Holistic approach for plant species circumscription integrating standard barcodes, chloroplast genomes, single-copy nuclear genes and micro-morphological data: a case study in Epimedium (Berberidaceae)

Ya-Jing Fan, Cheng-Guo Li, Chao Ma, Min He, Jun-Peng Ma, Min-Rong Luo, Richard G.J. Hodel, Florian Jabbour, Liang Zhao, Qian Yang

BMC Genomics

RESEARCH Open Access

Holistic approach for plant species circumscription integrating standard barcodes, chloroplast genomes, single-copy nuclear genes and micro-morphological data: a case study in *Epimedium* (Berberidaceae)

Ya-jing Fan[1,2], Cheng-guo Li[1,2], Chao Ma[1,2], Min He[1,2], Jun-peng Ma[1,2], Min-rong Luo[3], Richard G.J. Hodel[4], Florian Jabbour[5], Liang Zhao[1,2*] and Qian Yang[1,2*]

Abstract

Background Accurate plant species circumscription is fundamental to biodiversity conservation, medicinal resource development, and ecological research. Yet challenges such as sample incompleteness and reliance on limited molecular markers often hinder precise species circumscription. A single identification method—molecular or morphological—is in most cases not sufficient to accurately recognize plant species. Notably, single-copy nuclear genes, despite their critical importance in resolving species circumscription through higher evolutionary rates and biparental inheritance, remain underexplored in current research. A critical next step is developing nuclear genes as DNA barcodes. In some cases, micro-morphological characteristics mirror molecular evidence and confirm species identification. *Epimedium* (Berberidaceae), well known for its medical and horticultural significance, remains poorly understood taxonomically due to its phenotypic diversity. It is an ideal taxon to explore integrative plant species circumscription combining molecular and micro-morphological data.

Results Chloroplast genome structure analysis revealed that the variations near the IR/SC boundary, the unique *trnQ-UUG* gene rearrangement, and the repeat sequences in *Epimedium* hold significant evolutionary implications. It not only uncovered the conservation and specificity of genomic structures but also provided novel insights into the phylogeny and molecular evolution of this genus. We identified eight hypervariable regions in *Epimedium* species that emerged as strong candidates for potential DNA special barcodes. These regions and the whole chloroplast genome showed higher species discriminability compared to standard barcodes. Single-copy nuclear genes were more useful in species circumscription over chloroplast genomes. Furthermore, micro-morphological characteristics served as

*Correspondence:
Liang Zhao
biology_zhaoliang@126.com
Qian Yang
yang2094@vip.163.com

Full list of author information is available at the end of the article

strong complementary evidence for species circumscription and could help distinguish species that were unresolved using only molecular or genomic analyses.

Conclusions Using *Epimedium* as a case study, we propose a Multilayer Precision Species Circumscription Approach (MPSCA), a diagnostic framework that combines standard barcodes, chloroplast genome, single-copy nuclear genes, and micro-morphological data.

Keywords Berberidaceae, Circumscription, DNA barcodes, *Epimedium*, Single copy nuclear gene

Background

Accurate species circumscription is the cornerstone of biodiversity conservation, medicinal resource development, and ecological research [1, 2]. However, in practical applications, many challenges associated with species circumscription surpass the capabilities of traditional morphological methods (e.g. *Citrus* [3],; *Begonia* [4],; *Corydalis* [5],; *Rhododendron* [6]),. For instance, in cases involving the prevention of important plant resource theft, detection of medicinal plant adulteration, or differentiation of agricultural crop varieties, reliance solely on morphological traits often hinders the accurate identification of closely related species or the analysis of highly similar specimens [7, 8]. These limitations are further exacerbated when dealing with processed plant materials, such as crushed powders, dried strips, or other treated forms, where morphological features are often obscured or entirely lost, rendering traditional identification methods ineffective [9]. Such shortcomings not only hinder the protection and sustainable management of plant resources but also jeopardize the quality control of medicinal products and the regulation of their markets [10]. Thus, the development of robust and integrative species circumscription methods remains a task that needs to be continuously explored and improved.

In response to these challenges, molecular markers have been a reliable workhorse in species circumscription, offering higher resolution and objectivity [10–13]. Early efforts in molecular systematics relied on gene segments, such as the chloroplast DNA (cpDNA) markers *rbcL*, *matK*, *trnH-psbA*, and nuclear ITS, which were widely adopted for their utility in plant species identification [14–18]. These markers, nevertheless, often lacked sufficient resolution for closely related taxa due to their limited variability [19–21]. With the emergence of next-generation sequencing (NGS) technologies, the focus shifted to analyzing entire chloroplast genomes, providing a comprehensive dataset for resolving species boundaries [21–24]. Super barcodes are whole chloroplast or mitochondrial genomes used as barcodes, offering higher resolution and greater information content compared to traditional short-fragment barcodes. However, generating fully annotated super barcodes is costly and their assembly remains challenging [23]. Consequently, gene fragments screened from the entire chloroplast genome are used as special barcodes because they provide rich phylogenetic information and can offer species-level discrimination. However, the uniparental inheritance and slow evolution of chloroplast genomes restrict their utility [25, 26].

Nuclear genes are increasingly utilized in plant circumscription [27–29]. Their biparental inheritance and recombination make nuclear genes more appropriate than organellar genes when dealing with hybridization, polyploidization, rapid speciation, incomplete lineage sorting, and horizontal gene transfer among closely related species [30, 31].

Initially, only a handful of nuclear genes were successfully used to provide phylogenetic hypotheses (*C-mos* [32],; *LFY* [33],; *ALG11* [34]),. Nuclear markers can now be obtained more easily via transcriptome sequencing (RNA−seq) and restriction site associated DNA sequencing (RAD-seq) [35–37]. Nevertheless, these two methods have limitations due to the need for fresh tissue, and uncertainty in orthology inference [38, 39].Target enrichment sequencing (Hyb-seq) emerged as a promising alternative for plant species circumscription, yet its high cost for customized probes and complex experimental procedures remain significant drawbacks [40, 41].

Deep genome skimming (DGS) technology, as an alternative to Hyb-Seq, stands out for its ability to efficiently and economically recover hundreds of single-copy nuclear genes (SCNs), along with chloroplast, mitochondrial DNA (mtDNA), and entire nuclear ribosomal DNA (nrDNA) repeat sequences, from deep genome sequencing data [42, 43]. Notably, DGS overcomes material restrictions, as it does not require fresh tissue, and provides a cost-effective approach to obtaining a wealth of nuclear gene data [42, 43]. Therefore, with the decreasing cost of sequencing, the DNA barcoding concept is now poised to extend its reach into SCNs data, enabling us to address previously intractable species circumscription [44–46].

Traditional morphological methods have long played an important role in species circumscription [47–49]. With the rapid development of molecular systematics, morphological characteristics remain critical, often to corroborate with molecular evidence for enhanced species circumscription in the molecular era [50–52].

The genus *Epimedium* (Berberidaceae), commonly known as "horny goat weed", is a group of perennial herbs comprising approximately 60 to 70 species, most of which are endemic to China, with a few species occurring in Asia or Europe [53–57]. China serves as the center of its diversity, hosting ~ 80% of the genus's species, many of which are narrowly distributed in mountainous regions such as the Hengduan Mountains, Qinling Mountains, and karst landscapes of southwestern China [57, 58]. The selected species for this study include both widespread taxa (e.g., *E. brevicornu* and *E. sagittatum*) and regionally restricted endemics (e.g., *E. hunanense* and *E. qingchengshanense*), reflecting the genus's biogeographic and ecological heterogeneity [57].

The morphological classification of *Epimedium* species presents significant challenges, primarily due to overlapping traits such as leaflet shape, flower spur length, and sepal pigmentation [56]. For instance, although *E. acuminatum* and *E. koreanum* exhibit distinct typical leaflet morphology, spur length, and sepal coloration, they prove difficult to precisely distinguish due to continuous variation of these traits in natural populations [57]. These ambiguities are further exacerbated in processed medicinal materials, where diagnostic traits are often lost. Traditional classification criteria based on floral structures are insufficient to accurately distinguish species, especially those with overlapping morphological characteristics [59].

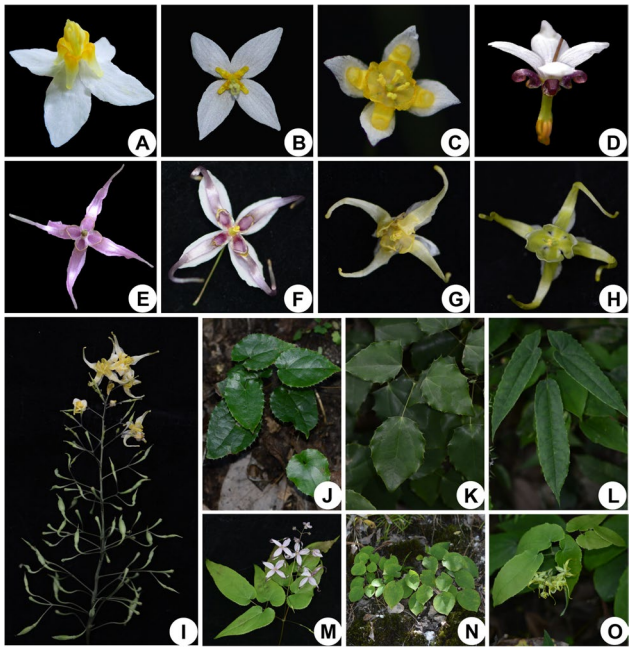

Fig. 1 Morphological diversity of *Epimedium* species. **A–H** Flowers of different forms. **A***E. brevicornu*.**B***E. pubescens*.**C***E. truncatum*.**D***E. dolichostemon*.**E***E. sutchuenense*.**F***E. acuminatum*.**G***E. chlorandrum*.**H***E. franchetii*.**I** Panicle of *E. chlorandrum*. **J–L** Pinnately compound leaves *E. leptorrhizum*, *E. truncatum*, and *E. mikinorii*.**M**, **O** Racemes of *E. acuminatum* and *E. hunanense*.**N***E. pubescens*

However, leaf epidermal characteristics, such as stomatal morphology and wax patterns, offer a promising alternative [60]. These traits are environmentally stable, accessible year-round, and preserved in herbarium specimens, even when floral structures are absent or degraded. Previous studies on *Epimedium* micro-morphology have primarily focused on floral structures, such as petal epidermal cell types and nectary anatomy [59]. Systematic investigations of leaf epidermal traits for species circumscription remain largely unexplored, highlighting a critical gap in resolving taxonomic ambiguities within this genus.

In traditional Chinese medicine, *Epimedium* leaves are prized for their bioactive compounds, yet adulteration is widespread due to morphological convergence. For example, *E. brevicornu* and *E. koreanum*—pharmacologically distinct species—are frequently confused in commercial markets [55]. Similarly, *E. wushanense* is often substituted with the morphologically similar but chemically divergent *E. pseudowushanense* [61]. Such misidentification not only undermines product efficacy but also threatens sustainable resource management.

Additionally, this genus also holds significant value in horticulture (Fig. 1) [54, 63]. The genus *Epimedium* was first described by Linnaeus in 1753 [62]. The 19th century saw an expansion in descriptions of *Epimedium* species, particularly from East Asia, though infrageneric classifications remained inconsistent due to overlapping morphological characters [63–65]. A Major milestone was achieved in the 20th century with Stearn's comprehensive revision, which provided a systematic framework and described numerous new species [56]. The integration of molecular phylogenetic methods in recent decades has further clarified relationships within the genus, resolving long-standing taxonomic ambiguities. Despite these advances, species circumscription in *Epimedium* remains challenging due to high morphological plasticity and frequent hybridization [63–69].

Here, we generate plastome sequences and single copy nuclear genes to combine with micro-morphological data in the Berberidaceae for species delimitation in this rapidly radiating group. Specifically, we address the following objectives: (1) assess the utility of complete plastomes as super barcodes; (2) assess the utility of SCNs as potential barcode candidates; (3) integrate leaf micro-morphological characteristics as diagnostic tools in *Epimedium* taxonomy, complementing molecular approaches to enhance species circumscription accuracy. We propose and validate a novel, integrated framework for plant species circumscription that combines standard barcodes, chloroplast genome, SCNs, and micro-morphological data, using *Epimedium* as an example. This method will

provide a generalizable technical framework for the rapid circumscription of other plant taxa, and promote the further development of plant taxonomy and biodiversity research. Additionally, while our primary focus is species circumscription, we concurrently analyze chloroplast genome structural features to explore their potential roles in elucidating genomic diversity and evolutionary dynamics within *Epimedium*.

Results

Characteristics of Epimedium plastomes

The 27 complete chloroplast genomes of *Epimedium* all showed a typical quadripartite structure consisting of a large single-copy region (LSC, 86882–89455 bp), a small single-copy region (SSC, 16,031 bp–17,224 bp), and two inverted-repeat regions (IR, 25,209 bp–26,236 bp), with the total length ranging from 156,720 bp (*E. leptorrhizum* ZY707) to 159,159 bp (*E. pauciflorum* ZY709) (Fig. 2; Table 1). The total Guanine-Cytosine (GC) content of each sequence varied from 38.7 to 38.9, with the highest content in the IR region, followed by the LSC and SSC regions (Table 1). The total number of unique genes annotated in Most of the sequences was 112, including 78 protein-coding genes, 4 ribosomal RNA genes and 30 transporter RNA genes (Table S1). For all species of this genus, there were 20 repeats genes in the IR region,

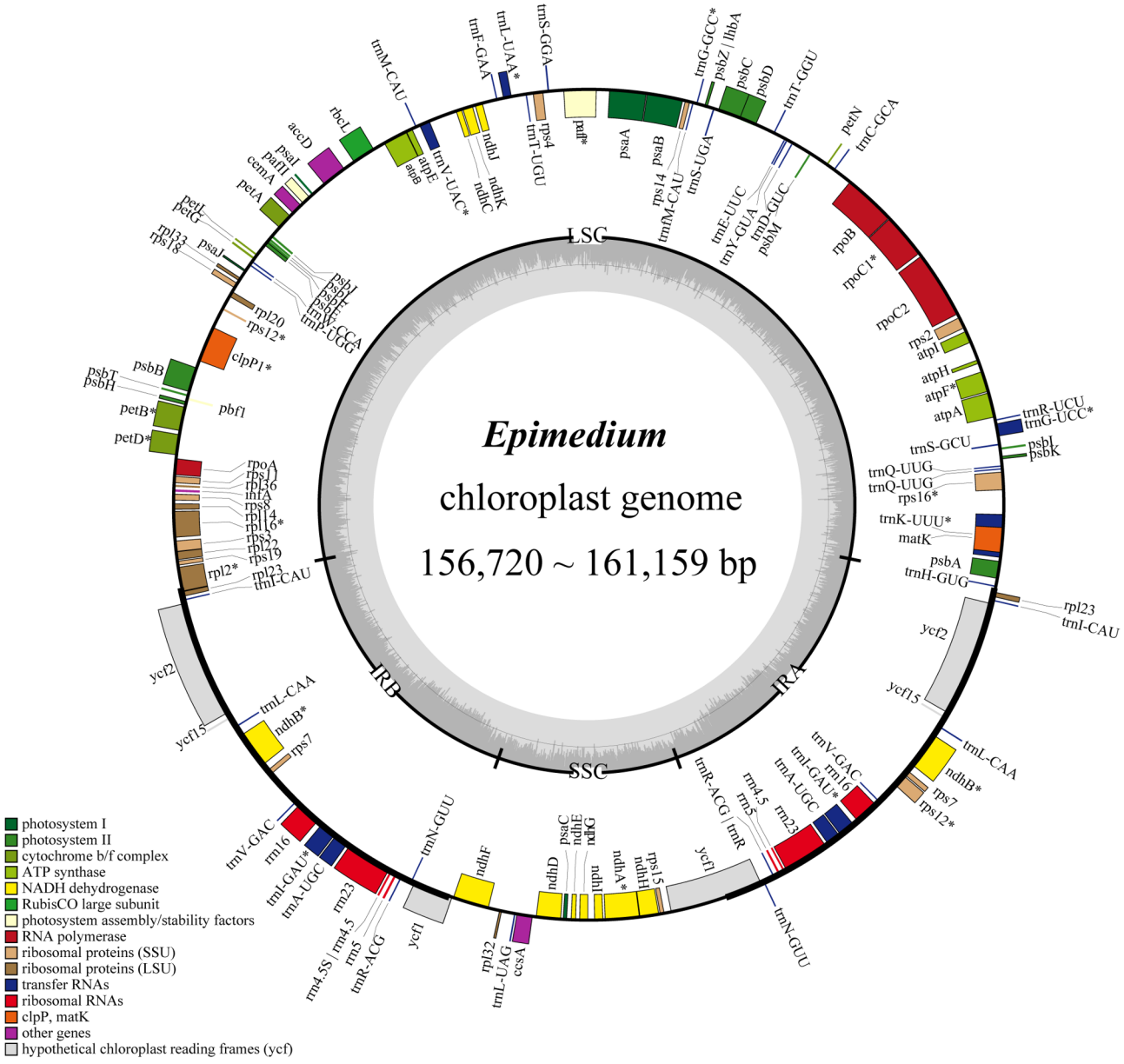

Fig. 2 Gene map of *Epimedium* chloroplast genome. The two gray arrows indicate the direction of gene transcription. The dashed area in the inner circle indicates the GC content of the plastome. LSC: large-single-copy; SSC: small-single-copy; IR: inverted repeat

Table 1 Genomic characteristics and source statistics of *Epimedium* and outgroups

Species	Voucher number	GenBank accession	SRA accession	Total size (bp)	Total GC%	LSC		SSC		IR	
						size (bp)	GC%	size (bp)	GC%	size (bp)	GC%
E. acuminatum	701	PV192894 OQ366391*	SRR31870119	156,711	38.8	89,043	37.4	17,022	32.8	25,323	43.3
E. borealiguizhouense	702	PV192895 MN873561* NC044889*	SRR31870118	156,697	38.8	89,030	37.4	17,033	32.8	25,317	43.3
E. brachyrrhizum	715	PV192904	SRR31870113	156,570	38.8	88,921	37.4	17,105	32.8	25,272	43.3
E. brevicornu	716	PV192900	SRR31870112	154,550	38.8	86,882	37.3	17,072	32.7	25,298	43.3
	717	PV192901	SRR31870111	156,647	38.8	88,979	37.4	17,072	32.7	25,298	43.3
	731	PV192902 MT560406* MN714008*	SRR31870099	156,637	38.8	88,993	37.4	17,056	32.7	25,292	43.3
E. coactum	713	PV192907 MT560395* MT560402*	SRR31870115	156,385	38.8	88,849	37.4	17,086	32.8	25,275	43.3
E. davidii	703	PV192910 MT560419*	SRR31870107	156,022	38.8	88,499	37.4	17,069	32.8	25,227	43.3
E. elongatum	733	PV192915 NC080247*	SRR31870097	156,439	38.8	88,911	37.4	17,068	32.8	25,230	43.3
E. flavum	704	PV192918 MW483093*	SRR31870096	155,961	38.9	89,007	37.4	16,506	33.0	25,224	43.3
E. franchetii	705	PV192906 NC065479*	SRR31870095	156,367	38.8	88,783	37.4	17,042	32.8	25,271	43.3
E. glandulosopilosum	706	PV192916 MT560393*	SRR31870094	156,447	38.8	88,927	37.4	17,072	32.7	25,224	43.3
E. koreanum	718	PV192919	SRR31870110	157,115	38.7	89,455	37.3	17,224	32.8	25,218	43.3
	719	PV192920 MT560411* KM207675*	SRR31870109	156,528	38.8	89,022	37.4	17,184	32.8	25,161	43.3
E. leptorrhizum	707	PV192909 NC062453*	SRR31870093	156,135	38.9	88,824	37.4	16,761	33.0	25,275	43.3
E. mikinorii	708	PV192896 NC044890*	SRR31870092	156,492	38.8	88,937	37.4	16,921	32.8	25,317	43.3
E. pauciflorum	709	PV192912 NC044890*	SRR31870091	156,585	38.8	89,045	37.4	17,092	32.7	25,224	43.3
E. platypetalum	710	PV192917 MW483078*	SRR31870090	156,523	38.8	88,983	37.4	17,068	32.8	25,236	43.3
E. pubescens	720	PV192903	SRR31870108	156,411	38.8	88,940	37.4	16,911	32.7	25,280	43.3
	721	PV192913 MT560420* NC053532* MN939633*	SRR31870106	156,550	38.8	88,980	37.4	17,026	32.8	25,272	43.3
E. qingchengshanense	714	PV192911 NC053539* MT560415*	SRR31870114	156,514	38.8	89,034	37.4	17,062	32.8	25,209	43.3
E. sagittatum	722	PV192905	SRR31870105	157,502	38.8	88,999	37.4	16,031	32.7	26,236	43.0
	723	PV192908 MN02726*	SRR31870104	156,510	38.8	88,898	37.4	17,081	32.8	25,266	43.3
E. simplicifolium	711	PV192897 NC053529*	SRR31870117	156,749	38.8	89,020	37.4	17,037	32.8	25,346	43.3
E. stellulatum	712	PV192898 MN027267*	SRR25207830*	156,697	38.8	89,048	37.4	17,031	32.8	25,309	43.3
E. wushanense	724	PV192914	SRR31870103	156,418	38.8	88,968	37.4	17,002	32.9	25,224	43.3

Table 1 (continued)

Species	Voucher number	GenBank accession	SRA accession	Total size (bp)	Total GC%	LSC		SSC		IR	
						size (bp)	GC%	size (bp)	GC%	size (bp)	GC%
	732	PV192899 MN857417*	SRR31870098	156,483	38.8	88,831	37.4	17,052	32.8	25,300	43.3
P. dubium	729	PV192922	SRR31870101								
P. peltatum	730	PV192923	SRR31870100								
V. hexandra	728	PV192921	SRR31870102								

* indicates sequences downloaded from National Center of Biotechnology Information. Newly sequenced data are listed without additional symbols. Full provenance details, including collection sites and voucher specimens, are provided in Supplementary Table S8

including 9 protein-coding genes, and 4 rRNA genes and 7 tRNA genes. *trnQ-UUG* gene was repeated in the LSC region. Thus, most of the *Epimedium* genomes have 133 genes. Due to the absence of the *rpl32* gene, the total number of unique genes annotated in *E. flavum* ZY704 was 111. A total of 17 genes contained unique introns, of which 15 contained one intron and two contained two introns.

Comparative genome analyses

Using *Epimedium acuminatum* ZY701 as the reference sequence, 21 sequences from different species were selected to visualize alignments. The results showed that the chloroplast genomes of different species in *Epimedium* were highly similar, but some species had gene insertions and deletions in some regions. Meanwhile, compared with the coding regions of CDS genes, the non-coding regions displayed more dissimilarity (Fig. S1). The whole chloroplast genome of *E. acuminatum* ZY701 was compared with that of other genera of Berberidaceae. The results indicated significant differences in sequence length, gene order, and content between *Epimedium* and its related genera (Fig. S2).

The Berberidoideae significantly exhibited amplification of the IR region with an increase of approximately 10,000 bp, which was not observed in the other subfamilies. In the Nandinoideae, the LSC-IRa boundary involved the gene *rps19*, while the IRb-SSC region boundary was located between the gene *rps19* and the gene *trnH-GUG*. However, the length of the IR region and the gene composition of its boundaries in the genus *Epimedium* differed significantly from those of the Berberidoideae and the Nandinoideae (Fig. 3, S3).

In *Epimedium*, we found that seven genes (*rpl2*, *rpl22*, *rpl23*, *ycf1*, *rpl19*, *trnI*, *trnH*) were located near these boundary regions (Fig. 3). Notably, the IR region of this genus exhibits significant length variation, with differences ranging up to approximately 1000 bp, and we found obvious contraction and expansion phenomena. Specifically, the LSC-IRa boundary of 12 species is positioned within *rpl2*, while seven species exhibited this boundary within *rpl22*, and in two species it was located in *rpl23*. IR-SSC boundaries were located within *ycf1*. The LSC-IRb boundaries were located between *rpl23* and *trnH* for 13 species, between *rpl19* and *trnH* for six species, and between *rpl2* and *trnH* for two other species. Therefore, the IR region of seven species, including *E. davidii*, *E. elongatum*, *E. flavum*, *E. pauciflorum*, *E. platypetalum*, *E. qingchengshanense*, and *E. wushanense*, show a significant expansion. On the contrary, *E. koreanum* and *E. leptorrhizum* showed significant IR contraction (Fig. 3).

The comparative analysis revealed significant variations in SSR counts among species, with *E. simplicifolium* and *E. mikinorii* exhibiting the lowest counts at 80, and *E. franchetii* the highest at 94 (Fig. 4A). Mononucleotides were the Most abundant SSR type, accounting for 78.9%, followed by dinucleotides, tetranucleotides, and trinucleotides. Pentanucleotides and hexanucleotides were less frequent, constituting 0.9% and 0.2%, respectively (Fig. 4A). Microsatellites (SSRs) with lengths between 10 bp and 13 bp Made up approximately 85% of the total, primarily located in the LSC region (85%), with the SSC region containing 12%, and the IR region the least at 3% (Fig. 4B and C).

In the analysis of 21 chloroplast genomes, a total of 706 tandem repeats were identified, with counts ranging from 30 (*E. coactum*) to 36 (*E. flavum*, *E. pauciflorum*, *E. simplicifolium*) across species (Fig. 4D). These tandem repeats were predominantly found in the IR region (64%), followed by the LSC region (32%), while only 1–2 were observed in the SSC region (Fig. 4D). Concerning large repeats, a total of 1964 were detected across the 21 chloroplast genomes, comprising 1021 forward repeats, 932 palindromic repeats, 5 complement repeats, and 6 reverse repeats (Fig. 4D). The distribution of these repeats varied among species, with forward and palindromic repeats displaying a range of counts, while complementary and reverse repeats were nearly absent. Most large repeats were 30–39 bp in length (57.4%), with longer repeats becoming less frequent. These large repeats were predominantly located in the LSC and IR regions, with very few observed in the SSC region (Fig. 4E).

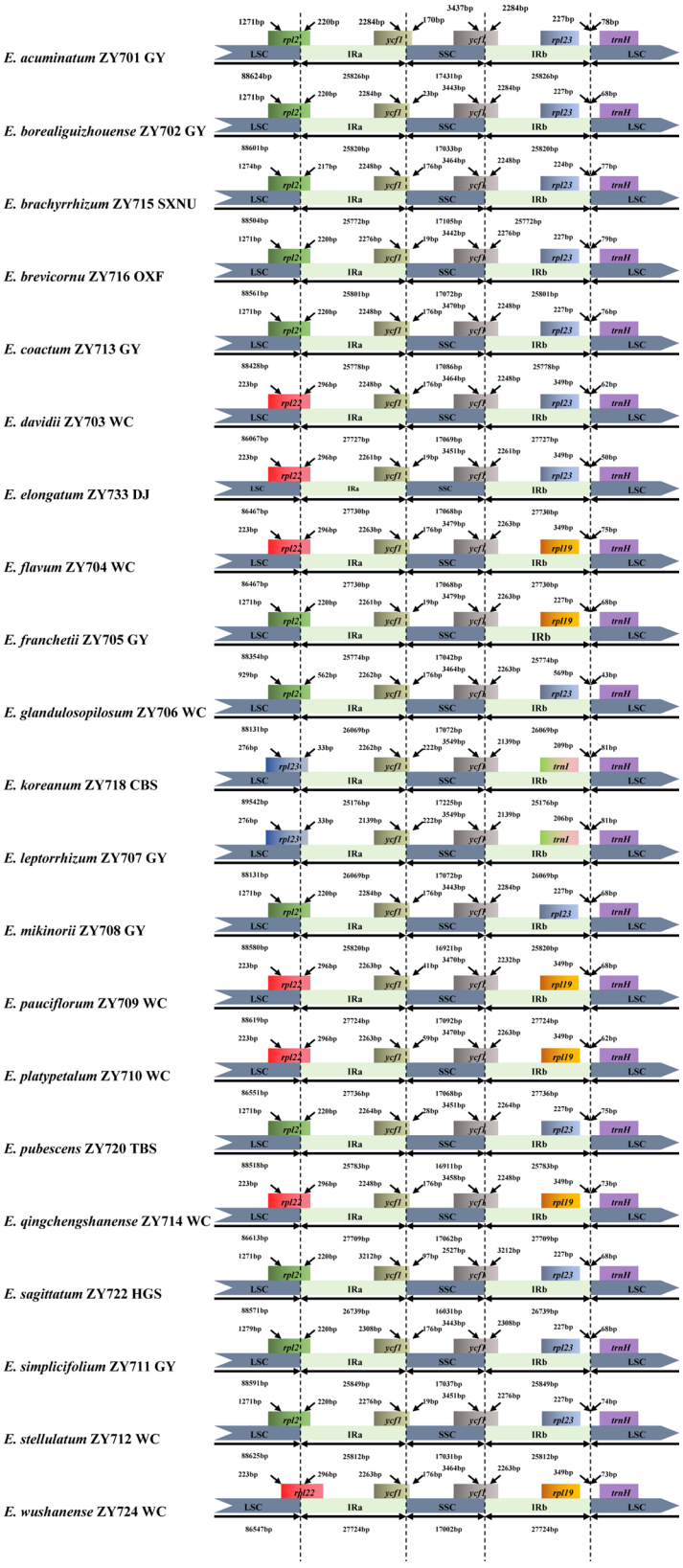

Fig. 3 Comparison of the LSC, IR and SSC boundary regions of the plastomes of 21 species of *Epimedium*

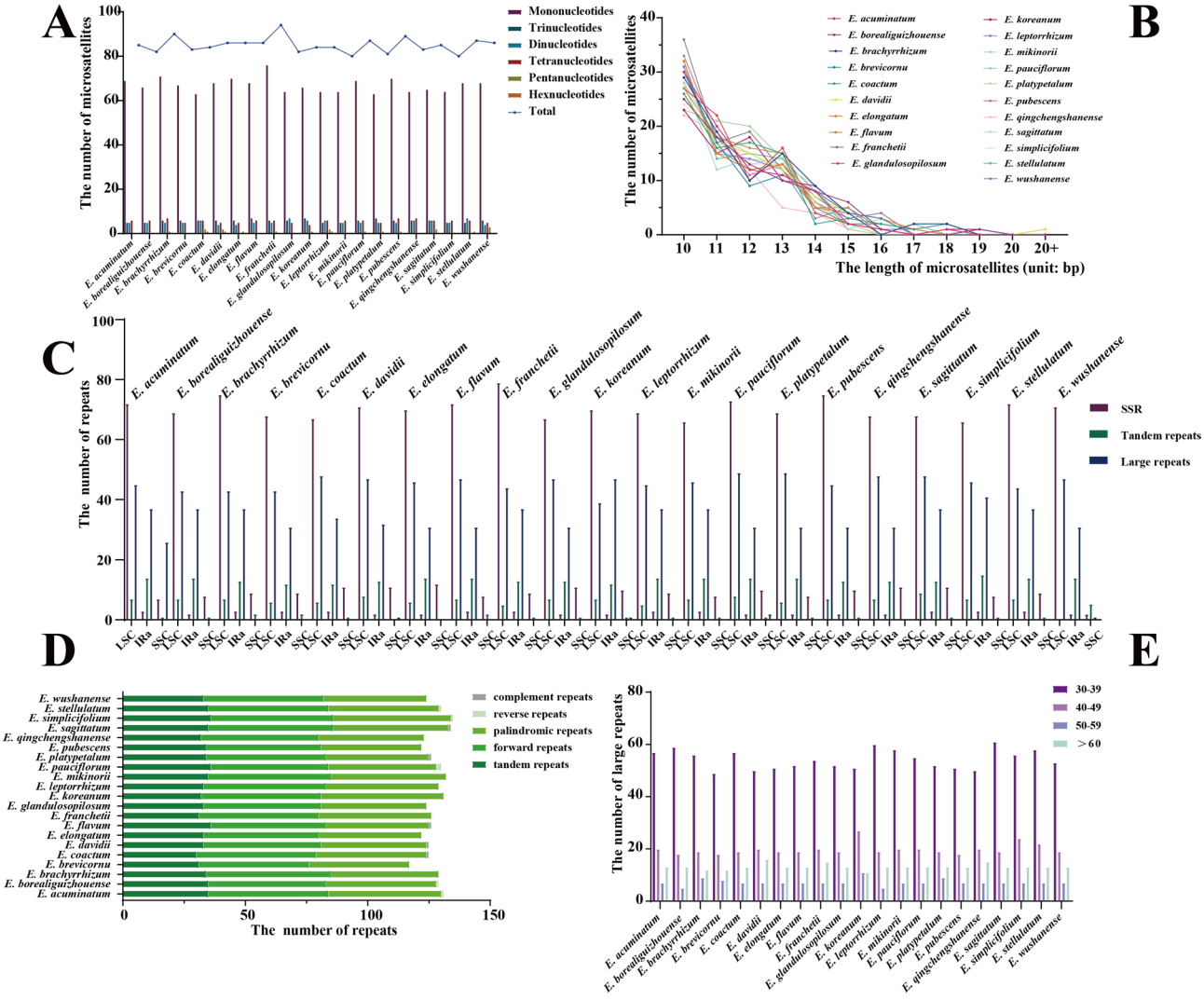

Fig. 4 Analyses of SSRs and repeated sequences in the plastomes of 21 species of *Epimedium*. **A** Frequency of microsatellites by the length of repeated units. **B** Frequency of microsatellites by length (unit: bp). **C** Frequency of all repeats by location. **D** Numbers of five different types of repeats. **E** Frequency of four types of dispersed repeats by length (unit: bp)

Hypervariable regions in the chloroplast genomes

The analysis revealed Pi values ranging from 0 to 0.08940, with an average of 0.00194 (Fig. 5). We identified eight regions where Pi values exceeded the 0.008 threshold, with the *ycf1-ndhF* region exhibiting the highest diversity at 0.04223. Notably, these high-Pi-value sites were situated in the spacer regions. Specifically, we identified eight regions of high variability with Pi values ranging from 0.00239 to 0.08940. These regions include *psbA-trnH*, *psbC-psbZ*, *psaA-pafI*, *accD-pafII*, *ycf1-ndhF*, *rpl32-trnL*, *ndhD-psaC*, and *rps15-ycf1* (Table 2). The distribution of these highly variable regions exhibits distinct characteristics: four (*psbA-trnH*, *psbC-psbZ*, *psaA-pafI*, *accD-pafII*) are clustered in the large single-copy region (LSC), while the other four (*ycf1-ndhF*, *rpl32-trnL*, *ndhD-psaC*, *rps15-ycf1*) are located in the small single-copy region (SSC). No regions of high variability were found within the inverted repeat region (IR).

Standard barcodes and chloroplast genomes in species discrimination

The variable site proportions and total site were 304/1578 (19.3%) for *matK*, 114/1428 (8.0%) for *rbcL*, 2/674 (0.3%) for *trnH-psbA*, 649/7483 (8.7%) for the tandem hypervariable region and 39,140/179,620 (24.8%) for the whole chloroplast genome (Table 3). Using the maximum likelihood (ML) method, the *rbcL* gene exhibited the lowest identification efficiency (9.5%), successfully distinguishing only *Epimedium acuminatum* and *E. koreanum* (Fig. S4A). The *matK* gene showed higher efficiency (14.3%), identifying *E. borealiguizhouense*, *E. flavum*, and *E. koreanum* (Fig. S5A). The *trnH-psbA* spacer achieved the highest single-locus efficiency (19.0%),

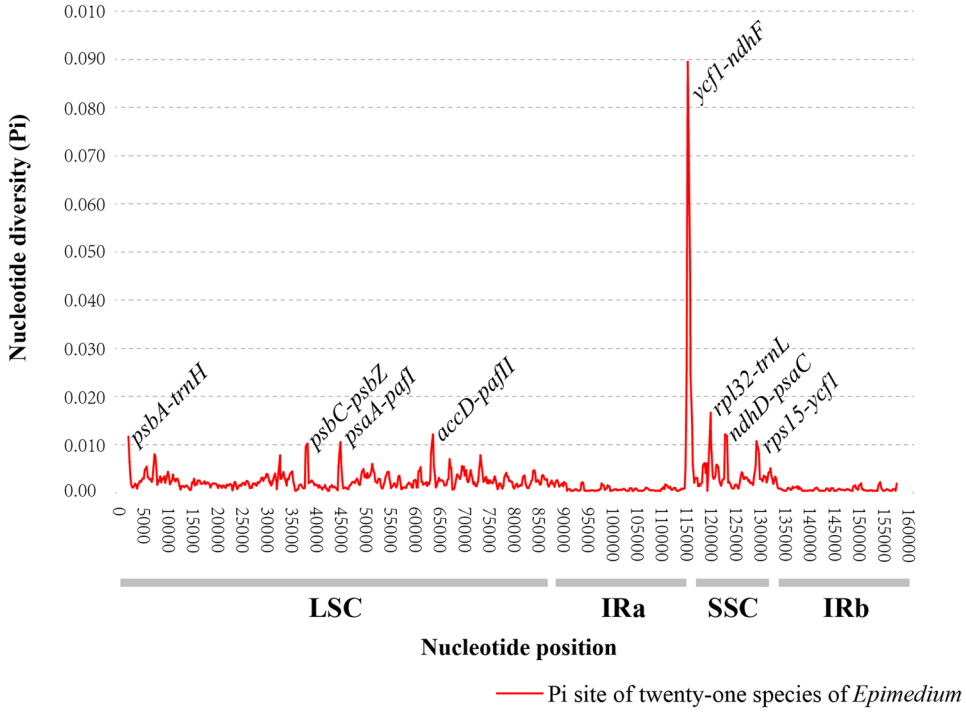

Fig. 5 Sliding window analysis of the 27 plastomes of samples of *Epimedium*. The horizontal axis indicates the location of the nucleotide, and the vertical axis indicates the Pi value

Table 2 Sequence characterization of eight hypervariable regions in 27 chloroplast genomes of *Epimedium*

Region	Aligned length	Variable sites		Indels		Nucleotide diversity (Pi)
		No.	%	No.	Average length	
trnH-psbA	600	32	5.33	18	5.600	0.01131
psbC-psbZ	953	21	2.20	7	5.571	0.00950
psaA-pafI	789	36	4.56	8	4.250	0.00788
accD-pafII	1210	43	3.56	15	4.800	0.00795
ycf1-ndhF	1010	35	3.47	8	120.000	0.04224
rpl32-trnL	987	21	2.13	8	5.750	0.01137
ndhD-psaC	1001	32	3.20	6	3.500	0.01161
rps15-ycf1	466	22	4.72	8	31.750	0.00810

Table 3 Sequence characterization of chloroplast genetic markers in 27 Chloroplast genomes of *Epimedium*

Marker	Length	Variable sites	Pairwise Identity%	Identical sites%	GC%
matK	1578	304	98.7	76.9	33.9
rbcL	1428	114	99.5	92.0	44.4
trnH-psbA	674	2	92.0	16.2	28.4
hypervariable regions	8483	649	91.8	30.0	32.9
chloroplast genome	179,620	19,605	97.5	66.0	38.8

discriminating *E. borealiguizhouense*, *E. flavum*, *E. koreanum*, and *E. qingchengshanense* (Fig. S5C). The concatenated *matK* + *rbcL* + *trnH-psbA* dataset identified the same four species as the *trnH-psbA* spacer (Fig. S5C). The tandem hypervariable region dataset distinguished six species (28.6%) (Fig. 6A). Similarly, the complete chloroplast genome Successfully distinguished six of 21 species (28.6%) (Fig. 6C). Under Bayesian inference (BI), both *rbcL* and *matK* demonstrated identical identification efficiency (14.3%): *rbcL* distinguished *E. acuminatum*, *E. flavum*, and *E. koreanum*, while *matK* identified *E. borealiguizhouense*, *E. flavum*, and *E. koreanum* (Fig. S4B, S5B). The *trnH-psbA* spacer again showed the highest efficiency (19.0%), resolving the same four species as under ML (*E. borealiguizhouense*, *E. flavum*, *E. koreanum*, and *E. qingchengshanense*) (Fig. S4D). The concatenated combination replicated this four-species identification outcome (Fig. S5D). The tandem highly variable region dataset again distinguished six species (28.6%), and the whole chloroplast genomes distinguished six species (28.6%) (Fig. 6B, D).

Compared to single-segment barcodes, the combination of two or three loci provided limited improvement in

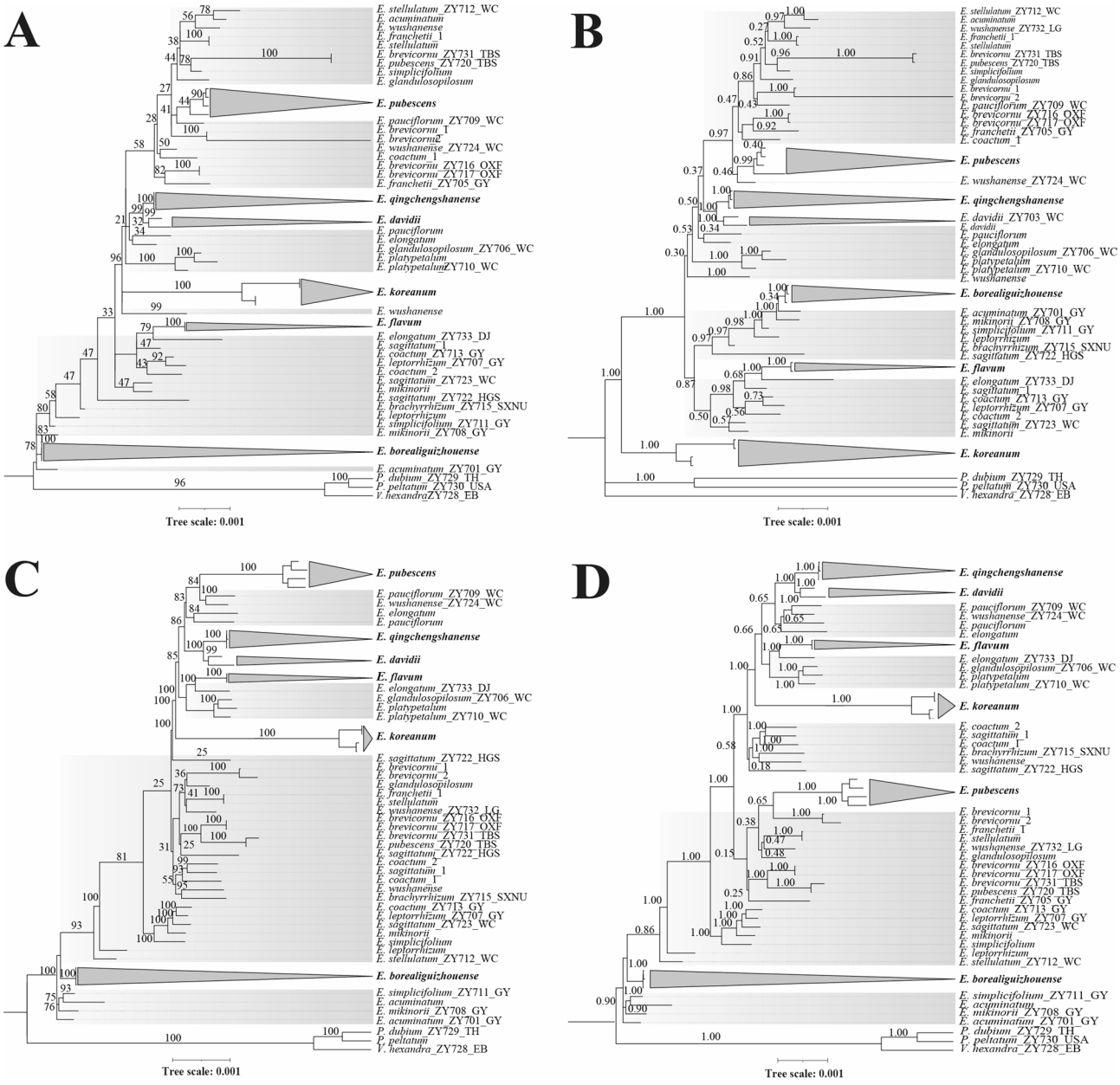

Fig. 6 Phylogenetic tree of *Epimedium* via the maximum-likelihood (ML) and Bayesian inference (BI) method, based on two datasets. **A** ML tree based on hypervariable regions. **B** BI tree based on hypervariable regions. **C** ML tree based on complete plastomes. **D** BI tree based on complete plastomes. Support values > 50% (ML bootstrap)/>0.5 (BI posterior probability) are indicated. All green indicates the species form a monophyletic group in the topological structure with node support values of ≥ 70% bootstrap (ML) and ≥ 0.95 posterior. probability (BI)

resolution. In conclusion, the standard barcodes, whether used alone or in combination, effectively identified at most four species: *E. borealiguizhouense*, *E. flavum*, *E. koreanum*, and *E. qingchengshanense*. The tandem hypervariable region dataset exhibited the same circumscription rate as the complete chloroplast genome and showed a slight improvement over the standard barcodes. Overall, both ML and BI methods yielded identical detection rates. Although DNA special barcode in *Epimedium* exhibit distinct advantages for species delineation, low branch support values were observed for certain clades (Fig. 6, S4, S5).

Single-copy nuclear genes assembly and comparative

Illumina sequencing produced raw data for each sample within a range of approximately 35.165 to 47.798 GB. The number of genes in *Epimedium* ranged from 780 (*Epimedium simplicifolium* 37) to 843 (*E. pubescens* 76D), and the number of genes in the outgroup *Vancouveria chrysantha* was 646 (Table S2). After gene length and

coverage Quality filtering, we obtained a dataset of 698 SCNs. The SCN datasets assessed using both ML and BI methods exhibited a significant improvement in support values compared to standard barcodes, complete chloroplast genomes, and variable regions. *Epimedium stellulatum* proves to be indistinguishable using Super barcodes or standard barcodes, whereas SCN data have favorable species discrimination power, with a support value of 100% (Fig. 7).

Phylogenetic discordance between chloroplast and nuclear data

Phylogenetic analyses of SCNs (Fig. 8A) and chloroplast genomes (Fig. 8B) revealed substantial topological

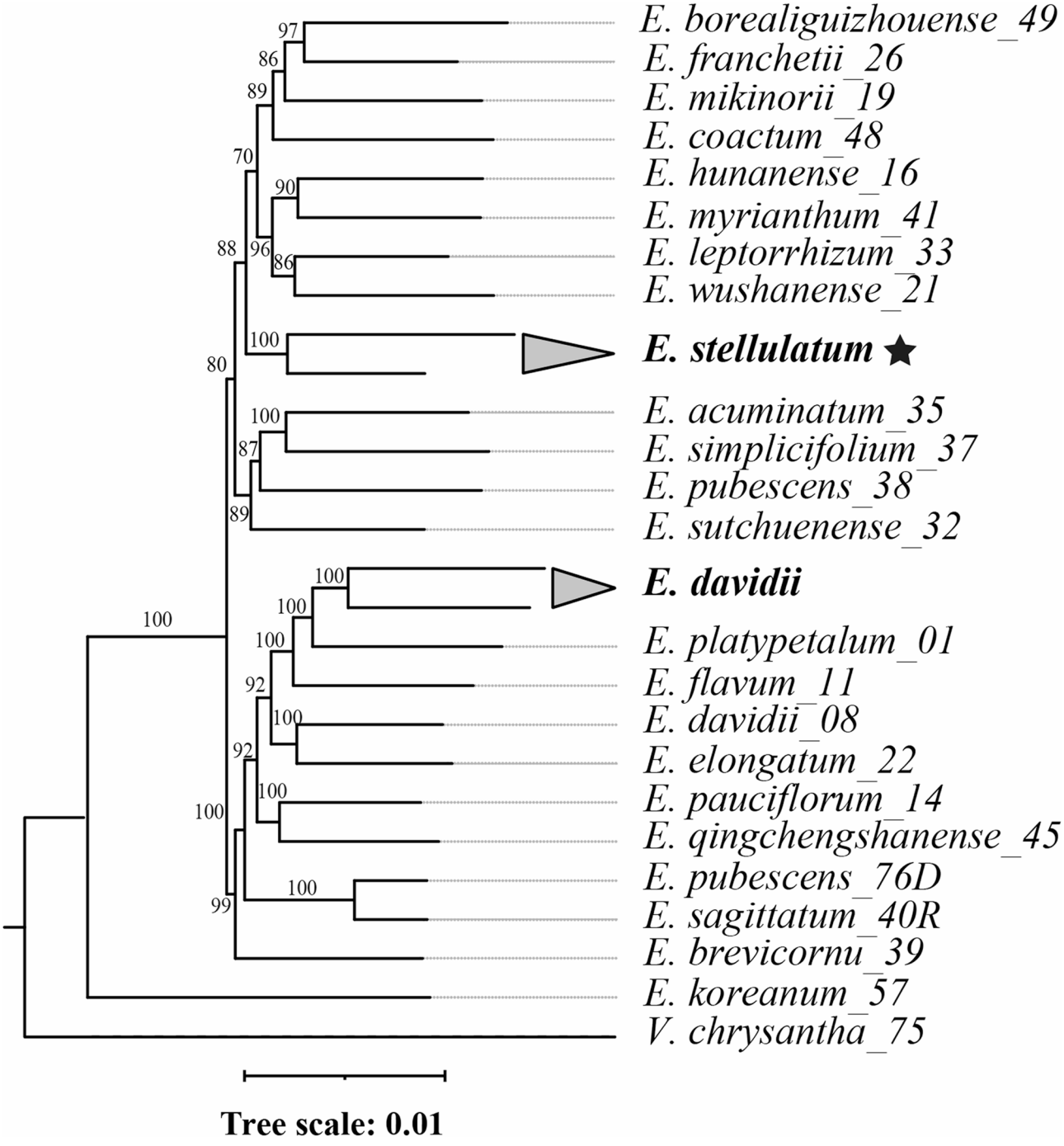

Fig. 7 Phylogenetic tree of *Epimedium* using maximum-likelihood (ML) based on single-copy nuclear genes. Support values > 50% (ML bootstrap) are indicated.All green indicates the species form a monophyletic group in the topological structure with node support values of ≥ 70% bootstrap (ML) and ≥ 0.95 posterior probability (BI)

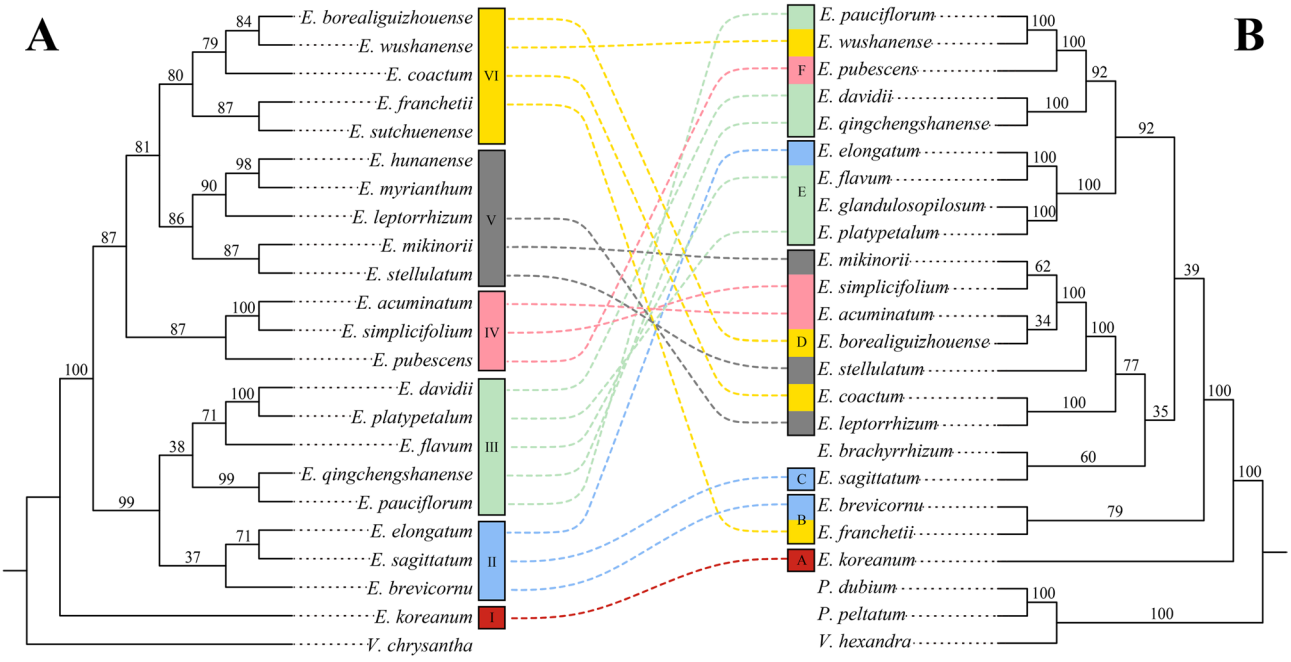

Fig. 8 Phylogenetic conflict between chloroplast and SCNs data in *Epimedium*.**A** ML tree based on single-copy nuclear genes. **B** ML tree based on complete plastomes. The six major clades in the SCN-based tree are labeled with Roman numerals (Ⅰ-Ⅵ), while the corresponding clades in the plastome-based tree are labeled with letters (A | F). The colors of the branches in the SCN tree correspond to the same species in the plastome tree, highlighting the phylogenetic discordance between the two datasets

discordance, underscoring complex evolutionary histories within *Epimedium*. Both trees resolved six major clades (labeled I-VI in SCNs tree and A-F in the chloroplast tree), yet with striking incongruences. For instances, the SCNs-based phylogenetic tree showed robust support (bootstrap support > 85%) for key groupings, with *E. mikinorii* and *E. stellulatum* forming a cohesive lineage in clade Ⅴ. In contrast, the chloroplast genome-based phylogenetic tree demonstrated a monophyletic grouping of *E. mikinorii* with *E. simplicifolium*. *Epimedium acuminatum* and *E. borealiguizhouense* constituted a distinct monophyletic clade, which together with the clade containing *E. stellulatum*, were nested within a higher-level monophyletic clade. Morever, *E. pauciflorum*, which clustered with *E. wushanense* in the chloroplast genome-based phylogenetic tree, formed a strongly supported clade (bootstrap support > 95%) with *E. qingchengshanense* in the SCNs phylogeny. Similarly, *E. davidii* shifted from a sister relationship with *E. qingchengshanense* in the chloroplast genome-based phylogenetic tree to grouping with *E. platypetalum* in the SCN-based phylogenetic tree.

Micro-morphological characteristics-based species discrimination

Leaves from 11 species in *Epimedium* disclosed extensive micro-morphological disparities (Table 4). The adaxial epidermal cells of all species are irregular in shape, with a few cells flattened (Fig. 9E1, S6D1, E1, F1), a few slightly elevated (Fig. 9B1, D1, S6A1, B1), and the others slightly concave (Fig. 9A1, C1, S6C1). The anticlinal walls of abaxial epidermal cells undulate in all species examined, with variations ranging from subtle protrusions (Fig. 9A2, B2, C2, D2, S6A2, B2, C2, D2) to flush alignment with the cell surface (Fig. 9E2, S6E2, F2).

Surfaces of most species presented epicuticular waxes, which were scarce in *E. elongatum* and *E. fangii* (Fig. 9C1–3, D1–3). A broad epidermal wax heterogeneity was observed across species, with higher concentrations found on the abaxial side, contrasting with the lower abundance of adaxial waxes (Fig. 9, S6). Variations in wax morphology consisted of granular (Fig. 9A1, B1, S6A1), membranous platelets (Fig. S6C1, D1, E1), and partial reticulations (Fig. 9E1, S6B1, F1). Meanwhile, the abaxial epidermal wax compositions in diverse species exhibited unique patterns, featuring membranous platelets (Fig. 9B2, S6B2, C2, E2), complemented by reticulate structures in some instances (Fig. 9A2, E2, S6A2), and thread wax patterns in specific situations (Fig. 9B2, S6F2).

Even the stomata of some species are covered with incomplete wax granules or platelets (Fig. 9A3, E3, S6C3, F3). Stomata are randomly distributed only on the abaxial surface, and the stomata are all elliptical in shape. In addition, some stomata were surrounded by distinct protuberances of different shapes, such as rhombic (Fig. 9E2), circular (Fig. S6B2) and triangular (Fig. S6E2), pentagonal (Fig. 9B2). Third, the thickness of the guard cells varied among the species studied. Some species were thick (Fig.

Table 4 Quantitative and qualitative characteristics of leaf epidermal cell across 11 *Epimedium* species

Plant name	Collection date	Collection location	Voucher	Epidermal cell		Stomata						Epidermal wax layer
				Cell shape AD/AB	Anticlinal cell wall pattern AD/AB	Distribution	Shape of stomata	Shape of stomatal pore	Shape of guard cell	Mean stomatal size (μm) L×W	Mean stomatal pore size (μm) L×W	
Epimedium brevicornu	19,590,609	China: Shaanxi	6694	irregular	undulate	abaxial	wide elliptic	elongated elliptic	kidney shaped	23.42×20.07	15.99×5.64	granular and reticulate
Epimedium davidii	193,506	China: Shaanxi	63,581	irregular	undulate	abaxial	wide elliptic	elliptic	kidney shaped	19.19×16.29	10.08×5.71	granular and membranous platelets
Epimedium elongatum	19,580,602	China: Sichuan	77,574	irregular	undulate	abaxial	circular	elongated elliptic	kidney shaped	25.23×23.35	17.59×8.01	nearly smooth
Epimedium fangii	1950	China: Sichuan	54,585	irregular	undulate	abaxial	circular	elliptic	narrow kidney shaped	17.37×16.93	8.41×4.98	nearly smooth
Epimedium franchetii	19,640,409	China: Chongqing	107,552	irregular	undulate	abaxial	wide elliptic	elliptic	narrow kidney shaped	22.08×17.84	11.64×6.04	partial reticulate
Epimedium hunanense	19,621,028	China: Hunan	015367	irregular	undulate	abaxial	wide elliptic	elongated elliptic	kidney shaped	24.79×19.97	15.54×4.98	granular and reticulate
Epimedium pubescens	19,640,507	China: Shaanxi	2588	irregular	undulate	abaxial	wide elliptic	elongated elliptic	kidney shaped	24.98×16.62	14.90×6.32	partial reticulate and membranous platelets
Epimedium sagittatum	20,060,810	China: Gansu	2,006,017	irregular	undulate	abaxial	wide elliptic	elongated elliptic	narrow kidney shaped	26.26×20.63	17.41×6.59	membranous platelets
Epimedium stellulatum	19,590,809	China: Hubei	10,797	irregular	undulate	abaxial	wide elliptic	elliptic	kidney shaped	27.98×23.15	15.18×8.32	membranous platelets
Epimedium sutchuenense	10,040,408	China: Shaanxi	258	irregular	undulate	abaxial	wide elliptic	elongated elliptic	kidney shaped	23.93×21.58	12.45×5.63	membranous platelets
Epimedium wushanense	20,050,612	China: Shaanxi	2590	irregular	undulate	abaxial	wide elliptic	elongated elliptic	kidney shaped	24.66×18.69	13.66×5.91	partial reticulate

AD and AB refer to adaxial and abaxial leaf surfaces, respectively. Cell dimensions (L×W) indicate length × width (μm)

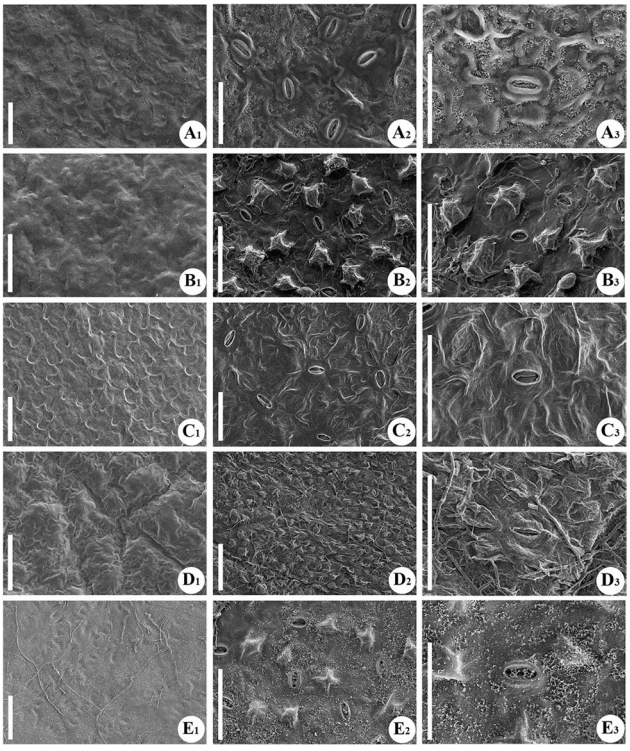

Fig. 9 Characteristics of the epidermal surface of leaves. **A_1–A_3***Epimedium brevicornu*. **B_1–B_3***Epimedium davidii*.**C_1–C_3***Epimedium elongatum*.**D_1–D_3***Epimedium fangii*.**E_1–E_3***Epimedium franchetii*. The images show the epidermis of adaxial leaves and abaxial leaves, and the magnified stomatal apparatus on the abaxial surface in each row from the left to right, respectively

S6B3, C3), moderately thick (Fig. 9A3, S6D3, E3, F3), and some were thin (Fig. 9B3, C3, D3, E3, S6A3).

Homogeneity of variance tests and univariate ANOVA confirmed that all variables met the assumptions of equal variance and exhibited significant interspecific differences (Table S3). Post hoc Tukey HSD tests further delineated species-specific contrasts (Fig. S7; Tables S4, 5, 6, 7). *Epimedium brevicornu* showed highly significant differences in the stomatal area compared to *E. davidii* (mean difference = 261.01, $p < 0.001$), *E. fangii* (mean difference = 252.50, $p < 0.001$), and *E. franchetii* (mean difference = 218.51, $p < 0.001$), but exhibited minor differences with *E. sagittatum* and *E. stellulatum* ($p \geq 0.05$). *Epimedium elongatum* differed significantly from *E. brevicornu* (mean difference = −101.49, $p < 0.001$) and *E. davidii* (mean difference = −362.49, $p < 0.001$), but not from other species, suggesting that stomatal area holds high diagnostic value in specific species pairs.

Stomatal length-to-width ratios also varied significantly among species. For instance, *E. brevicornu* differed markedly from *E. elongatum* (mean difference = 0.218, $p = 0.002$), *E. fangii* (mean difference = 0.262, $p < 0.001$), and *E. sutchuenense* (mean difference = 0.219, $p = 0.002$), but showed no significant differences with other species ($p \geq 0.05$). *Epimedium elongatum* also differed significantly from *E. hunanense* (mean difference = −0.189, $p = 0.014$) and *E. wushanense* (mean difference = −0.205, $p = 0.006$), indicating that stomatal length-to-width ratios serve as a key feature for distinguishing certain species.

In stomatal pore area, *E. elongatum* differed significantly from most species (e.g., *E. davidii*, *E. fangii*, *E. franchetii*, *E. hunanense*, *E. pubescens*, *E. sutchuenense*; $p < 0.001$), whereas *E. pubescens* showed no significant differences with *E. brevicornu*, *E. fangii*, *E. franchetii*, or *E. hunanense* ($p \geq 0.05$), suggesting strong discriminatory power of stomatal pore area among some closely related species (Fig. S7C). For stomatal pore length-to-width ratios, *E. brevicornu* differed significantly from *E. davidii* (mean difference = 1.024, $p = 0.003$), *E. fangii* (mean difference = 0.941, $p = 0.008$), *E. stellulatum* (mean difference = 0.848, $p = 0.024$), and *E. sutchuenense* (mean difference = 0.899, $p = 0.013$). *Epimedium hunanense* exhibited significant differences from most species (absolute mean difference > 1.2, $p < 0.001$), indicating its unique stomatal pore length-to-width ratio as a diagnostic trait (Fig. S7D).

Principal Component Analysis (PCA) extracted two principal components, cumulatively explaining 83.90% of the variance (Fig. S8). Principal Component 1 (PC1) accounted for 51.10% of the total variance, primarily reflecting stomatal size traits, dominated by stomatal area (loading = 0.962) and stomatal pore area (loading = 0.967), which were highly positively correlated ($r = 0.876$, $p < 0.001$). Principal Component 2 (PC2) explained 32.80% of the variance, capturing stomatal shape characteristics, driven by stomatal length-to-width ratio (loading = 0.856) and stomatal pore length-to-width ratio (loading = 0.856), which showed moderate positive correlation ($r = 0.479$, $p < 0.001$).

The PCA score plot (Fig. S8) displayed clustering patterns in the PC1-PC2 space. For example, *E. stellulatum* scored high on PC1 (indicating larger stomatal and pore areas) but low on PC2, while *E. hunanense* scored markedly high on PC2 (reflecting elongated stomatal morphology with higher length-to-width ratios). The above analyses collectively demonstrate that the leaf epidermal micro-morphological characteristics of *Epimedium* exhibit typical inter-specific specificity, serving as a complementary approach to molecular evidence for species delimitation (Fig. S9).

Discussion

The structure of the chloroplast genome in Epimedium and its evolutionary implications

The chloroplast genome of *Epimedium* exhibits a dynamic equilibrium between structural conservation and lineage-specific differentiation, providing critical clues for unraveling the evolutionary history of this genus [61, 63, 64]. Although the IR/LSC/SSC boundaries are

highly conserved in most species (Fig. 3, S3), consistent with the slow evolutionary rate of chloroplast genomes, two notable exceptions reveal adaptive differentiation phenomena. In our study, the IR region within the genus *Epimedium* of Podophylloideae varies significantly by up to about 1,000 bp, while the Subfamily Berberidoideae displays a remarkable amplification of the IR region, with an increase of approximately 10,000 bp (Fig. 3, S3). Secondly, there are interspecific differences in the localization of IR-SC boundary-adjacent genes (such as *rpl2*, *rpl22*, *rpl23*, *ycf1*, *rpl19*, *trnI*, and *trnH*). For example, *E. wushanense* is characterized by an expansion of the IR region compared to some species in *Epimedium*, a feature that can be used for preliminary species identification. However, samples of some species are subject to fragmentation in the plastid phylogenetic tree because the matrilineal genetic properties of chloroplasts may be affected by hybridization or incomplete genealogical sorting [29]. While these boundary shifts themselves lack direct species-discriminating power, they provide a contextual framework for evolutionary trajectories—conserved boundaries reflect deep genomic stability, whereas structural anomalies suggest lineage-specific adaptations [70–72].

A genus-defining *trnQ-UUG* rearrangement further underscores chloroplast architecture' s taxonomic relevance. Unique to *Epimedium* (Fig. 2 and 3), this tRNA-Q locus variation serves dual roles: its presence distinguishes the genus from other Berberidaceae [73, 74]. At the same time, intra-generic duplication events in the LSC region correlate with speciation patterns (Fig. 2 and 3). Unlike the homogenous SSR distributions (Fig. 4), which suggest recent radiation or conserved ancestral polymorphisms [75, 76]. This structural innovation clarifies species relationships through duplication mapping, bridging the gap between genomic architecture and taxonomic resolution.

Repeat element analyses reveal additional evolutionary pressures. Microsatellites, tandem repeats, and large repeats within the genus *Epimedium* exhibit variations in quantity, type, distribution, and length across different species (Fig. 4). This hints at either recent diversification preserving ancestral patterns or selective maintenance of functional repeats [76]. While insufficient for species delimitation alone, these repeat signatures complement nuclear and morphological data by exposing cryptic evolutionary forces shaping genome organization.

Collectively, chloroplast structural features resolve *Epimedium* taxonomic complexities through evolutionary contextualization. Conserved elements (e.g., majority IR boundaries, SSR abundance) reflect deep phylogenetic stability, while exceptions (*trnQ-UUG* rearrangements, lineage-specific IR contractions) mark adaptive innovations. Critically, these patterns gain taxonomic relevance when integrated with multilayered data: the *trnQ-UUG* rearrangement validates genus monophyly while guiding species-level distinctions, and repeat distributions expose evolutionary pressures invisible to single-marker analyses. This synthesis is consistent with the study's overall framework, demonstrating how structural conservation and divergence together reveal a dual narrative of ancient genomic stability and recent adaptive diversification in the genus.

Multi-source data for species circumscription in Epimedium

The limitations of standard barcodes became evident when addressing challenging taxa, prompting the investigation of more comprehensive genomic approaches [44]. The introduction of complete chloroplast genomes expanded the scope of species circumscription, offering enhanced resolution and detail [11]. Yet, the high costs associated with this method impeded its widespread adoption. Consequently, the focus shifted to hypervariable regions within specific taxa, designated as special barcodes [77]. In our study, we utilized the whole chloroplast genome and highly variable regions to improve the accuracy of the phylogenetic tree and species discrimination rate, surpassing traditional molecular markers such as *rbcL*, *matK*, and *psbA-trnH*. (Fig. S4, S5). Here, we were able to effectively identify *E. pubescens* when traditional molecular markers failed to, but this approach was not exhaustive for circumscribing other species within *Epimedium*. Specifically, compared with the core barcodes, circumscription based on the chloroplast genome is usually more reliable.

According to our findings, the adoption of SCNs led to a marked increase in the support values of the constructed phylogenetic tree of *Epimedium*, far surpassing the reliance on chloroplast DNA (Fig. 7). Despite the inherent limitations of our study, including the relatively sparse sampling across species and the absence of duplicate sequences for robust statistical validation, it is in line with other findings [78–80]. Collaborative efforts have collectively painted a compelling picture of the immense potential of SCNs in species discrimination. By tapping into the untapped potential of the nuclear genome, SCNs have the capacity to serve as robust nuclear molecular markers, offering an unprecedented level of precision and accuracy in species delimitation [81].

The nuclear-cytoplasmic discordance between chloroplast phylogenies and SCN topologies highlights the limitations of single-locus molecular approaches in delimiting species boundaries. Such incongruences likely stem from biological processes including interspecific hybridization, incomplete lineage sorting, and distinct evolutionary pressures on nuclear versus plastid genomes [82].

Complementing a genomic approach, we advocate for the inclusion of micro-morphological characters as micro-morphological barcodes in the species circumscription process (*Pulsatilla*: [83]; *Mahonia*: [84]). Here, our study reveals different micro-morphological features among different species in *Epimedium* (Fig. 9, S6). For instance, phylogenetic analyses using either chloroplast genomes or single-copy nuclear genes consistently demonstrate non-monophyly in *E. franchetii* and *E. sagittatum* samples. This indicates the inadequacy of these molecular markers for precise species delimitation. However, leaf epidermal micro-morphology revealed striking divergences. In *E. franchetii*, both adaxial and abaxial surfaces exhibit granular epidermal wax coalescing into incomplete reticulate plates. Its stomata are additionally covered by fragmented reticulate wax, surrounded by irregular quadrilateral protrusions. In contrast, *E. sagittatum* displays dense granular epidermal wax on both leaf surfaces, aggregating into partial reticulate structures in specific regions, with stomata lacking irregular protrusions. Principal component analysis (PCA) distinctly clustered these two species into separate groups (Fig. S8). Post hoc tests further demonstrated significant differences ($P < 0.01$) in stomatal area, stomatal aperture area, and aperture length-to-width ratio. These findings substantiate that leaf epidermal micro-morphology provides effective diagnostic criteria when molecular evidence is insufficient for taxonomic delimitation [60].

Drivers of phylogenetic discordance in Epimedium

Phylogenetic trees of *Epimedium* were constructed based on 698 single-copy nuclear genes (Figs. 7 and 8A), revealing significant topological discrepancies between them. Tree A (Fig. 7), which incorporated replicate samples within species, exhibited consistently high nodal support (all bootstrap values > 70%). In contrast, Tree B (Fig. 8A), constructed after removing replicate samples, showed substantially reduced support at specific nodes (e.g., down to 38%) and altered the clustering relationships of key species. This inconsistency was primarily driven by: (1) Intraspecific genetic variation introduced by replicate samples obscured interspecific differentiation signals [85]. In Tree A, multiple samples of the same species (e.g., *E. stellulatum* and *E. davidii*) formed highly supported monophyletic clades, but their intraspecific variation might be misinterpreted as interspecific divergence, leading to erroneous clustering (e.g., association between *E. stellulatum* and *E. mikinorii*). Tree B reduced intraspecific noise by excluding replicates, thereby exposing conflicting signals like incomplete lineage sorting (ILS) or gene flow among gene trees, which decreased nodal support. (2) Long-branch attraction (LBA) was more pronounced in Tree A [86]. Replicate samples formed long branches due to rapid intraspecific evolution, causing artifactual clustering with taxa having similar evolutionary rates. After removing long-branch samples, Tree B mitigated LBA effects and revealed the true topology, though support for some short branches declined due to weak signal. Additionally, data missingness and model bias exacerbated topological conflicts [87]. In Tree A, replicates filled missing data for certain genes, improving site occupancy and nodal support but potentially masking gene tree heterogeneity. In contrast, sparse data from single samples in Tree B made the algorithm sensitive to missing sites, resulting in support fluctuations. Future studies should integrate gene tree heterogeneity using species tree methods (e.g., ASTRAL) and validate contentious nodes with morphological or population genomic data.

The failure of individuals from the same species to cluster into a single branch on a phylogenetic tree (non-monophyly) typically stems from the complexity of evolutionary processes or methodological limitations. First, evolutionary mechanisms may play a dominant role. For instance, during early species divergence, incomplete lineage sorting (ILS) can occur if ancestral genetic polymorphisms are not fully fixed randomly, causing different individuals to carry distinct ancestral gene fragments and leading to chaotic genetic signals [88]. Natural hybridization or introgression between species may also disrupt species boundaries via gene flow, scattering individuals across divergent branches [89]. Second, biodiversity itself might be underestimated, such as unrecognized cryptic species (morphologically similar but genetically distinct) being misclassified, which distorts clustering outcomes [90]. Finally, convergent evolution or rapid radiative evolution can drive unrelated species to independently develop similar traits under analogous environmental pressures, despite significant genomic divergence, creating conflicts between morphological and molecular data [85]. These factors highlight the need for integrating multi-gene analyses, expanding sampling, and combining ecological and morphological evidence to resolve evolutionary relationships accurately.

Topological conflicts between organellar and nuclear phylogenies (e.g., chloroplast vs. SCNs) arise from inherent genetic mechanisms and methodological biases. The uniparental inheritance of chloroplasts fundamentally differs from biparental nuclear recombination, leading to divergent evolutionary trajectories. For instance, *E. pauciflorum* and *E. wushanense* cluster in chloroplast trees but form a sister clade with *E. qingchengshanense* in nuclear gene trees (Fig. 8), reflecting historical chloroplast capture via hybridization events [29, 30]. Incomplete lineage sorting (ILS) further explains unstable placements (e.g., *E. mikinorii* in Fig. 8), particularly in rapidly radiating groups like *Epimedium* that diversified during Himalayan uplift [58]. Methodological limitations

compound these conflicts: uneven taxonomic sampling (e.g., *E. hunanense* exclusively in SCN datasets) reduces topological comparability, while the low variability of chloroplast genomes (evidenced by limited barcode discrimination in Fig. 6) contrasts with the high resolution of SCNs capturing rapidly evolving loci (Fig. 7), necessitating integrated approaches for resolving recent radiations [42, 81].

Critically, these inconsistencies do not diminish the utility of individual datasets but instead highlight the complementary strengths of a multi-layered approach. The Multilayer Precision Species Circumscription Approach (MPSCA) leverages chloroplast genomes for deep evolutionary insights, SCNs for fine-scale species delimitation, and micro-morphology to validate molecular hypotheses—exemplified by distinguishing *E. franchetii* and *E. sagittatum* through epidermal wax patterns when genomic data alone proved insufficient (Fig. 9). By reconciling discordant signals across datasets, MPSCA transforms phylogenetic conflict into a diagnostic tool, advancing species delimitation in taxonomically complex groups where no single method suffices.

Towards a holistic approach to species circumscription

Our research now introduces a forward-looking strategy for species discrimination: the use of SCNs for precise species circumscription in plant groups where chloroplast genome has been proven inadequate. To overcome the limitations of individual methods of species circumscription, we propose a novel Multilayer Precision Species Circumscription Approach (MPSCA) that integrates chloroplast genome data, SCNs and micro-morphological features (Fig. 10). First, when we get several "unknown species" with processed, crushed, dried material, or other treated forms, we should use short molecular markers (such as ITS, *matK*, *rbcL*, and *trnH-psbA*) for preliminary delimitation. If the initial results were uncertain, we further employed chloroplast genome analysis to improve accuracy. If species delimitation remained unclear, we then utilized SCNs for delimitation. When identifying plants, we do not necessarily have to proceed to the very last step. When the range of plants determined at a certain step is small enough but still not precise, we can integrate micro-morphological evidence to improve the accuracy of identification.

To validate the MPSCA, this study implemented a hierarchical verification framework integrating molecular and morphological evidence. Initial screening using standard DNA barcodes successfully delineated four species (*E. borealiguizhouense*, *E. flavum*, *E. koreanum*, and *E. qingchengshanense*) as highly supported monophyletic clades, while ambiguous taxa like *E. pubescens* required further resolution through chloroplast genome analysis. Subsequent integration of hypervariable chloroplast regions established six species (including *E. pubescens*) as distinct lineages with enhanced nodal support. For taxa with persistent phylogenetic ambiguities in chloroplast phylogenies (e.g., *E. stellulatum*), single-copy nuclear genes (SCNs) provided critical discriminative power, resolving *E. stellulatum* as a strongly supported monophyletic group. Micro-morphological characteristics (e.g., stomatal area and size) were integrated to address cases with insufficient genetic resolution, effectively differentiating between *E. brevicornu* and *E. stellulatum*. Application scenario analysis revealed differential requirements: *E. koreanum* identification required only standard barcodes with micro-morphological verification, *E. pubescens* necessitated chloroplast genome and

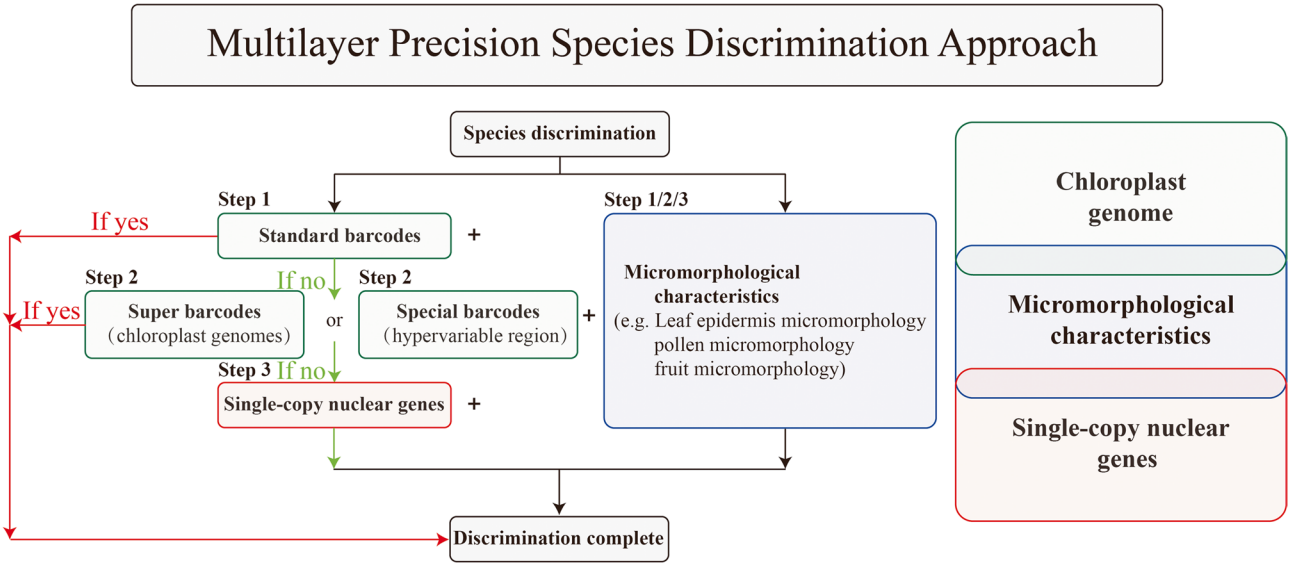

Fig. 10 Multilayer Precision Species Discrimination Approach

micro-morphological integration, while *E. stellulatum* demanded combined SCNs and micro-morphological analysis.

The significance of this approach lies in its capacity to effectively address the limitations inherent in traditional species identification methods [91, 92]. For instance, reliance solely on chloroplast genome data may result in inaccurate species delimitation, particularly in taxa characterized by hybridization or genetic introgression. In contrast, the incorporation of SCNs offers more reliable nuclear gene information, thereby compensating for the shortcomings of chloroplast genomes [93, 94]. Furthermore, the integration of micro-morphological traits enhances identification reliability, especially when molecular data fail to yield definitive conclusions for closely related species [94, 95]. By employing this multi-layered, evidence-based approach, we not only improve the accuracy of species identification but also provide robust data support for phylogenetic studies of complex taxa.

The application of this approach holds significant potential across various fields, including biodiversity research, conservation biology, and the development of medicinal plant resources. For instance, within the genus *Epimedium*, many species possess considerable medicinal value; however, their morphological similarities make them challenging to distinguish using traditional identification methods. Our multi-layered precise species delimitation approach (MPSCA) enables rapid and accurate identification of these species, thereby offering a scientific foundation for the sustainable development and conservation of medicinal resources. Moreover, this method can be adapted to other plant taxa, providing an efficient and standardized framework for advancing plant taxonomy and phylogenetic studies.

Limitations of this study include the insufficient number of replicate samples within the single-copy nuclear gene (SCN) dataset, which complicates the assessment of its species discrimination capability. For instance, uneven taxonomic sampling—such as *E. hunanense* being exclusively represented in the SCN dataset—limits cross-dataset comparability and may introduce biases in phylogenetic inference. Additionally, the extraction process for SCNs is complex, increasing both cost and labor [28]. Moreover, the computational demands associated with handling large-scale genomic data can be substantial [96]. These factors constrain the reliability and applicability of our findings. It should be noted that not all single-copy nuclear genes are used to identify species. Screening some nuclear gene fragments with hypervariable regions as much as possible is an ideal approach. Beyond methodological constraints, biological complexities such as morphological plasticity exacerbate taxonomic ambiguity. While expert validation and micro-morphological traits (e.g., stomatal wax patterns) were employed to mitigate errors, cryptic species detection in phenotypically plastic groups like *Epimedium* demands higher-resolution tools [90]. With ongoing technological advancements and increased data availability, this approach will be poised to further breakthroughs and advancements in the field of plant species circumscription, enabling a more accurate understanding of the phylogenetic relationships and species diversity of problematic groups.

Conclusion

This study shed light on the strong discriminatory potential of chloroplast and nuclear data, as well as the assistance of micro-morphological characteristics. We present the Multilayer Precision Species Discrimination Approach, which provides a rational approach to species circumscription, emphasizing its critical role in improving plant taxonomic accuracy and resource management efficiency. This comprehensive work breaks the boundaries of traditional DNA barcodes and improves the accuracy of species circumscription. Nonetheless, future research needs to address the problem of sparse sampling and simplify access to SCNs. At the same time, many plant taxa still lack chloroplast, single-copy nuclear gene barcodes and micro-morphological information, especially the latter two. The construction of molecular barcode and micro-morphological databases for global plants still faces huge challenges.

Materials and methods

Taxon sampling, DNA extraction, and sequencing

The plant samples in this study primarily comprise three parts: chloroplast genomes, single-copy nuclear genes, and leaf epidermal micro-morphology data.

Concerning the leaf epidermal micro-morphology data, we examined 11 samples from 11 species. The materials were sourced from the historical collections of Herbarium of Northwest A&F University (WUK), and voucher specimens are listed in Table 4.

Regarding chloroplast genome data, we assembled and annotated 30 chloroplast genome sequences, including 27 *Epimedium* sequences (covering 21 *Epimedium* species) and three outgroup sequences (*Plagiorhegma dubium*, *Podophyllum peltatum*, *Vancouveria chrysantha*) (Table 5). In addition, we got 27 fully assembled and annotated chloroplast genome sequences from the National Center of Biotechnology Information. Thus, the chloroplast genome dataset contains a total of 57 sequences. NCBI accession numbers for chloroplast genomes and raw sequencing data are provided in Table 1, with full specimen metadata available in Table S8. All voucher specimens were morphologically identified by Professor Zhao Liang and are deposited in WUK.

For the single-copy nuclear genes data, we collected a total of 27 samples from 22 species, including 26

Table 5 Voucher information and GenBank accession numbers for *Epimedium* and outgroup species from WUK and IMPD

Taxon	Collection Site	Collectors	Collection numbers	Date	Voucher specimen	Accession	SRA accession	Data analysis	
								Nuclear	Chloroplast
Epimedium acuminatum	China: Guizhou	Yu Zhang	701	20,210,403	WUK ZY701	PV192894	SRX27229214		√
	China: Guizhou	Mengyue Guo	0342	20,180,520	IMPD Guo0342	PV192895	SRX20954640		√
						OQ366391			√
Epimedium borealiguizhouense	China: Guizhou	Yu Zhang	702	20,210,403	WUK ZY702		SRX27229215		√
	China: Guizhou	Mengyue Guo	0338	20,180,520	IMPD Guo0338		SRX20954655	√	
						MN873561			√
						NC044889			√
Epimedium brachyrrhizum	China: Shaanxi	Yu Zhang	715	20,200,930	WUK ZY715	PV192904	SRX27229220		√
Epimedium brevicornu	UK	Yu Zhang	716	20,190,820	WUK ZY716	PV192900	SRX27229221		√
	China: Shaanxi	Yu Zhang	731	19,390,505	ZY731	PV192902	SRX27229234		√
	China: Gansu	Xiang Liu	18,042	20,180,520	Liu18042		SRX20954644	√	
						MT560406			√
						MN714008			√
Epimedium coactum	China: Guizhou	Yu Zhang	713	20,210,403	ZY713	PV192907	SRX27229218		√
	China: Guizhou	Mengyue Guo	0319	20,180,520	Guo0319		SRX20954654	√	
						MT560395			√
						MT560402			√
Epimedium davidii	China: Sichuan	Yu Zhang	703	20,210,422	ZY703	PV192910	SRX27229226		√
	China: Sichuan	Xiang Liu	18,030	20,180,421	Liu18030		SRX20954646	√	
	China: Sichuan	Xiang Liu	18,033	20,180,512	Liu18033		SRX20954679	√√	
	China: Sichuan	Xiang Liu	18,036	20,180,516	Liu18036		SRX20954690	√	
						MT560419			√
Epimedium elongatum	China: Sichuan	Yu Zhang	733	19,740,603	ZY733	PV192915	SRX27229236		√
	China: Sichuan		287	20,190,413	Zhang287		SRX20954626	√	
						NC080247			√
Epimedium flavum	China: Sichuan	Yu Zhang	704	20,210,422	ZY704	PV192918	SRX27229237		√
	China: Sichuan		289	20,190,413	Zhang289		SRX20954614	√	
						MW483093			√
Epimedium franchetii	China: Guizhou	Yu Zhang	705	20,210,403	ZY705	PV192906	SRX27229238		√
	China: Hubei	Xiang Liu	18,021	20,180,520	Liu18021		SRX20954659	√	
						NC065479			√
Epimedium glandulosopilosum	China: Sichuan	Yu Zhang	706	20,210,422	ZY706	PV192916	SRX27229239		√
						MT560393			√
Epimedium hunanense	China: Hunan	Mengyue Guo	0402	20,180,520	Guo0402		SRX20954619	√	
Epimedium koreanum	China: Jilin	Yu Zhang	718	20,140,427	ZY718	PV192919	SRX27229223		√
	China: Jilin	Yu Zhang	719	20,140,427	ZY719	PV192920	SRX27229224		√
	Korea		05	20,190,801	JP05		SRX20954664	√	
						MT560411			√
						KM207675			√
Epimedium leptorrhizum	China: Guizhou	Yu Zhang	707	20,210,403	ZY707	PV192909	SRX27229240		√
	China: Guizhou	Mengyue Guo	0321	20,190,413	Guo0321		SRX20954638	√	
						NC062453			√
Epimedium mikinorii	China: Guizhou	Yu Zhang	708	20,210,403	ZY708	PV192896	SRX27229241		√
	China: Hubei	Mengyue Guo	0431	20,180,520	Guo0431		SRX20954622	√	
						NC044890			√

Table 5 (continued)

Taxon	Collection Site	Collectors	Collection numbers	Date	Voucher specimen	Accession	SRA accession	Data analysis	
Epimedium myrianthum	China: Hunan	Mengyue Guo	0421	20,190,413	Guo0421		SRX20954647	√	
Epimedium pauciflorum	China: Sichuan	Yu Zhang	709	20,210,422	ZY709	PV192912	SRX27229242		√
	China: Sichuan		278	20,190,413	Zhang278		SRX20954617	√	
						NC044890			√
Epimedium platypetalum	China: Sichuan	Yu Zhang	710	20,210,422	ZY710	PV192917	SRX27229243		√
	China: Sichuan	Xiang Liu	18,038	20,180,518	Liu18038		SRX20954612	√	
						MW483078			√
Epimedium pubescens	China: Shaanxi	Yu Zhang	720	20,140,406	ZY720	PV192903	SRX27229225		√
	China: Sichuan	Yu Zhang	721	20,210,422	ZY721	PV192913	SRX27229227		√
	China: Shaanxi	Mengyue Guo	0456	20,180,520	Guo0456		SRX20954643	√	
	China: Sichuan			20,190,326	JYZRM		SRX20954684	√	
						MT560420			√
						NC053532			√
						MN939633			√
Epimedium qingchengshanense	China: Sichuan	Yu Zhang	714	20,210,422	ZY714	PV192911	SRX27229219		√
	China: Sichuan	Xiang Liu	16,025	20,190,413	Liu16025		SRX20954651	√	
						NC053539			√
						MT560415			√
Epimedium sagittatum	China: Shaanxi	Yu Zhang	722	20,210,729	ZY722	PV192905	SRX27229228		√
	China: Sichuan	Yu Zhang	723	20,210,422	ZY723	PV192908	SRX27229229		√
	China: Sichuan			20,190,326	JYZJY		SRX20954645	√	
						MN027267			√
Epimedium simplicifolium	China: Guizhou	Yu Zhang	711	20,210,403	ZY711	PV192897	SRX27229216		√
	China: Sichuan	Mengyue Guo	0342	20,190,413	Guo0343		SRX20954642	√	
						NC053529			√
Epimedium stellulatum	China: Sichuan	Yu Zhang	712	20,210,422	ZY712	PV192898			√
	China: Hubei	Xiang Liu	18,025	20,190,413	Liu18025		SRX20954661	√	
	China: Hubei	Mengyue Guo	0443	20,180,520	Guo0443		SRX20954648	√	
						MN027267			√
Epimedium sutchuenense	China: Hubei	Mengyue Guo	0437	20,180,520	Guo0437		SRX20954637	√	
Epimedium wushanense	China: Sichuan	Yu Zhang	724	20,210,422	ZY724	PV192914	SRX27229230		√
	China: Shaanxi	Yu Zhang	732	20,050,613	ZY732	PV192899	SRX27229235		√
	China: Hubei	Mengyue Guo	0434	20,180,520	Guo0434		SRX20954625	√	
						MN857417			√
Vancouveria chrysantha	England		09	20,180,706	Kew09		SRX20954685	√	
Vancouveria hexandra	USA	Yu Zhang	728	2021	ZY728	PV192921	SRX27229231		√
Plagiorhegma dubium	USA	Yu Zhang	729	20,160,518	ZY729	PV192922	SRX27229232		√
Podophyllum peltatum	USA	Yu Zhang	730	20,150,501	ZY730	PV192923	SRX27229233		√

Epimedium samples and one outgroup sample, *V. chrysantha*. All specimens were identified by Professor Baolin Guo, and the voucher specimens are deposited at the Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences (IMPD). Details are listed in Table 5 and S2.

The genus *Epimedium* comprises approximately 60-70 species globally, with China serving as its center of diversity, harboring ~ 80% of the described species, many of which are endemic and medicinally significant. Our sampling strategy prioritized Chinese endemic taxa and morphologically complex species frequently involved in medicinal adulteration. Specifically, species were selected to: (1) represent major geographic distributions across China, covering key biodiversity hotspots such as the Hengduan Mountains, Qinling Mountains,

and southwestern karst regions; (2) include taxonomically challenging groups with overlapping morphological traits (e.g., *E. sagittatum*, *E. acuminatum*); and (3) incorporate species of high medicinal value but prone to misidentification in commercial markets (e.g., *E. brevicornu*, *E. wushanense*). This approach ensured coverage of both phylogenetic diversity and practical identification challenges.

The total genomic DNA was extracted using the cetyltrimethylammonium bromide (CTAB) method [97] and sequenced using the Beijing Genetic Institute (BGI) Illumina Miseq platform (Illumina, San Diego, CA, USA). Raw sequences were trimmed using Trimmomatic v0.40 [98] to remove low-quality reads and adapter sequences.

Chloroplast genome assembly and annotation

We collected approximately 11 GB of high-quality data for each sample (Table 1) and assembled the quality-filtered reads using GetOrganelle [99], with *Epimedium coactum* (GenBank accession: MT560395) used as a reference. Chloroplast genomes were annotated using the Plastid Genome Annotator (PGA: [100]) with the default parameters, and verified for accuracy by GeSeq [101]. Finally, we performed minor manual corrections using Geneious v11.0.20 [102]. The circular plastome map of *Epimedium* was then visualized using the online tool OGDRAW [103].

Single-copy nuclear marker development and single-copy nuclear genes assembly

The SCN marker development and the assembly process are referenced in Liu et al. (2021) [42], and whole genome and transcriptome data were obtained for *E. pubescens* (China National Center for Bioinformation accession: GWHBECS00000000), *Nandina domestica* (OneKP), *Podophyllum peltatum* (OneKP), and *Coptis chinensis* (GenBank assembly accession: GCA 015680905.1). All coding regions were extracted from the genomes using gffread [104], and input into MarkerMiner v1.2 [105]. To identify putative single-copy genes in *Epimedium*, MarkerMiner v1.2 was used with the following specific parameters: -minTranscriptLen 600, -minProteinCoverage 50, -minTranscriptCoverage 50, and -minSimilarity 50. The genes extracted through these parameters were then utilized as reference genes for subsequent assembly.

We used Trimmomatic v. 0.40 to remove adapters and Quality filters on raw reads. The results are then input into FastQC v. 0.11.9 [106] for quality control. Finally, Hybpiper v2.1.2 [107] was utilized to assemble single-copy nuclear genes. Briefly, we used BWA v. 0.7.1 [108] to align reads from *Epimedium* samples to the reference genes. Reads were assembled into contiguous sequences (contigs) using SPAdes v. 3.15.0 [109]. Assembled contigs were aligned to the reference genes using Exonerate v. 2.2.0 [110].

Retrieving standard DNA barcodes

To assess the discriminatory power of standard DNA barcodes (*matK*, *rbcL*, and *trnH-psbA*) in *Epimedium*, we extracted these regions from 57 complete chloroplast genomes (27 newly sequenced and 30 publicly available; see Table 1 for GenBank accession numbers). Gene boundaries for *matK*, *rbcL*, and *trnH-psbA* were identified using Geneious v11.0.20. Extracted sequences were aligned with MAFFT v6.240 and trimmed to remove ambiguous regions [111].

Phylogenetic analyses were conducted using both maximum likelihood (ML) and Bayesian inference (BI) approaches. Single-marker datasets (*matK*, *rbcL*, *trnH-psbA*) and a concatenated dataset (combined *matK* + *rbcL* + *trnH-psbA*) were independently tested to assess resolution across methods. ML trees were reconstructed with RAxML v8 [112], while BI trees were generated using MrBayes 3.2 [113].

A species was deemed diagnosable only if all examined individuals formed a monophyletic group in topologies, with nodal support ≥ 70% bootstrap (ML) and ≥ 0.95 posterior probability (BI). Taxa violating this criterion were classified as "molecularly indeterminate".

Chloroplast genome-based analysis

For the comparative genomic analysis, one chloroplast genome from each species was chosen. These sequences were aligned using MAFFT v6.240, and then fine-tuned at the boundaries in Geneious. The comprehensive alignment matrix was visually displayed through the mVISTA v7.450 [114], with default settings and *E. acuminatum* ZY701 as the reference. We compared the characterization of IR-SC boundaries for 21 species of *Epimedium*. In addition, we also used *E. acuminatum* ZY701 as a representative of the genus *Epimedium* within the subfamily Podophylloideae, along with *Berberis aristata* (GenBank accession: MN746308), *Leontice incerta* (GenBank accession: NC043927), *Mahonia bodinieri* (GenBank accession: NC066184), *Nandina domestica* (GenBank accession: NC008336), *Plagiorhegma dubium* (GenBank accession: NC038103), *Podophyllum peltatum* (GenBank accession: NC037903), and *Vancouveria hexandra* (GenBank accession: NC042222), to conduct an IR-SC boundary analysis. Subsequently, we manually plotted the results of these analyses to visualize the data.

We conducted a genome-wide sliding window analysis on 27 *Epimedium* chloroplast genomes to evaluate nucleotide diversity (Pi values) and identify regions of high variation. Sliding window analysis was performed using DnaSP v6.12.03 [115] to assess nucleotide polymorphism and to screen hypervariable regions as DNA special

barcode candidates, with the following parameters: window length = 600, step size = 200.

We also performed a comparative analysis of simple sequence repeats (SSRs), tandem repeats, and four types of large repeats in 21 species of *Epimedium*. Microsatellites (SSRs) are identified by MISA [116], and the threshold is 10 repeat units. The threshold of SSRs is 6, 5, 5, 5 and 4 repeat units, respectively, for mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide. Tandem repeats were identified via Tandem repeats finder [117]. Furthermore, four distinct types of large repeats were examined using the online REPuter tool [118].

The chloroplast dataset was reconstructed through maximum likelihood (ML) and Bayesian inference (BI) methods. These inferences are based on two datasets: the complete chloroplast genome and the hypervariable region. Identification criteria as above.

Single-copy nuclear gene-based analysis

We aligned single-copy nuclear gene datasets using MAFFT to find similarities and differences between sequences. Low-quality and redundant sequences were removed using the default parameters of trimAl v1.5.0 [119]. Additionally, genes present in 80% of the 22 species and longer than 600 nucleotides were selected to identify genes with sufficient coverage across the taxa. We used ML approaches for constructing phylogenetic trees from the SCNs tandem dataset. The resulting ML tree was compared with the chloroplast genome-based ML tree to assess congruence between nuclear and plastid phylogenies (Fig. 8). Identification criteria as above. Our main purpose was to only demonstrate the nuclear-cytoplasmic conflict at the species level, and the samples of each species in Fig. 8 were randomly selected.

Micro-morphological trait observation

Micro-morphological traits play a pivotal role in addressing taxonomic ambiguities in *Epimedium*, especially when molecular data exhibit incomplete lineage sorting or hybridization. Diagnostic traits encompass leaf epidermal characteristics, trichome types, pollen exine sculpturing, seed coat patterns, floral micro-morphology, and fruit/seed appendages [120–122]. Among these, leaf epidermal traits were emphasized for detailed investigation due to their year-round accessibility, suitability for non-destructive sampling, and direct relevance to traditional medicinal applications, as leaves are the primary source of bioactive compounds in this genus [123, 124]. Critically, leaf epidermal characteristics establish a critical link between practical taxonomic identification and phylogenetic frameworks, underscoring the necessity of integrating environmentally stable micro-morphological features into species delineation criteria [125].

First, the leaves on the dried specimens were rehydrated gradually in a decreasing alcohol series (Table 4). The rehydrated leaflets were then immersed in 70% alcohol and cleaned of surface impurities using an ultrasonic cleaner (Kunshan Ultrasonic Instrument Co., Ltd., China). Next, the leaflets were dehydrated in an incremental alcohol and isoamyl acetate series followed by critical point drying in CO_2 using a K850 critical point dryer (EMITECH, Ashford, UK). The leaf blades were then mounted on a column and sputter coated with gold-palladium using a JS-1600 sputter coater (HTCY, China). The materials were photographed using a Hitachi S-3400 scanning electron microscope (SEM, Hitachi, Japan) at 5 kV. Images were captured, and stomatal-related data were measured using ImageJ [126].

Statistical analyses were conducted using IBM SPSS Statistics v27.0. One-way analysis of variance (ANOVA) followed by Tukey HSD post hoc tests were applied to assess interspecific differences in stomatal traits (area, length-to-width ratio, pore area, and pore length-to-width ratio). Principal component analysis (PCA) was performed to reduce dimensionality, using a varimax rotation to extract components with eigenvalues > 1 [127]. Hierarchical clustering analysis (HCA) was conducted based on Euclidean distance, with the between-groups linkage method (UPGMA) to generate dendrograms. All micro-morphological variables were standardized (z-scores) prior to PCA and HCA to ensure equal weighting.

Abbreviations

NGS	Next-generation sequencing
DGS	Deep genome skimming
ITS	Internal transcribed spacer
SCN	Single-copy nuclear gene
mtDNA	Mitochondrial DNA
nrDNA	Nuclear ribosomal DNA
LSC	Large single copy
SSC	Small single copy
IR	Inverted repeat
tRNA	Transfer RNA
rRNA	Ribosomal RNA
SSRs	Simple sequence repeats
ML	Maximum likelihood
BI	Bayesian inference

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-12108-5.

Supplementary Material 1

Acknowledgements

We thank Xiaohua He (Plant Protection Science and Technology Innovation and Talent Cultivating Platform), Guoyun Zhang (State Key Laboratory of Crop Stress Biology for Arid Areas) of Northwest A&F University for assistance with SEM.

Authors' contributions
Y.-J.F. conducted the investigation, developed methodology, performed software work, conducted formal analysis, created visualizations, and wrote/edited the original draft. C.-G.L. and C.M. developed methodology, performed software work, and wrote the original draft. M.H. conducted formal analysis and wrote the original draft. J.-P.M. developed methodology and conducted formal analysis. M.-R.L. performed software work and conducted formal analysis. F.J. and R.G.J.H. contributed to conceptualization, methodology, supervision, validation, and manuscript review/editing. L.Z. and Q.Y. contributed to conceptualization, funding acquisition, investigation, methodology, project administration, supervision, validation, and manuscript review/editing. All authors read and approved the final manuscript.

Funding
This work was supported by the Natural Science Fundamental Research Plan of Shaanxi Province (grant number 2025JC-YBQN-252) and National Natural Science Foundation of China (grant numbers 32570249, 32170381 and 31770200).

Data availability
Data generated or analyzed during this study are provided within the manuscript or supplementary information files. Raw Illumina sequencing reads have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession codes PRJNA1202853 and PRJNA990516. Sequence data supporting this study are available in the GenBank database under accession numbers PV192894–PV192923. The single-copy nuclear gene sequence alignment matrices are publicly accessible via the Figshare repository at https://doi.org/10.6084/m9.figshare.285097.

Declarations

Ethics approval and consent to participate
This study's material collections and experimental research complied with relevant institutional, national, and international guidelines and legislation. No specific permissions or licenses were required. All voucher specimens were morphologically authenticated by Prof. Liang Zhao and Prof. Bao-lin Guo, and deposited in the Herbarium of Northwest A&F University (WUK) and the Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences (IMPD) for permanent preservation.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]College of Life Sciences, Northwest A&F University, Yangling 712100, China
[2]Herbarium of Northwest A&F University, Yangling 712100, China
[3]College of Horticulture, Northwest A&F University, Yangling 712100, China
[4]Department of Biological Sciences, Northern Arizona University, 617 S. Beaver St., Biological Sciences Room 227, PO Box 5640, Flagstaff, AZ 86011, USA
[5]Institut de Systématique Evolution Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, 57 rue Cuvier CP39, Paris 75005, France

Received: 5 March 2025 / Accepted: 10 September 2025
Published online: 17 October 2025

References

1. Lumbsch HT, Leavitt SD. Goodbye morphology? A paradigm shift in the delimitation of species in lichenized fungi. Fungal Divers. 2011;50:59–72.
2. Sokoloff DD, Degtjareva GV, Valiejo-Roman CM, Severova EE, Barinova S, Chepinoga VV, et al. Kazakhstan has an unexpected diversity of medicinal plants of the genus *Acorus* (Acoraceae) and could be a cradle of the triploid species *A. calamus*. Plants. 2024;13:1978.
3. Wu GA, Terol J, Ibanez V, López-García A, Pérez-Román E, Borredá C, et al. Genomics of the origin and evolution of *Citrus*. Nature. 2018;554:311–6.
4. Xiong C, Huang Y, Li ZL, Wu L, Liu ZG, Zhu WJ, et al. Comparative chloroplast genomics reveals the phylogeny and the adaptive evolution of *Begonia* in China. BMC Genomics. 2023;24:648.
5. Chen JT, Lidén M, Huang XH, Zhang L, Zhang XJ, Kuang TH, et al. An updated classification for the hyper-diverse genus *Corydalis* (Papaveraceae: Fumarioideae) based on phylogenomic and morphological evidence. J Integr Plant Biol. 2023;65:2138–56.
6. Mo ZQ, Fu CN, Zhu MS, Milne RI, Yang JB, Cai J, et al. Resolution, conflict and rate shifts: insights from a densely sampled plastome phylogeny for *Rhododendron* (Ericaceae). Ann Bot. 2022;130:687–701.
7. Lourenço-Lopes C, Fraga-Corral M, Jimenez-Lopez C, Carpena M, Pereira AG, Garcia-Oliveira P, et al. Biological action mechanisms of Fucoxanthin extracted from algae for application in food and cosmetic industries. Trends Food Sci Technol. 2021;112:711–22.
8. Mahima K, Sunil Kumar KN, Rakhesh KV, Rajeswaran PS, Sharma A, Sathishkumar R. Advancements and future prospective of DNA barcodes in the herbal drug industry. Front Pharmacol. 2022;13:947512.
9. Seethapathy GS, Raclariu-Manolica AC, Anmarkrud JA, Wangensteen H, de Boer HJ. DNA metabarcoding authentication of ayurvedic herbal products on the European market raises concerns of quality and fidelity. Front Plant Sci. 2019;10:68.
10. Michel M, Eldridge AL, Hartmann C, Klassen P, Ingram J, Meijer GW. Benefits and challenges of food processing in the context of food systems, value chains and sustainable development goals. Trends Food Sci Technol. 2024;153:104703.
11. Bucklin A, Steinke D, Blanco-Bercial L. DNA barcoding of marine metazoa. Annu Rev Mar Sci. 2011;3:471–508.
12. Padilla-García N, Rojas-Andrés BM, López-González N, Castro M, Castro S, Loureiro J, et al. The challenge of species delimitation in the diploid-polyploid complex *Veronica* subsection *Pentasepalae*. Mol Phylogenet Evol. 2018;119:196–209.
13. Wagner F, Härtl S, Vogt R, Oberprieler C. Fix me another marguerite! Species delimitation in a group of intensively hybridizing lineages of ox-eye daisies (*Leucanthemum* mill., Compositae-Anthemideae). Mol Ecol. 2017;26:4260–83.
14. China Plant BOL Group. A DNA barcode for land plants. Proc Natl Acad Sci USA. 2009;106:12794–7.
15. China Plant BOL Group, Li DZ, Gao LM, Li HT, Wang H, Ge XJ, et al. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proc Natl Acad Sci U S A. 2011;108:19641–6.
16. Chen SL, Yao H, Han JP, Liu C, Song JY, Shi LC, et al. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS ONE. 2010;5:e8613.
17. Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. Identification of birds through DNA barcodes. PLoS Biol. 2004;2:e312.
18. Hollingsworth PM, Graham SW, Little DP. Choosing and using a plant DNA barcode. PLoS ONE. 2011;6:e19254.
19. Ashfaq M, Asif M, Anjum ZI, Zafar Y. Evaluating the capacity of plant DNA barcodes to discriminate species of cotton (*Gossypium*: Malvaceae). Mol Ecol Resour. 2013;13:573–82.
20. Rubinoff D, Cameron S, Will K. Are plant DNA barcodes a search for the holy grail?? Trends Ecol Evol. 2006;21:1–2.
21. Hollingsworth PM, Li DZ, van der Bank M, Twyford AD. Telling plant species apart with DNA: from barcodes to genomes. Philos Trans R Soc Lond B Biol Sci. 2016;371:20150338.
22. Li XW, Yang Y, Henry RJ, Rossetto M, Wang YT, Chen SL. Plant DNA barcoding: from gene to genome. Biol Rev. 2015;90:157–66.
23. Ślipiko M, Myszczyński K, Buczkowska K, Bączkiewicz A, Szczecińska M, Sawicki J. Molecular delimitation of European leafy liverworts of the genus *Calypogeia* based on plastid super-barcodes. BMC Plant Biol. 2020;20:243.
24. Yu XQ, Jiang YZ, Folk RA, Zhao JL, Fu CN, Fang L, et al. Species discrimination in *Schima* (Theaceae): next-generation super-barcodes meet evolutionary complexity. Mol Ecol Resour. 2022;22:3161–75.
25. Du YP, Bi Y, Yang FP, Zhang MF, Chen XQ, Xue J, et al. Complete chloroplast genome sequences of *Lilium*: insights into evolutionary dynamics and phylogenetic analyses. Sci Rep. 2017;7:5751.
26. Song BN, Liu CK, Zhao AQ, Tian RM, Xie DF, Xiao YL, et al. Phylogeny and diversification of genus *Sanicula* L. (Apiaceae): novel insights from plastid phylogenomic analyses. BMC Plant Biol. 2024;24:70.

27. Liu YY, Cao JL, Kan SL, Wang PH, Wang JL, Cao YN, et al. Phylogenomic analyses sheds new light on the phylogeny and diversification of *Corydalis* DC. in Himalaya-Hengduan mountains and adjacent regions. Mol Phylogenet Evol. 2024;193:108023.
28. Zimmer EA, Wen J. Reprint of: using nuclear gene data for plant phylogenetics: progress and prospects. Mol Phylogenet Evol. 2013;66:539–50.
29. Zimmer EA, Wen J. Using nuclear gene data for plant phylogenetics: progress and prospects Ⅱ. Next-gen approaches. J Syst Evol. 2015;53:371–9.
30. McLay TGB, Fowler RM, Fahey PS, Murphy DJ, Udovicic F, Cantrill DJ, et al. Phylogenomics reveals extreme gene tree discordance in a lineage of dominant trees: hybridization, introgression, and incomplete lineage sorting blur deep evolutionary relationships despite clear species groupings in *Eucalyptus* subgenus *Eudesmia*. Mol Phylogenet Evol. 2023;187:107869.
31. Strand AE, Leebens-Mack J, Milligan BG. Nuclear DNA-based markers for plant evolutionary biology. Mol Ecol. 1997;6:113–8.
32. Saint KM, Austin CC, Donnellan SC, Hutchinson MN. *C-mos*, a nuclear marker useful for squamate phylogenetic analysis. Mol Phylogenet Evol. 1998;10:259–63.
33. Schlüter PM, Kohl G, Stuessy TF, Paulus HF. A screen of low-copy nuclear genes reveals the *LFY* gene as phylogenetically informative in closely related species of orchids (*Ophrys*). Taxon. 2007;56:493–504.
34. Belinky F, Szitenberg A, Goldfarb I, Feldstein T, Wörheide G, Ilan M, et al. ALG11-a new variable DNA marker for sponge phylogeny: comparison of phylogenetic performances with the 18S rDNA and the COI gene. Mol Phylogenet Evol. 2012;63:702–13.
35. Han FM, Peng Y, Xu LJ, Xiao PG. Identification, characterization, and utilization of single copy genes in 29 angiosperm genomes. BMC Genomics. 2014;15:504.
36. Su N, Hodel RGJ, Wang X, Wang JR, Xie SY, Gui CX, et al. Molecular phylogeny and inflorescence evolution of *Prunus* (Rosaceae) based on RAD-seq and genome skimming analyses. Plant Divers. 2023;45:397–408.
37. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10:57–63.
38. Cheon S, Zhang J, Park C. Is phylotranscriptomics as reliable as phylogenomics? Mol Biol Evol. 2020;37:3672–83.
39. Wang XQ, Ye XY, Zhao L, Li DZ, Guo ZH, Zhuang HF. Genome-wide RAD sequencing data provide unprecedented resolution of the phylogeny of temperate bamboos (Poaceae: Bambusoideae). Sci Rep. 2017;7:11546.
40. Dodsworth S, Pokorny L, Johnson MG, Kim JT, Maurin O, Wickett NJ, et al. Hyb-seq for flowering plant systematics. Trends Plant Sci. 2019;24:887–91.
41. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, et al. Target-enrichment strategies for next-generation sequencing. Nat Methods. 2010;7:111–8.
42. Liu BB, Ma ZY, Ren C, Hodel RGJ, Sun M, Liu XQ, et al. Capturing single-copy nuclear genes, organellar genomes, and nuclear ribosomal DNA from deep genome skimming data for plant phylogenetics: a case study in vitaceae. J Syst Evol. 2021;59:1124–38.
43. Liu BB, Ren C, Kwak M, Hodel RGJ, Xu C, He J, et al. Phylogenomic conflict analyses in the Apple genus *Malus* s.l. reveal widespread hybridization and allopolyploidy driving diversification, with insights into the complex biogeographic history in the Northern hemisphere. J Integr Plant Biol. 2022;64:1020–43.
44. Coissac E, Hollingsworth PM, Lavergne S, Taberlet P. From barcodes to genomes: extending the concept of DNA barcoding. Mol Ecol. 2016;25:1423–8.
45. Eberle J, Ahrens D, Mayer C, Niehuis O, Misof B. A plea for standardized nuclear markers in metazoan DNA taxonomy. Trends Ecol Evol. 2020;35:336–45.
46. Zhang L, Huang YW, Huang JL, Ya JD, Zhe MQ, Zeng CX, et al. DNA barcoding of *Cymbidium* by genome skimming: call for next-generation nuclear barcodes. Mol Ecol Resour. 2023;23:424–39.
47. Chase MW, Fay MF. Barcoding of plants and fungi. Science. 2009;325:682–3.
48. Salis R, Sunde J, Gubonin N, Franzén M, Forsman A. Performance of DNA metabarcoding, standard barcoding and morphological approaches in the circumscription of insect biodiversity. Mol Ecol Resour. 2024;00:e14018.
49. Yuan YD, Liu X, Wang JP, Zhang JC. Morphological and microscopic circumscription of three major medicinal *Dendrobium* species in Ta-pieh mountains area. Microsc Res Tech. 2019;82:483–93.
50. Duminil J, Michele MD. Plant species delimitation: a comparison of morphological and molecular markers. Plant Biosyst. 2009;143:528–42.
51. Ebach MC, Holdrege C. DNA barcoding is no substitute for taxonomy. Nature. 2005;434:697.
52. Endress PK. Morphology and angiosperm systematics in the molecular era. Bot Rev. 2003;68:545–70.
53. Guo BL, He SZ, Zhong GY, Xiao PG. Two new species of *Epimedium* (Berberidaceae) from China. J Syst Evol. 2007;45:813–21.
54. Loconte H. Berberidaceae. In: Kubitzki K, Rohwer JG, Bittrich V, editors. The families and genera of vascular plants II. Berlin: Springer; 1993. pp. 147–52.
55. Ma HP, He XR, Yang Y, Li MX, Hao DJ, Jia ZP. The genus *Epimedium*: an ethnopharmacological and phytochemical review. J Ethnopharmacol. 2011;134:519–41.
56. Stearn WT. The genus Epimedium and other herbaceous Berberidaceae. Portland: Timber; 2002.
57. Ying TS, Boufford DE, Brach AR. *Epimedium* L. Flora of China. Beijing: Science & St. Louis: Missouri Botanical Garden Press;; 2011. pp. 787–99.
58. Zhang C, Meng R, Meng Y, Guo BL, Liu QR, Nie ZL. Parallel evolution, atavism, and extensive introgression explain the radiation of *Epimedium* sect. *Diphyllon* (Berberidaceae) in southern east Asia. Front Plant Sci. 2023;14:1234148.
59. Xie SY, Hou XQ, Zhang XH. Are the Spurs more complex than other petal types in *Epimedium*? Evidence from development, micromorphology, and nectary structure. Flora. 2022;293:152101.
60. Zhao X, Hou QZ, Du MN, Zhang H, Jia LY, Zhang ZH, et al. Micromorphological leaf epidermal traits as potential taxonomic markers for infrageneric classification of *Oxytropis* (Fabaceae). Phytokeys. 2022;201:51–76.
61. Guo MY, Ren L, Xu YQ, Liao BS, Song JY, Li Y, et al. Development of plastid genomic resources for discrimination and classification of *Epimedium Wushanense* (Berberidaceae). Int J Mol Sci. 2019;20:4003.
62. Linnaeus C. Species plantarum. Stockholm: Laurentius Salvius; 1753.
63. Zhang YJ, Yang LL, Chen JJ, Sun W, Wang Y. Taxonomic and phylogenetic analysis of *Epimedium* L. based on amplified fragment length polymorphisms. Sci Hortic. 2014;170:284–92.
64. Zhang YJ, Du LW, Liu A, Chen JJ, Wu L, Hu WM, et al. The complete chloroplast genome sequences of five *Epimedium* species: lights into phylogenetic and taxonomic analyses. Front Plant Sci. 2016;7:306.
65. Liu JQ. The integrative species concept and species on the speciation way. Biodivers Sci. 2016;24(9):1004–8.
66. Jiang Y, Ding CB, Zhang L, Yang RW, Zhou YH, Tang L. Identification of the genus *Epimedium* with DNA barcodes. J Med Plants Res. 2011;5:6413–7.
67. Guo MY, Xu YQ, Ren L, He SZ, Pang XH. A systematic study on DNA barcoding of medicinally important genus *Epimedium* L. Genes (Basel). 2018;9:637. Berberidaceae.
68. Guo MY, Pang XH, Xu YQ, Jiang WJ, Liao BS, Yu JS, et al. Plastid genome data provide new insights into the phylogeny and evolution of the genus *Epimedium*. J Adv Res. 2021;36:175–85.
69. Lone SA, Hassan QP, Gupta S. Development of DNA barcode for rapid identification of *Epimedium elatum* (Morren & Decne) from Northwestern Himalayas in India. J Appl Res Med Aromat Plants. 2019;13:100205.
70. Liu Q, Li XY, Li MZ, Xu WK, Schwarzacher T, Heslop-Harrison JS. Comparative chloroplast genome analyses of *Avena*: insights into evolutionary dynamics and phylogeny. BMC Plant Biol. 2020;20:406.
71. Guo YY, Yang JX, Bai MZ, Zhang GQ, Liu ZJ. The chloroplast genome evolution of Venus slipper (*Paphiopedilum*): IR expansion, SSC contraction, and highly rearranged SSC regions. BMC Plant Biol. 2021;21:248.
72. Moghaddam M, Ohta A, Shimizu M, Terauchi R, Kazempour-Osaloo S. The complete chloroplast genome of *Onobrychis Gaubae* (Fabaceae-Papilionoideae): comparative analysis with related IR-lacking clade species. BMC Plant Biol. 2022;22:75.
73. Chery M, Drouard L. Plant tRNA functions beyond their major role in translation. J Exp Bot. 2023;74:2352–63.
74. Noureen M, Tada I, Kawashima T, Arita M. Rearrangement analysis of multiple bacterial genomes. BMC Bioinformatics. 2019;20(Suppl 23):631.
75. Chen JL, Wang F, Zhao Z, Li MH, Liu ZJ, Peng DH. Complete Chloroplast genomes and comparative analyses of three *Paraphalaenopsis* (Aeridinae, Orchidaceae) species. Int J Mol Sci. 2023;24:11167.
76. Shi ZY, Zhao WQ, Li ZA, Kang DR, Ai PH, Ding HX, et al. Development and validation of SSR markers related to flower color based on full-length transcriptome sequencing in *Chrysanthemum*. Sci Rep. 2022;12:22310.
77. Zhang L, Zhang ED, Wei YQ, Zheng GQ. Phylogenetic analysis and divergence time estimation of *Lycium* species in China based on the chloroplast genomes. BMC Genomics. 2024;25:569.
78. Antonelli A, Clarkson JJ, Kainulainen K, Maurin O, Brewer GE, Davis AP, et al. Settling a family feud: a high-level phylogenomic framework for the Gentianales based on 353 nuclear genes and partial plastomes. Am J Bot. 2021;108:1143–65.

79. Duan HN, Jiang YZ, Yang JB, Cai J, Zhao JL, Li L, et al. Skmer approach improves species discrimination in taxonomically problematic genus *Schima* (Theaceae). Plant Divers. 2024;46:713–22.
80. Ruhsam M, Rai HS, Mathews S, Ross TG, Graham SW, Raubeson LA, et al. Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*? Mol Ecol Resour. 2015;15:1067–78.
81. Pillon Y, Johansen J, Sakishima T, Chamala S, Barbazuk WB, Roalson EH, et al. Potential use of low-copy nuclear genes in DNA barcoding: a comparison with plastid genes in two Hawaiian plant radiations. BMC Evol Biol. 2013;13:35.
82. Wang K, Lenstra JA, Liu L, Hu QJ, Ma T, Qiu Q, et al. Incomplete lineage sorting rather than hybridization explains the inconsistent phylogeny of the wisent. Commun Biol. 2018;1(1):169.
83. Li QJ, Wang X, Wang JR, Su N, Zhang L, Ma YP, et al. Efficient circumscription of *Pulsatilla* (Ranunculaceae) using DNA barcodes and Micro-Morphological characters. Front Plant Sci. 2019;10:1196.
84. Tong RC, Gui CX, Zhang Y, Su N, Hou XQ, Liu M, et al. Phylogenomics, plastome structure and species circumscription in *Mahonia* (Berberidaceae). BMC Genomics. 2022;23:766.
85. Edelaar P, Roques S, Hobson EA, Gonçalves da Silva A, Avery ML, Russello MA, et al. Shared genetic diversity across the global invasive range of the monk parakeet suggests a common restricted geographic origin and the possibility of convergent selection. Mol Ecol. 2015;24:2164–76.
86. Susko E, Roger AJ. Long branch attraction biases in phylogenetics. Syst Biol. 2021;70:838–43.
87. Bowler DE, Boyd RJ, Callaghan CT, Robinson RA, Isaac NJ, Pocock MJ. Treating gaps and biases in biodiversity data as a missing data problem. Biol Rev Camb Philos Soc. 2025;100:50–67.
88. Pezzi PH, Wheeler LC, Freitas LB, Smith SD. Incomplete lineage sorting and hybridization underlie tree discordance in *Petunia* and related genera (Petunieae, Solanaceae). Mol Phylogenet Evol. 2024;198:108136.
89. Li X, Wei GM, El-Kassaby YA, Fang YM. Hybridization and introgression in sympatric and allopatric populations of four oak species. BMC Plant Biol. 2021;21:266.
90. Hending D. Cryptic species conservation: a review. Biol Rev Camb Philos Soc. 2025;100:258–74.
91. Azizi MMF, Lau HY, Abu-Bakar N. Integration of advanced technologies for plant variety and cultivar identification. J Biosci. 2021;46:91.
92. Liu J, Milne RI, Möller M, Zhu GF, Ye LG, Luo YH, et al. Integrating a comprehensive DNA barcode reference library with a global map of yews (*Taxus* L.) for forensic identification. Mol Ecol Resour. 2018;18:1115–31.
93. Denay G, Preckel L, Petersen H, Pietsch K, Wöhlke A, Brünen-Nieweler C. Benchmarking and validation of a bioinformatics workflow for meat species identification using 16S rDNA metabarcoding. Foods. 2023;12:968.
94. Shen YY, Chen X, Murphy RW. Assessing DNA barcoding as a tool for species identification and data quality control. PLoS ONE. 2013;8:e57125.
95. Fikáček M, Hu FS, Le MH, Huang JP. Can immature stages be ignored in studies of forest leaf litter arthropod diversity? A test using Oxford nanopore DNA barcoding. Insect Conserv Divers. 2024;17:31–50.
96. Teufel A, Krupp M, Weinmann A, Galle PR. Current bioinformatics tools in genomic biomedical research (Review). Int J Mol Med. 2006;17:967–73.
97. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry. 1987;19:11–5.
98. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014;30:2114–20.
99. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, et al. Getorganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21:241.
100. Qu XJ, Moore MJ, Li DZ, Yi TS. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods. 2019;15:50.
101. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq-Versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45:6–11.
102. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28:1647–9.
103. Lohse M, Drechsel O, Kahlau S, Bock R. Organellargenomedraw-a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res. 2013;41:575–81.
104. Pertea G, Pertea M. GFF utilities: GffRead and gffcompare. F1000Res. 2020;9:304.
105. Chamala S, García N, Godden GT, Krishnakumar V, Jordon-Thaden IE, De Smet R, et al. Markerminer 1.0: a new application for phylogenetic marker development using angiosperm transcriptomes. Appl Plant Sci. 2015;3:1400115.
106. Babraham Institute. FastQC. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 20 May 2024.
107. Johnson MG, Gardner EM, Liu Y, Medina R, Goffinet B, Shaw AJ, et al. Hybpiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. Appl Plant Sci. 2016;4:1600016.
108. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
109. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77.
110. Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. 2005;6:31.
111. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
112. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.
113. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. Mrbayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61:539–42.
114. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. Vista: computational tools for comparative genomics. Nucleic Acids Res. 2004;32:273–9.
115. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017;34:3299–302.
116. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33:2583–5.
117. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27:573–80.
118. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. Reputer: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29:4633–42.
119. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–3.
120. Su S, Zhao L, Ren Y, Zhang XH. Diversity of petals in berberidaceae: development, micromorphology, and structure of floral nectaries. Protoplasma. 2021;258:905–22.
121. Zhang HM, Wang H, Wei J, Chen XP, Sun MJ, Ouyang HZ, et al. Comparison of the active compositions between raw and processed *Epimedium* from different species. Molecules. 2018;23:1656.
122. Li ZN, Zeng SH, Li YB, Li MY, Souer E. Leaf-like sepals induced by ectopic expression of a SHORT VEGETATIVE PHASE (*SVP*)-Like MADS-Box gene from the basal eudicot *Epimedium sagittatum*. Front Plant Sci. 2016;7:1461.
123. Zhang WJ, Chen HX, Wang ZS, Lan GS, Zhang LK. Comparative studies on antioxidant activities of extracts and fractions from the leaves and stem of *Epimedium Koreanum* Nakai. J Food Sci Technol. 2013;50:1122–9.
124. Yu DY, Huang RQ, Yu SX, Liang Q, Wang Y, Dang HS, et al. Construction of the first high-density genetic linkage map and QTL mapping of flavonoid and leaf-size related traits in *Epimedium*. BMC Plant Biol. 2023;23:278.
125. Bayat N, Attar F, Sotoodeh A. Pollen and leaf micromorphological characteristics of spiny almonds (*Prunus* subgenus *Amygdalus*) in Iran. Microsc Res Tech. 2025;88:575–94.
126. Collins TJ. Imagej for microscopy. Biotechniques. 2007;43(1 Suppl):25–30.
127. Kaiser HF. The varimax criterion for analytic rotation in factor analysis. Psychometrika. 1958;23:187–200.

Publisher's Note