.Don var. <i>elatum</i> 3 [ <i>Acomastylis elata</i> (Wall. ex G.Don) F.Bolle var. <i>elata</i> ]	Li QQ 20150729073 (NMTC)	China, Sichuan, Yajiang	PX414093	156,145	85,544	18,509	26,046	129	84 (6)	37 (7)	8 (4)	36.7	34.3	30.8	42.7	This article
<i>Geum elatum</i> Wall. ex G.Don var. <i>elatum</i> 4 [ <i>Acomastylis elata</i> (Wall. ex G.Don) F.Bolle var. <i>elata</i> ]	Li QQ 20230720004 (NMTC)	China, Xizang, Yadong	PX414094	156,248	85,610	18,530	26,054	129	84 (6)	37 (7)	8 (4)	36.7	34.3	30.8	42.7	This article
<i>Geum elatum</i> Wall. ex G.Don var. <i>elatum</i> 5 [ <i>Acomastylis elata</i> (Wall. ex G.Don) F.Bolle var. <i>elata</i> ]			KY419976													Zhang et al. (2017)
<i>Geum elatum</i> Wall. ex G.Don var. <i>humile</i> (Royle) Hook.f. [ <i>Acomastylis elata</i> (Wall. ex G.Don) F.Bolle var. <i>humilis</i> (Royle) F.Bolle]	Li QQ LWQ0819002 (NMTC)	China, Xizang, Leiwuqi	PX414095	156,121	85,508	18,501	26,056	129	84 (6)	37 (7)	8 (4)	36.7	34.3	30.8	42.7	This article

(Continued)

TABLE 1 Continued

Species name (synonym)	Voucher	Locality	GenBank accession	Size (bp)				Number of genes				GC content (%)				References
				Total	LSC	SSC	IR	Total	Protein coding	tRNA	rRNA	Total	LSC	SSC	IR	
<i>Geum henryi</i> (Batalin) Smedmark 1 [ <i>Coluria henryi</i> Batalin]	Li QQ 20220717008 (NMTC)	China, Sichuan, Guangwu Mountain	PX414096	155,175	85,563	18,466	25,573	129	84 (6)	37 (7)	8 (4)	36.7	34.3	30.7	42.8	This article
<i>Geum henryi</i> (Batalin) Smedmark 2 [ <i>Coluria henryi</i> Batalin]	Li QQ 20220718002 (NMTC)	China, Sichuan, Micang Mountain	PX414097	155,291	85,740	18,419	25,566	129	84 (6)	37 (7)	8 (4)	36.7	34.3	30.8	42.8	This article
<i>Geum japonicum</i> Thunb. var. <i>chinense</i> F.Bolle 1			MW770453	155,999	85,333	18,410	26,128	129	84 (6)	37 (7)	8 (4)	36.8	34.4	30.7	42.6	Unpublished
<i>Geum japonicum</i> Thunb. var. <i>chinense</i> F.Bolle 2			MW770454	155,912	85,375	18,297	26,120	129	84 (6)	37 (7)	8 (4)	36.8	34.4	30.7	42.7	Unpublished
<i>Geum japonicum</i> Thunb. var. <i>chinense</i> F.Bolle 3	Li QQ 20220711012 (NMTC)	China, Sichuan, Dayi	PX414098	156,007	85,340	18,411	26,128	129	84 (6)	37 (7)	8 (4)	36.8	34.5	30.7	42.6	This article
<i>Geum japonicum</i> Thunb. var. <i>chinense</i> F.Bolle 4	Li QQ 20220709002 (NMTC)	China, Sichuan, Duijiangyan	PX414099	156,009	85,342	18,411	26,128	129	84 (6)	37 (7)	8 (4)	36.8	34.4	30.7	42.6	This article
<i>Geum japonicum</i> Thunb. var. <i>chinense</i> F.Bolle 5	Li QQ 20220713001 (NMTC)	China, Sichuan, Mount Emei	PX414100	156,013	85,347	18,410	26,128	129	84 (6)	37 (7)	8 (4)	36.8	34.4	30.7	42.6	This article
<i>Geum japonicum</i> Thunb. var. <i>chinense</i> F.Bolle 6	Li QQ 20160807032 (NMTC)	China, Yunnan, Luquan	PX414101	155,856	85,375	18,271	26,105	129	84 (6)	37 (7)	8 (4)	36.8	34.4	30.8	42.7	This article
<i>Geum japonicum</i> Thunb. var. <i>chinense</i> F.Bolle 7	Li QQ 20150728012 (NMTC)	China, Sichuan, Luding	PX414102	155,831	85,350	18,271	26,105	129	84 (6)	37 (7)	8 (4)	36.8	34.4	30.8	42.7	This article
<i>Geum longifolium</i> (Maxim.) Smedmark [ <i>Coluria longifolia</i> Maxim.]			OP161499	155,884	85,338	18,358	26,094	129	84 (6)	37 (7)	8 (4)	36.7	34.4	30.9	42.6	Guo et al. (2023)
<i>Geum macrophyllum</i> Willd.			MT774132	155,940	85,307	18,329	26,152	129	84 (6)	37 (7)	8 (4)	36.6	34.3	30.6	42.6	Li and Wen (2021)
<i>Geum omeiense</i> (T.C.Ku) Smedmark 1 [ <i>Coluria omeiensis</i> T.C.Ku]	LCH 1146 (NMTC)	China, Sichuan, Mount Emei	PX414103	155,388	85,538	18,140	25,855	129	84 (6)	37 (7)	8 (4)	36.6	34.3	30.8	42.6	This article

(Continued)

TABLE 1 Continued

Species name (synonym)	Voucher	Locality	GenBank accession	Size (bp)				Number of genes				GC content (%)				References
				Total	LSC	SSC	IR	Total	Protein coding	tRNA	rRNA	Total	LSC	SSC	IR	
<i>Geum omeiense</i> (T.C.Ku) Smedmark 2 [ <i>Coluria omeiensis</i> T.C.Ku]	LCH 1148 (NMTC)	China, Sichuan, Mount Emei	PX414104	155,388	85,538	18,140	25,855	129	84 (6)	37 (7)	8 (4)	36.6	34.3	30.8	42.6	This article
<i>Geum rupestre</i> (T.T.Yü & C.L.Li) Smedmark 1 [ <i>Taihangia rupestris</i> T.T.Yü & C.L.Li]			MZ151697	155,558	85,857	18,543	25,579	129	84 (6)	37 (7)	8 (4)	36.8	34.5	30.8	42.8	<a href="#">Feng et al. (2022)</a>
<i>Geum rupestre</i> (T.T.Yü & C.L.Li) Smedmark 2 [ <i>Taihangia rupestris</i> T.T.Yü & C.L.Li]			MG262388	155,479	85,771	18,550	25,579	129	84 (6)	37 (7)	8 (4)	36.8	34.5	30.8	42.8	<a href="#">Duan et al. (2018)</a>
<i>Geum rupestre</i> (T.T.Yü & C.L.Li) Smedmark 3 [ <i>Taihangia rupestris</i> T.T.Yü & C.L.Li]			ON873898	155,514	85,823	18,533	25,579	129	84 (6)	37 (7)	8 (4)	36.8	34.5	30.8	42.8	<a href="#">Liu et al. (2025)</a>
<i>Geum rupestre</i> (T.T.Yü & C.L.Li) Smedmark 4 [ <i>Taihangia rupestris</i> T.T.Yü & C.L.Li]			ON873891	155,512	85,838	18,496	25,589	129	84 (6)	37 (7)	8 (4)	36.8	34.5	30.9	42.8	<a href="#">Liu et al. (2025)</a>
<i>Geum triflorum</i> Pursh			KY419977													<a href="#">Zhang et al. (2017)</a>
<i>Geum urbanum</i> L. 1			OX327019													Unpublished
<i>Geum urbanum</i> L. 2			ON556622													Unpublished



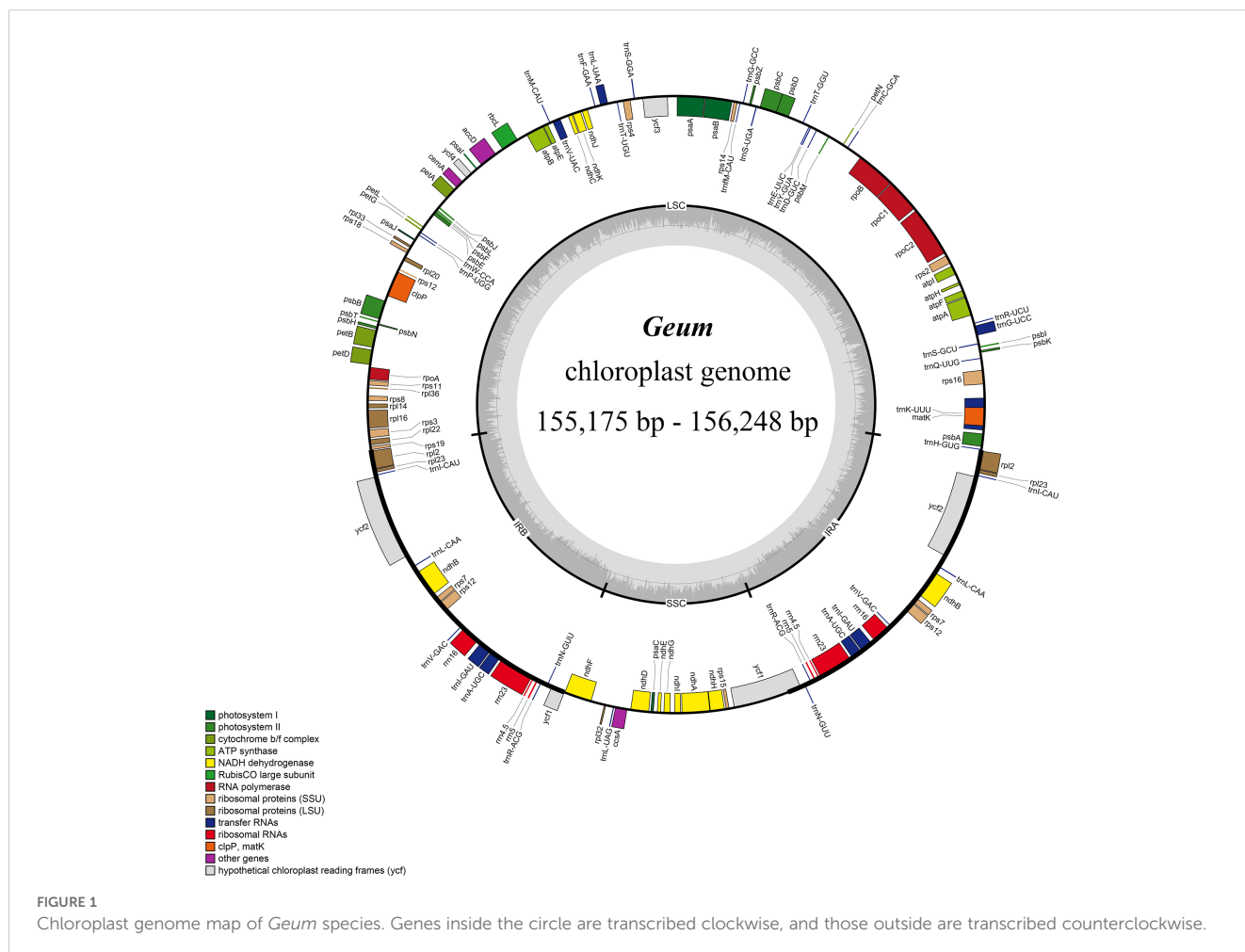
was then aligned separately by codons using MAFFT, with stop codons manually removed from each alignment. The alignments of the 78 PCGs were concatenated into a supermatrix for subsequent analysis. The FASTA format of the supermatrix was used as the input file in EasyCodeML. The ML tree established based on the supermatrix by RAxML v. 8.2.12 (Stamatakis, 2014), using the same parameters in the phylogenetic analyses, was used as an input tree (Supplementary Figure S1). The likelihood ratio test (LRT) was performed to detect adaptation signatures under four comparison site models: M0 (one-ratio) vs. M3 (discrete), M1a (neutral) vs. M2a (positive selection), M7 (beta) vs. M8 (beta and  $\omega > 1$ ), and M8a (beta and  $\omega = 1$ ) vs. M8, with significance threshold of  $p < 0.05$ . Bayesian empirical Bayes (BEB) (Yang et al., 2005) or Naïve empirical Bayes (NEB) (Nielsen and Yang, 1998) analysis was performed to detect sites under positive selection with posterior probabilities  $\geq 0.95$ .

### 3 Results and discussion

#### 3.1 Chloroplast genome characteristics

The size of the 28 *Geum* cp genomes ranged from 155,175 bp (*Geum henryi* 1) to 156,248 bp (*G. elatum* 4) (Table 1; Figure 1).

The genomes exhibited a typical quadripartite structure, as observed in most land plants (Daniell et al., 2016), comprising an LSC region of 85,307 bp (*G. macrophyllum*) to 85,857 bp (*Geum rupestre* 1), an SSC region of 18,140 bp (*Geum omeiense*) to 18,550 bp (*G. rupestre* 2), and a pair of IR regions of 25,566 bp (*G. henryi* 2) to 26,152 bp (*G. macrophyllum*). The total GC content of the 28 *Geum* cp genomes ranged from 36.6% to 36.8%, with the IR regions (42.6%–42.8%) showing higher GC content than the LSC (34.3%–34.5%) and SSC (30.7%–30.9%) regions (Table 1), likely due to the presence of ribosomal RNA (rRNA) genes (Ravi et al., 2008). All 28 *Geum* cp genomes encoded 129 genes, including 112 unique genes and 17 duplicated genes. The 112 unique genes consisted of 78 PCGs, 30 transfer RNA (tRNA) genes, and four rRNA genes (Figure 1; Table 1; Supplementary Table S2). The 17 genes duplicated in the IR regions comprised seven tRNA genes (*trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, and *trnV-GAC*), six PCGs (*ndhB*, *rpl2*, *rpl23*, *rps7*, *rps12*, and *ycf2*), and four rRNA genes (*rrn4.5*, *rrn5*, *rrn16*, and *rrn23*). Among the 17 genes containing introns, three genes (*clpP*, *rps12*, and *ycf3*) had two introns, while the remaining 14 genes (*ndhA*, *ndhB*, *petB*, *petD*, *rpl2*, *rpl16*, *rpoC1*, *rps16*, *trnA-UGC*, *trnG-UCC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, and *trnV-UAC*) each contained a single intron (Supplementary Table S2).



No gene rearrangements or inversions were detected in the 28 *Geum* cp genomes based on Mauve alignment analysis, indicating strong collinearity among these genomes. Although the *Geum* cp genomes are highly conserved in gene content, organization, and order, minor variations were visible in the IR/SC boundary regions (Figure 2). Expansion and contraction of the IR regions are the primary drivers of cp genome size variation in terrestrial plants (Ravi et al., 2008; Mower and Vickrey, 2018). All 28 *Geum* cp genomes contained identical genes and pseudogenes at the

boundary regions, including *rps19*, *rpl2*, *ψycf1*, *ndhF*, *ycf1*, and *trnH-GUG*. The *rps19-rpl2-trnH-GUG* genes were located in the LSC/IR boundary regions. In *G. henryi* and *G. omeiense*, *rps19* crossed the LSC/IRb junction (JLB), extending 2 and 32 bp in the IRb region, respectively, whereas in the other taxa, *rps19* was entirely located in the LSC region, 0–8 bp from the JLB. The duplicated *rpl2* gene was located in both IRb and IRa regions, 59–88 bp away from the JLB and IRa/LSC junction (JLA), respectively. Gene *trnH-GUG* was positioned in the LSC region, 4–79 bp from

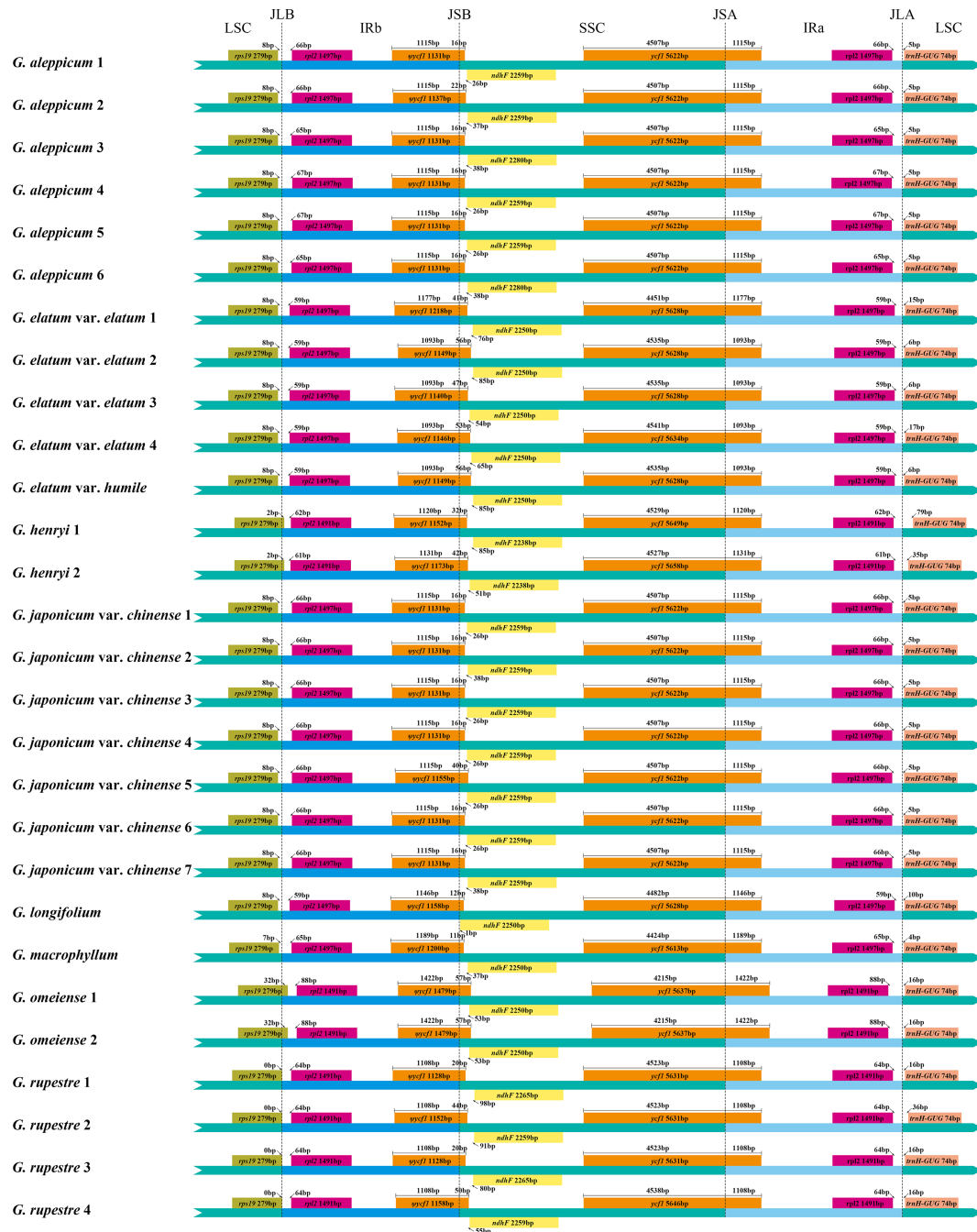


FIGURE 2

Comparison of the boundary regions of the large single-copy (LSC), small single-copy (SSC), and inverted repeat (IR) regions among 28 *Geum* chloroplast genomes.



the JLA. The pseudogene *ψycf1* and the *ndhF* gene were located around the IRb/SSC junction (JSB). The *ψycf1* pseudogene spanned the JSB, extending 11–56 bp into the SSC region, whereas *ndhF* was located in the SSC region, with 1–98 bp from the JSB region. In *Geum longifolium*, *G. japonicum* var. *chinense* 5, and *G. omeiense*, *ψycf1* overlapped with the *ndhF* gene by 4–14 bp. The *ycf1* gene crossed the SSC/IRa junction (JSA), with a length of 4,215–4,542 bp in the SSC region and 1,093–1,422 bp in the IRa region. The results indicated no significant expansion or contraction of the IR region in *Geum* cp genomes, supporting minor IR boundary shifts among closely related species (Shen et al., 2022; Li et al., 2024b; Jiang et al., 2025).

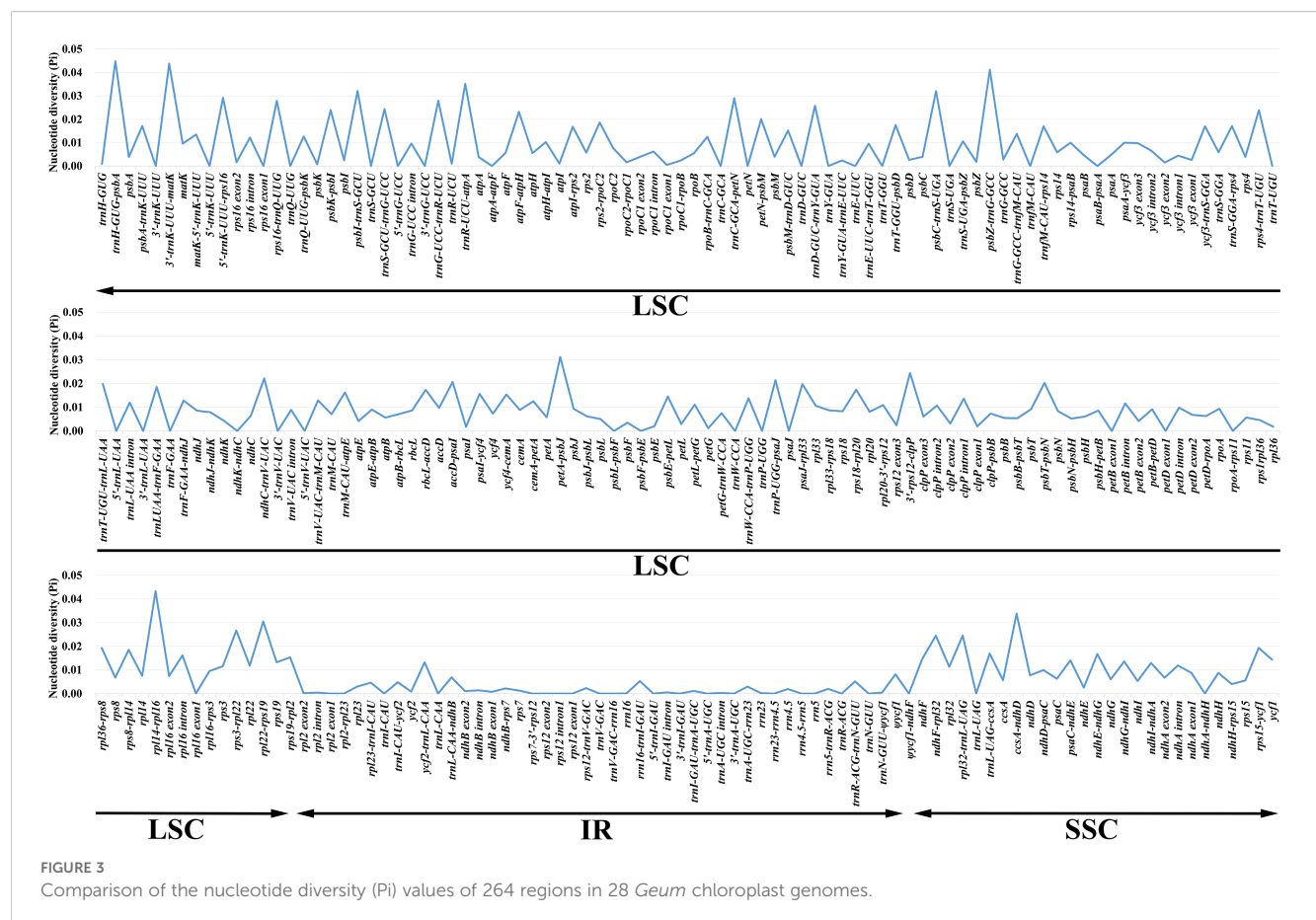
The mVISTA analysis revealed that *Geum* cp genomes were generally conserved at the genome-wide level, although several highly divergent regions were identified (Supplementary Figure S2). Overall, the LSC and SSC regions exhibited greater divergence than the IR regions. Noncoding regions, particularly intergenic spacers (IGS), were more variable than coding regions, consistent with observations in other angiosperms (e.g., Wu et al., 2020; Xu et al., 2021; Hang et al., 2025).

### 3.2 Divergence hotspots

Nucleotide variability (Pi) values for 264 regions in the 28 *Geum* cp genomes were analyzed using DnaSP v.6.12.03 (Rozas et al.,

2017). Pi values ranged from 0 to 0.04487, with an average of 0.00843, indicating high similarity among *Geum* cp genomes (Figure 3; Supplementary Table S3). Four regions had Pi > 0.04, six regions had 0.03 < Pi ≤ 0.04, 18 regions had 0.02 < Pi ≤ 0.03, 53 regions had 0.01 < Pi ≤ 0.02, 128 regions had 0 < Pi ≤ 0.01, and 55 regions had Pi = 0. The four highly variable regions with Pi > 0.04 (*trnH-GUG-psbA*, 3'-*trnK-UUU-matK*, *rpl14-rpl16*, and *psbZ-trnG-GCC*) were all located in the LSC region. Among the six regions with 0.03 < Pi ≤ 0.04 (*trnR-UCU-atpA*, *ccsA-ndhD*, *psbI-trnS-GCU*, *psbC-trnS-UGA*, *petA-psbJ*, and *rpl22-rps19*), *ccsA-ndhD* was in the SSC region, while the other five regions were in the LSC region. Of the 18 regions with 0.02 < Pi ≤ 0.03 (5'-*trnK-UUU-rps16*, *trnC-GCA-petN*, *trnG-UCC-trnR-UCU*, *rps16-trnQ-UUG*, *rps3-rpl22*, *trnD-GUC-trnY-GUA*, *rpl32-trnL-UAG*, *ndhF-rpl32*, 3'-*rps12-clpP*, *trnS-GCU-trnG-UCC*, *rps4-trnT-UGU*, *psbK-psbI*, *atpF-atpH*, *ndhC-trnV-UAC*, *trnP-UGG-psaJ*, *accD-psaI*, *psbT-psbN*, and *petN-psbM*), two regions (*rpl32-trnL-UAG*, *ndhF-rpl32*) were in the SSC region, and the remaining 16 regions were located in the LSC region. In general, regions located in the IR region exhibited lower Pi values compared with those in the LSC and SSC regions, indicating that the IR region is relatively more conserved. Moreover, coding regions were less variable than noncoding regions, with the most highly variable regions located in the intergenic spacers.

Considering both sequence length and variability, among the 16 regions with Pi > 2% and alignment lengths > 400 bp, 11 regions (3'-



*trnK*-UUU-*matK*, *psbZ*-*trnG*-GCC, *trnR*-UCU-*atpA*, *petA*-*psbJ*, 5'-*trnK*-UUU-*rps16*, *rps16*-*trnQ*-UUG, *rpl32*-*trnL*-UAG, *ndhF*-*rpl32*, *trnS*-GCU-*trnG*-UCC, *ndhC*-*trnV*-UAC, and *petN*-*psbM*) were proposed as candidate molecular markers for *Geum*, suitable for developing specific DNA barcodes. In our study, the core DNA barcodes *trnH*-GUG-*psbA* exhibited the highest Pi value of 0.04487, whereas *matK* and *rbcL* showed relatively low Pi values of 0.00962 and 0.00869, respectively. Although *trnH*-GUG-*psbA* exhibited the highest Pi value, its relatively short length limits the number of informative sites. The cp molecular markers *trnL* intron and *trnL*-*trnF* intergenic spacer, previously used in phylogenetic studies of *Geum* species (Smedmark and Eriksson, 2002; Faghir et al., 2018; Lv et al., 2020; Protopopova et al., 2023), had Pi values of only 0.01198 and 0.01856 in our dataset, respectively. To further assess the resolving power of the 11 candidate molecular markers compared with the core DNA barcodes, NJ trees were reconstructed individually for each sequence (Supplementary Figures S3–S16). The resolution ability of these sequences was evaluated based on both the number of successfully identified species and the support values in the NJ tree. Overall, the 11 candidate molecular markers demonstrated better resolution than the core DNA barcodes *trnH*-GUG-*psbA* and *rbcL*. Among the 11 candidates, nine markers—excluding *psbZ*-*trnG*-GCC and 5'-*trnK*-UUU-*rps16*—also outperformed the core DNA barcode *matK*. In

this study, the relatively lower resolution of *psbZ*-*trnG*-GCC and 5'-*trnK*-UUU-*rps16* compared with *matK* was mainly due to their inability to correctly identify *G. henryi*. In future research, the utility of the 11 candidate molecular markers for *Geum* can be further evaluated with more detailed taxon sampling and larger numbers of population samples per species. In conclusion, the development of specific molecular markers for particular taxonomic groups is necessary, and the new candidate markers identified in this study will facilitate future research on species identification and the phylogeny of *Geum*.

### 3.3 Phylogenetic analyses

Overall, compared with previous phylogenetic studies on *Geum* using cp molecular markers (Smedmark and Eriksson, 2002; Faghir et al., 2018; Protopopova et al., 2023), our study achieved higher phylogenetic resolution based on plastid genome data. Phylogenetic trees inferred using ML and BI analyses were consistent in topology (Figure 4; Supplementary Figures S17, S18). In all analyses, outgroups were robustly separated from the tribe Colurieae (ML BS = 100%, PP = 1.00). All sampled *Geum* species formed a well-supported clade (ML BS = 100%, PP = 1.00), confirming the monophyly of the current

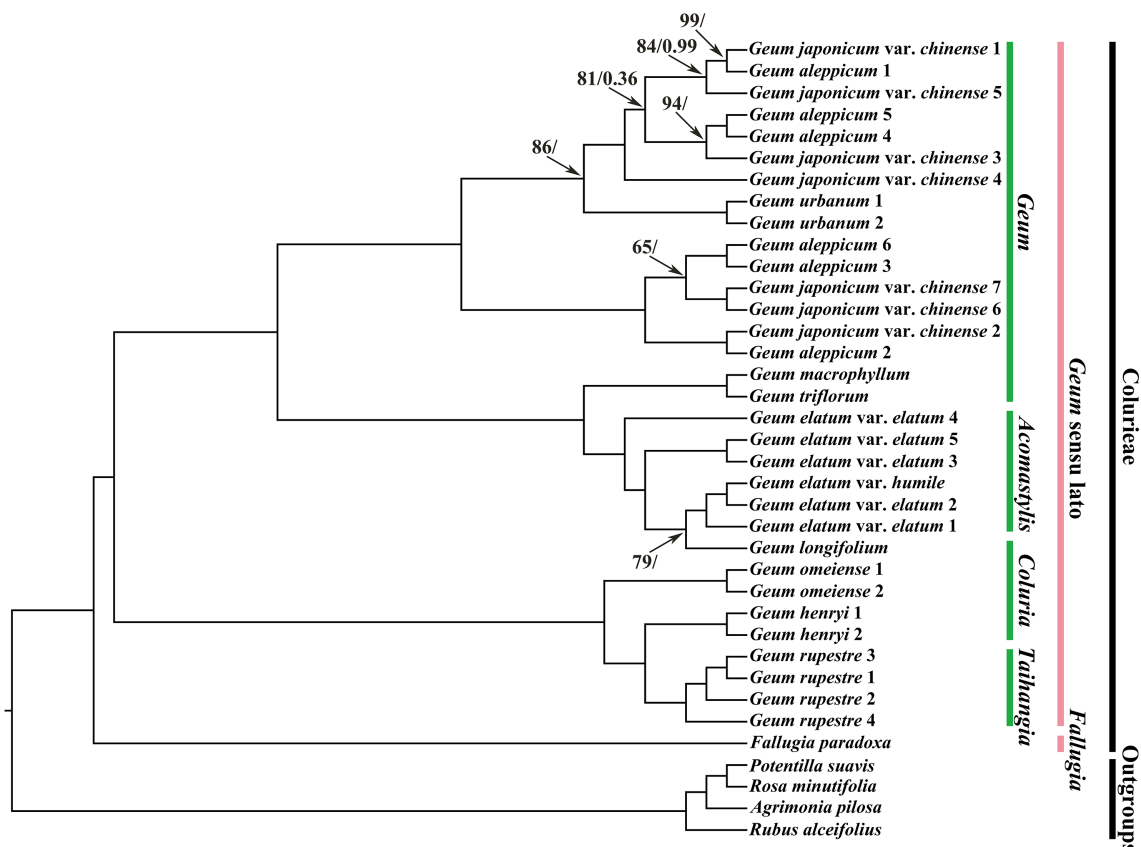


FIGURE 4

Phylogenetic trees of *Geum* and its related taxa based on 37 chloroplast genome sequences with the inverted repeat region IRa removed. *Agrimonia pilosa*, *Potentilla suavis*, *Rosa minutifolia*, and *Rubus alceifolius* were used to root the trees. Values along branches represent ML bootstrap percentages (only values < 100% are shown) and Bayesian posterior probabilities (only PP < 1.00 are shown), respectively.

delimitation of *Geum*. Under the current sampling, two major clades were recovered within *Geum*. One comprised *G. rupestre*, *G. henryi*, and *G. omeiense*, with *G. omeiense* being sister to *G. henryi* and *G. rupestre*. The other consisted of two subclades: one including *G. longifolium*, *G. elatum* var. *elatum*, and *G. elatum* var. *humile*. *Geum triflorum*, and *G. macrophyllum*; the other consisting of *G. aleppicum*, *G. japonicum* var. *chinense*, and *Geum urbanum*. Multiple samples of *G. rupestre*, *G. henryi*, and *G. omeiense* clustered into separate branches, corroborating the monophyly of the three species and supporting their treatment as distinct species. A sample of *G. longifolium* was deeply nested within multiple samples of *G. elatum*, supporting the inclusion of *G. longifolium* in *G. elatum* as a conspecific taxon from a molecular phylogenetic perspective. *Geum longifolium* appears highly similar in morphology to *G. elatum*, especially in prominent features such as yellow flowers and basal, interrupted, pinnately compound leaves, but the most significant difference is that the former has wholly deciduous styles, whereas the latter has straight, nonplumose, persistent styles (Yü and Kuan, 1985; Yü and Li, 1985; Smedmark and Eriksson, 2002; Li et al., 2003). Future detailed morphological and phylogeographic studies based on a more comprehensive sampling strategy are necessary for any further possible taxonomic treatment of *G. longifolium*. Two samples of *G. urbanum*, which clustered together, were embedded within a large clade comprising different samples of *G. aleppicum* and *G. japonicum* var. *chinense* that were intermixed. The results showed that these three *Geum* taxa are nonmonophyletic. Morphologically, they are distinguished by differences such as petal size versus sepal size, shape of the capitulum of fruitlets, and the hair case of the receptacle (Juzepchuk, 1941; Yü and Kuan, 1985; Yü and Li, 1985; Li et al., 2003; Rohrer, 2014). Geographically, *G. urbanum* is native to Europe to Central Asia and Iran, NW. Africa; *G. aleppicum* is native to the temperate Northern Hemisphere to Mexico; and *G. japonicum* var. *chinense* is native to China, according to Plants of the World Online (<http://powo.science.kew.org>). The complicated evolutionary history and relationships among these three taxa still need to be clarified through expanded taxon sampling and the use of single-copy nuclear genes in the future. *Geum elatum*, which was previously placed in *Acomastylis*, was nested within the species of *Geum*. Consistent with previous studies (Smedmark and Eriksson, 2002; Smedmark, 2006), our results supported the inclusion of *Acomastylis* species within *Geum*. *Geum longifolium*, *G. henryi*, and *G. omeiense*, which were once placed in *Coluria*, did not cluster together in our phylogenetic tree. *Geum longifolium* was closer to *G. elatum* than to *G. henryi* and *G. omeiense*, whereas *G. henryi* and *G. omeiense* showed a close affinity with *G. rupestre*, which was formerly treated as a member of the genus *Taihangia*.

Consistent with the studies of Smedmark and Eriksson (2002) and Faghir et al. (2018), our analyses indicated that *Coluria* is not a monophyletic group. The phylogenetic results presented here support treating *Coluria* as part of *Geum*. The genus *Taihangia* was nested within *Geum* species in Smedmark and Eriksson (2002). In conclusion, the results presented here supported Smedmark's recircumscription of *Geum* in a broad sense and corroborate the inclusion of *Acomastylis*, *Coluria*, and *Taihangia* within *Geum*.

### 3.4 Adaptive evolution

In cases where the *p*-values of LRTs were below the threshold of 0.05 for the four compared models—M0 vs. M3, M1a vs. M2a, M7 vs. M8, and M8a vs. M8—the NEB analysis (Nielsen and Yang, 1998) was used to identify sites under positive selection with posterior probabilities  $\geq 0.95$  in model M3, while the BEB analysis (Yang et al., 2005) was used to identify sites under positive selection with posterior probabilities  $\geq 0.95$  in models M2a and M8, according to the PAML manual (Yang, 2007) (Table 2; Supplementary Table S4). The results showed that 48 genes had positive selection sites under model M3, none under model M2a, and 23 under model M8. The PAML manual (Yang, 2007) stated that the compared model M1a vs. M2a is more stringent than M7 vs. M8, which is corroborated by the results of our study. In addition, the PAML manual (Yang, 2007) suggested that the compared model M0 vs. M3 should be utilized to test for variable  $\omega$  among sites, rather than as a test of positive selection. Therefore, we relied on the results from M7 vs. M8 and M8a vs. M8 to identify positively selected sites in *Geum* cp genomes. The BEB analysis based on the M8 model detected 90 positive selection sites across a total of 23 genes (Table 2). The number of the positive selection sites among these genes ranged from one to 39: *ycf1* with 39 sites; *ndhF* with 14 sites; three genes (*matK*, *rbcL*, and *accD*) with four sites; two genes (*rpl22* and *ndhD*) with three sites; three genes (*atpA*, *rpoC2*, and *psaA*) with two sites; and 13 genes (*atpF*, *rpoC1*, *rps4*, *cemA*, *psbJ*, *clpP*, *rps3*, *ycf2*, *ccsA*, *ndhI*, *ndhA*, *ndhH*, and *rps15*) with one site. These 23 PCGs with positively selected sites included three small subunit ribosomal genes (*rps3*, *rps4*, *rps15*); one large subunit ribosomal gene (*rpl22*); two DNA-dependent RNA polymerase (plastid-encoded bacterial-type RNA polymerase [PEP]) subunit genes (*rpoC1*, *rpoC2*); two ATP synthase subunit genes (*atpA*, *atpF*); two photosystem I (PSI) complex genes (*psaA*, *psaB*); a photosystem II (PSII) core complex gene (*psbJ*); five NAD (P)H dehydrogenase genes subunit genes (*ndhA*, *ndhD*, *ndhF*, *ndhI*, *ndhH*); the ribulose-1,5-bisphosphate carboxylase (Rubisco) large subunit gene (*rbcL*); the acetyl-CoA-carboxylase subunit gene (*accD*); the c-type cytochrome synthesis gene (*ccsA*); the cp membrane protein gene (*cemA*); the ATP-dependent Clp protease proteolytic subunit gene (*clpP*); the maturase gene (*matK*); as well as *ycf1* and *ycf2*.

*Geum* species occur in various habitats such as hillside grasslands, moist meadows, swamps, riverine scrub, rocky slopes, moist woods, rocky cliffs and ledges, alpine meadows, and arctic tundra, ranging from low to high altitudes (0–5,400 m) and often at high elevations (Juzepchuk, 1941; Smedmark and Eriksson, 2002; Li et al., 2003; Rohrer, 2014). The adaptive evolution of these 23 genes may contribute to the ability of *Geum* species to thrive in such diverse habitats. Genes such as *rps3*, *rps4*, *rps15*, and *rpl22* are ribosomal protein subunit genes that encode ribosomal proteins. The cp ribosomal proteins are important components of the protein synthesis machinery in all living cells, influencing plant growth and development and facilitating responses to stress conditions (Tiller and Bock, 2014; Zhang et al., 2016; Robles and Quesada, 2022).

TABLE 2 Positively selected sites (\**p* > 95%; \*\**p* > 99%) identified in the chloroplast genomes of *Geum* in comparisons of M7 vs. M8 and M8a vs. M8 under Bayes empirical Bayes (BEB) analysis.

Gene	Positively selected sites	Pr( <i>w</i> > 1)	Number of sites
<i>matk</i>	472 V/604 K/643 K/830 P	0.969*/0.977*/0.974*/0.988*	4
<i>atpA</i>	1475 A/1628 A/	0.992**/0.969*	2
<i>atpF</i>	1645 T	0.985*	1
<i>rpoC2</i>	3303 Q/3319 D	0.961*/0.987*	2
<i>rpoC1</i>	4304 L	0.967*	1
<i>psaA</i>	7395 N/7452 S	0.975*/0.999**	2
<i>rps4</i>	8316 S	0.973*	1
<i>rbcL</i>	9578 H/9739 C/9941 C/9967 I	0.989*/0.993**/0.984*/0.997**	4
<i>accD</i>	10001 I/10034 D/10055 G/10107 M	0.952*/0.970*/0.980*/0.985*	4
<i>cemA</i>	10723 R	0.955*	1
<i>psbJ</i>	11266 A	0.989*	1
<i>clpP</i>	11989 Y	0.956*	1
<i>rps3</i>	14126 D	0.951*	1
<i>rpl22</i>	14398 L/14468 I/14473 G	0.954*/0.985*/0.998**	3
<i>ycf2</i>	15897 W	0.971*	1
<i>ndhF</i>	18372 V/18416 R/18418 S/18476 F/18493 Q/18517 I/18526 I/18544 S/18557 L/18574 L/18597 L/18664 C/18667 Y/18671–	0.980*/0.958*/0.983*/0.992**/0.967*/0.969*/0.969*/0.969*/0.966*/0.997**/0.964*/0.998**/1.000**/0.994**	14
<i>ccsA</i>	19009 F	0.979*	1
<i>ndhD</i>	19092 L/19500 K/19536 S	0.998**/0.987*/0.959*	3
<i>ndhI</i>	20069 F	0.978*	1
<i>ndhA</i>	20087 S	0.951*	1
<i>ndhH</i>	20743 K	0.963*	1
<i>rps15</i>	20924 L	0.961*	1
<i>ycf1</i>	21375 T/21392 K/21397 K/21420 A/21464 R/21477 I/21507 R/21534 S/21537 –/21562 L/21610 A/21614 P/21616 E/21649 I/21705 K/21706 K/21707 S/21709 L/21725 W/21730 K/21945 L/22009 V/22058 E/22106 I/22173 D/22180 K/22206 I/22242 Q/22288 R/22293 K/22322 M/22465 I/22469 D/22488 R/22544 L/22701 L/22742 N/22771 N/22878 N	0.995**/0.969*/0.965*/0.981*/0.968*/0.982*/0.968*/0.975*/0.986*/0.991**/0.970*/0.970*/0.983*/0.965*/0.999**/0.993**/0.993**/0.987*/0.997**/0.962*/0.999**/0.964*/0.968*/0.988*/0.981*/0.955*/0.991**/0.985*/0.984*/0.987*/0.975*/0.958*/0.984*/0.997**/0.981*/0.961*/0.967*/0.981*/0.980*	39

Amino acids refer to the sequence of *G. macrophyllum*.

The plastid RNA polymerase subunits β' and β", encoded by genes *rpoC1* and *rpoC2*, respectively, are two of the four enzymatic subunits that constitute the catalytic core of the PEP (Zhelyazkova et al., 2012; Pfalz et al., 2015). The *atpA* and *atpF* genes encode two of the six ATP synthase subunits encoded by the plastome, and the cp ATP synthase generates the ATP needed for plant growth and photosynthesis (Wicke et al., 2011; Yamamoto et al., 2023). The *psaA* and *psaB* genes encode two major subunits of PSI, which bind to the iron–sulfur reaction center that mediates the majority of the electron transfer events (Nelson and Yocum, 2006; Wicke et al., 2011). The subunit PsbJ, encoded by the gene *psbJ*, is essential for the stable formation of PSII–light-harvesting complex (LHCII) supercomplexes, thereby enabling the higher-order organization of PSII complexes (Suorsa et al., 2004). The *ndh*

genes, including *ndhA*, *ndhD*, *ndhF*, *ndhI*, and *ndhH*, encode subunits of the Ndh-1 complex, which plays a significant role in plant adaptation to environmental stress (Endo et al., 1999; Rumeau et al., 2007; Yamori et al., 2011). The gene *rbcL* encodes the large subunit of Rubisco (Wicke et al., 2011), and Rubisco mediates the fixation of inorganic carbon from CO<sub>2</sub> into usable sugars during photosynthesis (Wilson and Hayer-Hartl, 2018; Whitney and Sharwood, 2021). The gene *cemA* encodes cp envelope membrane protein A, which is localized in the inner cp envelope membrane and mediates CO<sub>2</sub> uptake (Sasaki et al., 1993; Katoh et al., 1996; Rolland et al., 1997). The gene *clpP* in the cp encodes one of the proteolytic subunits of the ATP-dependent Clp protease. Clp protease is involved in the degradation of polypeptides and is important for cp function, plant development, and stress



acclimation (Clarke, 1999; Adam and Clarke, 2002; Kuroda and Maliga, 2003; Clarke et al., 2005). The gene *accD* encodes the beta-carboxyl transferase subunit of acetyl-CoA carboxylase (ACCase) (Wakasugi et al., 2001). ACCase in plastids is the regulatory enzyme of *de novo* fatty acid synthesis, which is crucial for leaf and seed development, storage metabolism, and cp division (Rawsthorne, 2002; Kode et al., 2005; Caroca et al., 2021). Cytochrome *c* biosynthesis protein, encoded by the *ccsA* gene, is essential for *c*-type cytochrome biosynthesis at the step of heme attachment (Xie and Merchant, 1996). Maturase K, encoded by the *matK* gene, is involved in the posttranscriptional processing of chloroplasts and is related to plant development and photosynthesis (Barthet and Hilu, 2007). The *ycf1* and *ycf2* genes are essential genes in the cp genomes of higher plants and encode products necessary for cell survival (Drescher et al., 2000). The origin of *Geum* was dated to the Miocene, 17 million years before the present (MYBP), with a 95% confidence interval from 10 to 26 MYBP (Smedmark et al., 2003; Smedmark, 2006). In conclusion, these 23 genes with sites under positive selection in the *Geum* cp genomes are associated with biological processes such as photosynthesis, biosynthesis, and self-replication, which may be key factors enabling the adaptation of *Geum* species to their habitats over evolutionary history.

## 4 Conclusion

In summary, comparative analyses showed that the 28 *Geum* cp genomes were conserved in structure, size, GC content, gene order, and gene content. Eleven regions (3'-*trnK*-UUU-*matK*, *psbZ*-*trnG*-GCC, *trnR*-UCU-*atpA*, *petA*-*psbJ*, 5'-*trnK*-UUU-*rps16*, *rps16*-*trnQ*-UUG, *rpl32*-*trnL*-UAG, *ndhF*-*rpl32*, *trnS*-GCU-*trnG*-UCC, *ndhC*-*trnV*-UAC, and *petN*-*psbM*) may serve as candidate DNA molecular markers for future studies on population genetics and systematic evolution of *Geum* species. Our phylogenetic analyses provided new insights into the relationships among *Geum* species, supported Smedmark's broad recircumscription of *Geum*, and corroborated the inclusion of *Acomastylis*, *Coluria*, and *Taihangia* within the genus. A total of 23 genes with positively selected sites were identified, suggesting that adaptive evolution of these genes may play important roles in the adaptation of *Geum* species to their habitats. Overall, this study offers valuable insights into cp genome characteristics, phylogeny, and adaptive evolution in *Geum*. Broader taxon sampling at a global scale and incorporation of single-copy nuclear genes will further clarify the phylogenetic relationships and evolutionary history of this group.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

## Author contributions

W-TF: Formal analysis, Investigation, Writing – original draft. Z-PZ: Formal analysis, Investigation, Writing – original draft. J-JG: Investigation, Writing – original draft. JW: Conceptualization, Writing – review & editing. Q-QL: Conceptualization, Investigation, Project administration, Writing – original draft.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the National Natural Science Foundation of China (No. 32260053, 31460051), the Natural Science Foundation of Inner Mongolia, China (No. 2022LHQN03004), and the Fundamental Research Funds for Inner Mongolia Normal University (No. 2022JBBJ012).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Correction note

This article has been corrected with minor changes. These changes do not impact the scientific content of the article.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2025.1713809/full#supplementary-material>

## References

- Adam, Z., and Clarke, A. K. (2002). Cutting edge of chloroplast proteolysis. *Trends Plant Sci.* 7, 451–456. doi: 10.1016/s1360-1385(02)02326-9
- Barthet, M. M., and Hilu, K. W. (2007). Expression of *matK*: functional and evolutionary implications. *Am. J. Bot.* 94, 1402–1412. doi: 10.3732/ajb.94.8.1402
- Birnesser, H., and Stolt, P. (2007). The homeopathic antiarthritic preparation Zeel comp. N: a review of molecular and clinical data. *Explore* 3, 16–22. doi: 10.1016/j.explore.2006.10.002
- Blaschek, W., Eble, S., Hilgenfeldt, U., Holzgrabe, U., Reichling, J., and Schulz, V. (2018). Hagers Enzyklopädie der Arzneistoffe und Drogen (Stuttgart, Germany: Wiss. Verl.-Ges).
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bolle, F. (1933). Eine Übersicht über die Gattung *Geum* L. und die ihr nahestehenden Gattungen. *Feddes Repert.* 72, 1–119.
- Brouillet, L. (2014). “Colurieae Rydb,” in *Flora of North America North of Mexico*, vol. 9. (Oxford University Press, New York and Oxford). Flora of North America Editorial Committee, 57.
- Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Caroca, R., Howell, K. A., Malinova, I., Burgos, A., Tiller, N., Pelizzier, T., et al. (2021). Knockdown of the plastid-encoded acetyl-CoA carboxylase gene uncovers functions in metabolism and development. *Plant Physiol.* 185, 1091–1110. doi: 10.1093/plphys/kiab106
- Clarke, A. K. (1999). ATP-dependent Clp proteases in photosynthetic organisms—a cut above the rest! *Ann. Bot.* 83, 593–599. doi: 10.1006/anbo.1999.0878
- Clarke, A. K., MacDonald, T. M., and Sjögren, L. L. (2005). The ATP-dependent Clp protease in chloroplasts of higher plants. *Physiol. Plantarum* 123, 406–412. doi: 10.1111/j.1399-3054.2005.00452.x
- Daniell, H., Lin, C. S., Yu, M., and Chang, W. J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17, 134. doi: 10.1186/s13059-016-1004-2
- Darling, A. C., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704
- Darling, A. E., Mau, B., and Perna, N. T. (2010). ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS One* 5, e11147. doi: 10.1371/journal.pone.0011147
- Dierckxsens, N., Mardulyn, P., and Smits, G. (2017). NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45, e18. doi: 10.1093/nar/gkw955
- Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.
- Drescher, A., Ruf, S., Calsa, T. Jr., Carrer, H., and Bock, R. (2000). The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant J.* 22, 97–104. doi: 10.1046/j.1365-313x.2000.00722.x
- Drouin, G., Daoud, H., and Xia, J. (2008). Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol. Phylogenet. Evol.* 49, 827–831. doi: 10.1016/j.ympev.2008.09.009
- Duan, N., Liu, S., and Liu, B. B. (2018). Complete chloroplast genome of *Taihangia rupestris* var. *rupestris* (Rosaceae), a rare cliff flower endemic to China. *Conserv. Genet. Resour.* 10, 809–811. doi: 10.1007/s12686-017-0936-5
- Endo, T., Shikanai, T., Takabayashi, A., Asada, K., and Sato, F. (1999). The role of chloroplastic NAD(P)H dehydrogenase in photoprotection. *FEBS Lett.* 457, 5–8. doi: 10.1016/s0014-5793(99)00989-8
- Faghir, M. B., Pourmojib, R., and Shavvan, R. S. (2018). Phylogeny and character evolution of the genus *Geum* L. (Family Rosaceae) from Iran: evidence from analyses of plastid and nuclear DNA sequences. *Taxon. Biosyst.* 10, 1–22.
- Feng, Z., Zheng, Y., Jiang, Y., Li, L., Luo, G., and Huang, L. (2022). The chloroplast genomes comparative analysis of *Taihangia rupestris* var. *rupestris* and *Taihangia rupestris* var. *ciliata*, two endangered and endemic cliff plants in Taihang Mountain of China. *S. Afr. J. Bot.* 148, 499–509. doi: 10.1016/j.sajb.2022.05.022
- Focke, W. O. (1894). “Rosaceae,” in *Die natürlichen Pflanzenfamilien*, vol. 3. Ed. A. Engler (Wilhelm Engelmann, Leipzig), 1–60.
- Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 32, W273–W279. doi: 10.1093/nar/gkh458
- Gajewski, W. (1957). A cytogenetic study on the genus *Geum*. *Monogr. Bot.* 4, 3–414. doi: 10.5586/mb.1957.001
- Gajewski, W. (1959). Evolution in the genus *Geum*. *Evolution* 13, 378–388. doi: 10.2307/2406114
- Gajewski, W. (1968). “*Geum* L.,” in *Flora Europaea*, vol. 2. Eds. T. G. Tutin, V. H. Heywood, N. A. Burges, D. M. Moore, D. H. Valentine, S. M. Walters and D. A. Webb (Cambridge University Press, Cambridge), 34–36.
- Gao, F., Chen, C., Arab, D. A., Du, Z., He, Y., and Ho, S. Y. W. (2019). EasyCodeML: a visual tool for analysis of selection using CodeML. *Ecol. Evol.* 9, 3891–3898. doi: 10.1002/ece3.5015
- Granic, S., Kłębowska, A., Kosiński, M., Piwowarski, J. P., Dudek, M. K., Kaźmierski, S., et al. (2016). Effects of *Geum urbanum* L. root extracts and its constituents on polymorphonuclear leucocytes functions. Significance in periodontal diseases. *J. Ethnopharmacol.* 188, 1–12. doi: 10.1016/j.jep.2016.04.030
- Guo, J. J., Zhang, Z. P., Khasbagan, Soyolt, and Li, Q. Q. (2023). The complete chloroplast genome of *Geum longifolium* (Maxim.) Smedmark 2006 (Rosaceae: Colurieae) and its phylogenomic implications. *Mitochondrial DNA B* 8, 1124–1127. doi: 10.1080/23802359.2023.2270212
- Hänsel, R., Keller, K., Rimpler, H., and Schneider, G. (1993). *Hagers Handbuch der Pharmazeutischen Praxis* (Berlin Heidelberg: Springer-Verlag).
- Hang, L. M., Zhang, Z. P., Zhang, X. H., Li, Y. K., and Li, Q. Q. (2025). Comparative chloroplast genome analyses of Potentilleae: insights into genome characteristics, mutational hotspots, and adaptive evolution. *Genetica* 153, 22. doi: 10.1007/s10709-025-00236-5
- Henrickson, J., and Parfitt, B. D. (2014). “*Fallugia* Endl.,” in *Flora of North America North of Mexico*, vol. 9. (Oxford University Press, New York and Oxford). Flora of North America Editorial Committee, 73–74.
- Hu, H. S., Mao, J. Y., Wang, X., Liang, Y. Z., Jiang, B., and Zhang, D. Q. (2023). Plastid phylogenomics and species discrimination in the “Chinese” clade of *Roscoea* (Zingiberaceae). *Plant Diversity* 45, 523–534. doi: 10.1016/j.pld.2023.03.012
- Hutchinson, J. (1967). *The genera of flowering plants* (Oxford: Oxford University Press).
- Jiang, H., He, S., He, J., Zuo, Y., Guan, W., Zhao, Y., et al. (2025). Plastid genomic features and phylogenetic placement in *Rosa* (Rosaceae) through comparative analysis. *BMC Plant Biol.* 25, 752. doi: 10.1186/s12870-025-06734-0
- Juzepchuk, S. V. (1941). “Rosoidae,” in *Flora of the USSR*, vol. 10. Eds. V. L. Komarov, B. K. Schischkin and S. V. Juzepczuk (Izdatel'stvo Akademii Nauk SSSR, Moskva-Leningrad), 1–508.
- Kalkman, C. (1988). The phylogeny of the Rosaceae. *Bot. J. Linn. Soc.* 98, 37–59. doi: 10.1111/j.1095-8339.1988.tb01693.x
- Katoh, A., Lee, K. S., Fukuzawa, H., Ohshima, K., and Ogawa, T. (1996). *cema* homologue essential to CO<sub>2</sub> transport in the cyanobacterium *Synechocystis* PCC6803. *P. Natl. Acad. Sci. U.S.A.* 93, 4006–4010. doi: 10.1073/pnas.93.9.4006
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 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, 1647–1649. doi: 10.1093/bioinformatics/bts199
- Kode, V., Mudd, E. A., Iamtham, S., and Day, A. (2005). The tobacco plastid *accD* gene is essential and is required for leaf development. *Plant J.* 44, 237–244. doi: 10.1111/j.1365-313X.2005.02533.x
- Kumar, S., Stecher, G., Suleski, M., Sanderford, M., Sharma, S., and Tamura, K. (2024). MEGA12: molecular evolutionary genetic analysis version 12 for adaptive and green computing. *Mol. Biol. Evol.* 41, msae263. doi: 10.1093/molbev/msae263
- Kuroda, H., and Maliga, P. (2003). The plastid *clpP1* protease gene is essential for plant development. *Nature* 425, 86–89. doi: 10.1038/nature01909
- Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. (2017). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34, 772–773. doi: 10.1093/molbev/msw260
- Li, C. L., Ikeda, H., and Ohba, H. (2003). “*Geum* L., *Acomastylis* Greene, *Taihangia* Yü et Li, *Coluria* R.Br., *Waldsteinia* Willd.,” in *Flora of China*, vol. 9. Eds. Z. Y. Wu, P. H. Raven and D. Y. Hong. (Science Press/Missouri Botanical Garden Press, Beijing/St. Louis). 286–291.
- Li, Q. Q., and Wen, J. (2021). The complete chloroplast genome of *Geum macrophyllum* (Rosaceae: Colurieae). *Mitochondrial DNA B* 6, 297–298. doi: 10.1080/23802359.2020.1861562
- Li, Q. Q., Yu, Y., Zhang, Z. P., and Wen, J. (2021). Comparison among the chloroplast genomes of five species of *Chamaerhodos* (Rosaceae: Potentilleae): phylogenetic implications. *Nord. J. Bot.* 39, e03121. doi: 10.1111/njb.03121
- Li, Q. Q., Zhang, Z. P., Aogan, and Wen, J. (2024b). Comparative chloroplast genomes of *Argentina* species: genome evolution and phylogenomic implications. *Front. Plant Sci.* 15, 1349358. doi: 10.3389/fpls.2024.1349358
- Li, Q. Q., Zhang, Z. P., Wen, J., and Yu, Y. (2024a). Plastid phylogenomics of the tribe Potentilleae (Rosaceae). *Mol. Phylogenet. Evol.* 190, 107961. doi: 10.1016/j.ympev.2023.107961



- Liu, M., Xing, W. L., Zhang, B., Wen, M. L., Cheng, Y. Q., Liu, Y. Y., et al. (2025). Integrated genomic analysis reveals the fine-scale population genetic structure and variety differentiation of *Taihangia rupestris*, a rare cliff plant. *J. Syst. Evol.* 63, 536–550. doi: 10.1111/jse.13148
- Lv, Z. Y., Zhang, D. G., Huang, X. H., Wang, H. C., Yang, J. Y., Tojibaev, K., et al. (2020). *Geum sunhangii* (Rosaceae), a new species from Hubei Province, China. *PhytoKeys* 156, 113–124. doi: 10.3897/phytokeys.156.37277
- Menković, N., Šavikin, K., Tasić, S., Zdunić, G., Stešević, D., Milosavljević, S., et al. (2011). Ethnobotanical study on traditional uses of wild medicinal plants in Prokletije Mountains (Montenegro). *J. Ethnopharmacol.* 133, 97–107. doi: 10.1016/j.jep.2010.09.008
- Mower, J. P., and Vickrey, T. L. (2018). Structural diversity among plastid genomes of land plants. *Adv. Bot. Res.* 85, 263–292. doi: 10.1016/bs.abr.2017.11.013
- Nelson, N., and Yocum, C. F. (2006). Structure and function of photosystems I and II. *Annu. Rev. Plant Biol.* 57, 521–565. doi: 10.1146/annurev.arplant.57.032905.105350
- Neuhaus, H. E., and Emes, M. J. (2000). Nonphotosynthetic metabolism in plastids. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 51, 111–140. doi: 10.1146/annurev.arplant.51.1.111
- Nielsen, R., and Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148, 929–936. doi: 10.1093/genetics/148.3.929
- Palmer, J. D. (1985). Comparative organization of chloroplast genomes. *Annu. Rev. Genet.* 19, 325–354. doi: 10.1146/annurev.ge.19.120185.001545
- Pfalz, J., Holtzge, U., Barkan, A., Weisheit, W., Mittag, M., and Pfannschmidt, T. (2015). ZmpTAC12 binds single-stranded nucleic acids and is essential for accumulation of the plastid-encoded polymerase complex in maize. *New Phytol.* 206, 1024–1037. doi: 10.1111/nph.13248
- Phipps, J. B. (2014). “*Waldsteinia* Willd.” in *Flora of North America North of Mexico*, vol. 9. (Oxford University Press, New York and Oxford). Flora of North America Editorial Committee, 71–72.
- Posada, D., and Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst. Biol.* 53, 793–808. doi: 10.1080/10635150490522304
- Protopopova, M., Pavlichenko, V., Chepinga, V., Gnutikov, A., and Adelshein, R. (2023). *Waldsteinia* within *Geum* s.l. (Rosaceae): main aspects of phylogeny and speciation history. *Diversity* 15, 479. doi: 10.3390/d15040479
- Rana, H. K., Rana, S. K., Sun, H., and Luo, D. (2025). Genomic signatures of habitat isolation and paleo-climate unveil the “island-like” pattern in the glasshouse plant *Rheum nobile*. *Glob. Ecol. Conserv.* 58, e03471. doi: 10.1016/j.gecco.2025.e03471
- Ravi, V., Khurana, J. P., Tyagi, A. K., and Khurana, P. (2008). An update on chloroplast genomes. *Plant Syst. Evol.* 271, 101–122. doi: 10.1007/s00606-007-0608-0
- Rawsthorne, S. (2002). Carbon flux and fatty acid synthesis in plants. *Prog. Lipid Res.* 41, 182–196. doi: 10.1016/s0163-7827(01)00023-6
- Redžić, S. S. (2007). The ecological aspect of ethnobotany and ethnopharmacology of population in Bosnia and Herzegovina. *Coll. Antropol.* 31, 869–890.
- Robles, P., and Quesada, V. (2022). Unveiling the functions of plastid ribosomal proteins in plant development and abiotic stress tolerance. *Plant Physiol. Bioch.* 189, 35–45. doi: 10.1016/j.plaphy.2022.07.029
- Rohrer, J. R. (2014). “*Sieversia* Willd., *Geum* L.” in *Flora of North America North of Mexico*, vol. 9. (Oxford University Press, New York and Oxford). Flora of North America Editorial Committee, 57–70.
- Rolland, N., Dorne, A. J., Amoroso, G., Sütemeyer, D. F., Joyard, J., and Rochaix, J. D. (1997). Disruption of the plastid *ycf10* open reading frame affects uptake of inorganic carbon in the chloroplast of *Chlamydomonas*. *EMBO J.* 16, 6713–6726. doi: 10.1093/emboj/16.22.6713
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029
- Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248
- Rumeau, D., Peltier, G., and Cornac, L. (2007). Chlororespiration and cyclic electron flow around PSI during photosynthesis and plant stress response. *Plant Cell Environ.* 30, 1041–1051. doi: 10.1111/j.1365-3040.2007.01675.x
- Rydberg, P. A. (1913). “Rosaceae,” in *North American Flora*, vol. 22. (The New York Botanical Garden, New York), 389–480.
- Sasaki, Y., Sekiguchi, K., Nagano, Y., and Matsuno, R. (1993). Chloroplast envelope protein encoded by chloroplast genome. *FEBS Lett.* 316, 93–98. doi: 10.1016/0014-5793(93)81743-j
- Sato, N., Terasawa, K., Miyajima, K., and Kabeya, Y. (2003). Organization, developmental dynamics, and evolution of plastid nucleoids. *Int. Rev. Cytol.* 232, 217–262. doi: 10.1016/s0074-7696(03)32006-6
- Scheutz, N. J. (1870). *Monographiae Georum* (Uppsala: Berling).
- Schulze-Menz, G. K. (1964). “Rosales,” in *Engler's Syllabus der Pflanzenfamilien*, vol. 2. Ed. H. Melchior (Gebrüder Borntraeger, Berlin), 193–242.
- Shen, W., Dong, Z., Zhao, W., Ma, L., Wang, F., Li, W., et al. (2022). Complete chloroplast genome sequence of *Rosa luciae* and its characteristics. *Horticulturae* 8, 788. doi: 10.3390/horticulturae8090788
- Smedmark, J. E. (2006). Recircumscription of *Geum* (Coluriae: Rosaceae). *Bot. Jahrb. Syst.* 126, 409–417. doi: 10.1127/0006-8152/2006/0126-0409
- Smedmark, J. E., and Eriksson, T. (2002). Phylogenetic relationships of *Geum* (Rosaceae) and relatives inferred from the nrITS and *trnL-trnF* regions. *Syst. Bot.* 27, 303–317. doi: 10.1043/0363-6445-27.2.303
- Smedmark, J. E., Eriksson, T., and Bremer, B. (2005). Allopolyploid evolution in Geinae (Coluriae: Rosaceae)—building reticulate species trees from bifurcating gene trees. *Org. Divers. Evol.* 5, 275–283. doi: 10.1016/j.ode.2004.12.003
- Smedmark, J. E., Eriksson, T., Evans, R. C., and Campbell, C. S. (2003). Ancient allopolyploid speciation in Geinae (Rosaceae): evidence from nuclear granule-bound starch synthase (GBSSI) gene sequences. *Syst. Biol.* 52, 374–385. doi: 10.1080/10635150390197000
- Stamatakis, A. (2014). RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Commun. Stat. Theory Methods* A7, 13–26. doi: 10.1080/03610927808827599
- Suorsa, M., Regel, R. E., Paakkari, V., Battchikova, N., Herrmann, R. G., and Aro, E. M. (2004). Protein assembly of photosystem II and accumulation of subcomplexes in the absence of low molecular mass subunits PsbL and PsbJ. *Eur. J. Biochem.* 271, 96–107. doi: 10.1046/j.1432-1033.2003.03906.x
- Tiller, N., and Bock, R. (2014). The translational apparatus of plastids and its role in plant development. *Mol. Plant* 7, 1105–1120. doi: 10.1093/mp/ssu022
- Vollmann, C., and Schultze, W. (1995). Composition of the root essential oils of several *Geum* species and related members of the subtribus Geinae (Rosaceae). *Flavour Frag. J.* 10, 173–178. doi: 10.1002/ffj.2730100309
- Wakasugi, T., Tsudzuki, T., and Sugiura, M. (2001). The genomics of land plant chloroplasts: gene content and alteration of genomic information by RNA editing. *Photosynth. Res.* 70, 107–118. doi: 10.1023/A:1013892009589
- Wang, J., Fu, C. N., Mo, Z. Q., Möller, M., Yang, J. B., Zhang, Z. R., et al. (2022). Testing the complete plastome for species discrimination, cryptic species discovery and phylogenetic resolution in *Cephalotaxus* (Cephalotaxaceae). *Front. Plant Sci.* 13, 768810. doi: 10.3389/fpls.2022.768810
- Whitney, S. M., and Sharwood, R. E. (2021). Rubisco engineering by plastid transformation and protocols for assessing expression. *Methods Mol. Biol.* 2317, 195–214. doi: 10.1007/978-1-0716-1472-3\_10
- Wicke, S., Schneeweiss, G. M., dePamphilis, C. W., Müller, K. F., and Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol. Biol.* 76, 273–297. doi: 10.1007/s11103-011-9762-4
- Wilson, R. H., and Hayer-Hartl, M. (2018). Complex chaperone dependence of rubisco biogenesis. *Biochemistry* 57, 3210–3216. doi: 10.1021/acs.biochem.8b00132
- Wu, L., Nie, L., Xu, Z., Li, P., Wang, Y., He, C., et al. (2020). Comparative and phylogenetic analysis of the complete chloroplast genomes of three *Paonia* section *Moutan* species (Paeoniaceae). *Front. Genet.* 11, 980. doi: 10.3389/fgene.2020.00980
- Xiang, Y., Huang, C. H., Hu, Y., Wen, J., Li, S., Yi, T., et al. (2017). Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Mol. Biol. Evol.* 34, 262–281. doi: 10.1093/molbev/msw242
- Xie, Z., and Merchant, S. (1996). The plastid-encoded *ccsA* gene is required for heme attachment to chloroplast c-type cytochromes. *J. Biol. Chem.* 271, 4632–4639. doi: 10.1074/jbc.271.9.4632
- Xu, J., Liu, C., Song, Y., and Li, M. (2021). Comparative analysis of the chloroplast genome for four *Pennisetum* species: molecular structure and phylogenetic relationships. *Front. Genet.* 12, 687844. doi: 10.3389/fgene.2021.687844
- Yamamoto, H., Cheuk, A., Shearman, J., Nixon, P. J., Meier, T., and Shikanai, T. (2023). Impact of engineering the ATP synthase rotor ring on photosynthesis in tobacco chloroplasts. *Plant Physiol.* 192, 1221–1233. doi: 10.1093/plphys/kiad043
- Yamori, W., Sakata, N., Suzuki, Y., Shikanai, T., and Makino, A. (2011). Cyclic electron flow around photosystem I via chloroplast NAD(P)H dehydrogenase (NDH) complex performs a significant physiological role during photosynthesis and plant growth at low temperature in rice. *Plant J.* 68, 966–976. doi: 10.1111/j.1365-3113.2011.04747.x
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Yang, Z., Wong, W. S., and Nielsen, R. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22, 1107–1118. doi: 10.1093/molbev/msi097
- Yü, T. T., and Kuan, K. C. (1985). “*Coluria* R.Br.,” in *Flora of China*, vol. 37. Ed. T. T. Yü (Science Press, Beijing), 229–232.
- Yü, T. T., and Li, C. L. (1985). “*Geum* L., *Acomastylis* Greene, *Taihangia* Yü et Li, *Waldsteinia* Willd.,” in *Flora of China*, vol. 37. Ed. T. T. Yü. (Science Press, Beijing), pp. 221–229, 232–233.

- Zhang, S. D., Jin, J. J., Chen, S. Y., Chase, M. W., Soltis, D. E., Li, H. T., et al. (2017). Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. *New Phytol.* 214, 1355–1367. doi: 10.1111/nph.14461
- Zhang, P., Wang, L., and Lu, X. (2022). Complete chloroplast genome of *Geum aleppicum* (Rosaceae). *Mitochondrial DNA B.* 7, 234–235. doi: 10.1080/23802359.2021.2024461
- Zhang, J., Yuan, H., Yang, Y., Fish, T., Lyi, S. M., Thannhauser, T. W., et al. (2016). Plastid ribosomal protein S5 is involved in photosynthesis, plant development, and cold stress tolerance in *Arabidopsis*. *J. Exp. Bot.* 67, 2731–2744. doi: 10.1093/jxb/erw106
- Zhang, G. J., Zhang, Z. P., and Li, Q. Q. (2022). Comparative analysis of chloroplast genomes of *Sanguisorba* species and insights into phylogenetic implications and molecular dating. *Nord. J. Bot.* 2022, e03719. doi: 10.1111/njb.03719
- Zhelyazkova, P., Sharma, C. M., Förstner, K. U., Liere, K., Vogel, J., and Börner, T. (2012). The primary transcriptome of barley chloroplasts: numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. *Plant Cell* 24, 123–136. doi: 10.1105/tpc.111.089441
- Zhou, N., Miao, K., Liu, C., Jia, L., Hu, J., Huang, Y., et al. (2024). Historical biogeography and evolutionary diversification of *Lilium* (Liliaceae): new insights from plastome phylogenomics. *Plant Diversity* 46, 219–228. doi: 10.1016/j.pld.2023.07.009