BMC Plant Biology

RESEARCH

Open Access

Plastome-based phylogeny and historical biogeography of *Erythronium* L.

Yu Gyeom Kim[1], Joonhyung Jung[2] and Joo-Hwan Kim[1*]

Abstract

The genus *Erythronium* L. (Liliaceae) comprises approximately 30 species of spring-flowering perennial herbs that are widely distributed in temperate areas across the Northern Hemisphere. Although species within the genus exhibit considerable morphological similarity, their phylogenetic relationships remain insufficiently resolved. Owing to the limited scope of available datasets and insufficient taxon sampling, previous molecular studies have struggled to fully clarify phylogenetic relationships within the genus. This study presents a plastome-based analysis of phylogeny and historical biogeography in *Erythronium*, incorporating ten newly sequenced plastomes. All plastomes retained the typical quadripartite structure and presented low overall nucleotide diversity, with *inf*A consistently being pseudogenized across all the taxa. Phylogenetic analyses based on 78 plastid protein-coding genes resolved major infrageneric lineages within *Erythronium*, corresponding to three well-supported, geographically disjunct clades: Eastern North America, Western North America, and Eurasia. *Amana* Honda was placed as a strongly supported sister lineage in all analyses. Molecular dating via BEAST estimated that the crown age of *Erythronium* was 16.7 million years ago (Mya), with subsequent divergences of the Eurasian (8.71 Mya), Eastern North American (5.80 Mya), and Western North American (3.35 Mya) clades occurring after the late Miocene. Ancestral area reconstruction via the BBM indicated East Asia as the most likely origin. The geographically disjunct pattern of *Erythronium* diversification appears to have resulted from a combination of Miocene-era dispersal via the Bering Land Bridge, global climate cooling, and continental-scale geographic isolation. These results provide new insights into the evolutionary history and biogeographic diversification of *Erythronium*.

Keywords *Erythronium*, Liliaceae, Plastome, Next-generation sequencing, Phylogenomics, Biogeography

Background

Species in the genus *Erythronium* L. are commonly known by various names, such as dogtooth violet, trout lily, or fawn lily, which are derived from the tooth-like shape of their bulbs. It is a genus of the tribe Tulipeae (subfamily Lilioideae; family Liliaceae). The genus *Erythronium* is widely distributed throughout mid-latitude (30°N–60°N) temperate zones of the Northern Hemisphere. Species of *Erythronium* occur in North America, Europe, Eastern Siberia, Central Asia, and Eastern Asia, with 23 species occurring in Western North America, accounting for the majority of *Erythronium* species [4]. The flowers of most species of *Erythronium* have solitary inflorescences, but some Western North American species have cymose inflorescences. The flowers consisted of six free lanceolate tepals with various patterns of shading, spots and lines. All species of *Erythronium* have a cone-shaped tunicate bulb. This bulb produces one to three stolons. From these stolons, new

*Correspondence:
Joo-Hwan Kim
kimjh2009@gachon.ac.kr
[1]Department of Life Science, Gachon University, Seongnam 13120, Republic of Korea
[2]Division of Taxonomy, Honam National Institute of Biological Resources, 99, Gohadoan-gil, Mokpo-si, Jeollanam-do 58762, Republic of Korea

bulbs are produced. The genus *Erythronium* has a pair of ovate leaves [4]. Most species have characteristic dark or light-colored spots on their leaves. Some researchers have used this feature as a diagnostic characteristic, but the spots are highly variable between individuals and often disappear as the plant grows [9]. The tribe Tulipeae consists of four genera. In this tribe, the genus *Erythronium* is distinguished in that the position of the flower is almost always nodding and oriented toward the ground. Several species of *Erythronium* share the same habitat, and they generally have similar morphological characteristics. For example, yellow-flowered *E. americanum* Ker Gawl. and *E. umbilicatum* C.R.Parks & Hardin, both of which are found in Eastern North America, share a very wide range of habitats and are very similar in morphotype [18]. Additionally, Lin [24], argued that the differences in petal color and pattern were due to mutations that occurred during gene duplication rather than genetic inheritance. This overlap of habitats is the reason why hybrids and intermediate traits are so common in North America [34]. Asia has fewer species than the Americas do and has even been classified as European *E. dens-canis* L. by some researchers [27]. However, extensive morphological similarity and plasticity among species, especially in North America, have complicated species delimitation and limited the resolution of infrageneric relationships based solely on morphology.

Most plants in the world are autotrophs, and for them, chloroplasts are among the key organelles essential for survival and photosynthesis. The chloroplast genome (plastome) is mostly composed of a large single-copy (LSC), a small single-copy (SSC), and two inverted repeat (IR) regions [33]. Many molecular phylogenetic studies of plants use chloroplast DNA sequences. Chloroplast DNA does not undergo genetic recombination due to paternal inheritance and therefore has a high degree of conservation of protein-coding sequences (CDSs) within the sequence. Therefore, structural abnormalities such as deletions, translocations, inversions, and duplications in the four main sections of the chloroplast genome sequence provide insight into molecular phylogenetic studies. The chloroplast genome of plants is typically 115–165 kb in size [33]. Some researchers have developed sequencing tools to analyze such a large number of sequences. With the advent of next-generation sequencing (NGS) technology, which has many advantages, such as low cost, time savings, and high data throughput, compared with conventional methods, it has been used in many studies to maximize the efficiency of genome analysis. Owing to their excellent accessibility, the completed plastome sequences are uploaded to GenBank at the National Center for Biotechnology Information (NCBI) and have become the basis for analysis by researchers around the world.

North America and Eurasia exhibit many botanical similarities despite their great distance from each other. Among many plants, a few genera have disjunct distributions in both regions [43]. The genus *Erythronium* shows a distinctly disjunct distribution between North America and Eurasia, so this could be a very important key to understanding the biogeographical history of Liliales.

Many studies have performed phylogenetic analyses of Liliaceae. However, phylogenetic analyses of the genus *Erythronium* have shown that taxa within each clade have weak support values in common [10, 22]. Allen et al. [1], analyzed the phylogeny of 16 species of *Erythronium* via the plastidial *mat*K gene from the chloroplast genome and the ITS region of nuclear ribosomal DNA. They reported that *Erythronium* is divided into three well-resolved clades, each corresponding to a distinct geographic region: Eastern North America, Western North America, and Eurasia. However, low support values in the analyses with ITS sequences and polytomy in each clade in the analyses with *mat*K indicated uncertainty in the phylogenetic position of each species. Clennett et al., [10] investigated the phylogenetic relationships between *Erythronium* and its putative close relative *Amana* Honda via chloroplast markers (*rps*16 intron and the 5′ region of the *trn*K intron excluding *mat*K) and nuclear ITS sequences. While their results supported the division of *Erythronium* into three geographically structured clades, species boundaries within the Western North American clade remained unclear, suggesting limited resolution among taxa in that region. However, this study also revealed lower support values in analyses using ITS sequences. These low support values make each cluster between species less clear. This result suggests that increasing the support value of phylogenetic studies by identifying genes that are useful for molecular taxonomic phylogenetic studies of the genus *Erythronium* is necessary. Although much research has been conducted, the phylogenetic structure among species within each clade of the genus *Erythronium* remains unresolved. Identifying the phylogenetic position of a species provides fundamental research data not only for phylogenetic taxonomy but also for many other fields of study, including evolutionary biology, plant physiology, and biogeography. Molecular phylogenetic analyses of *Erythronium* have been conducted by several researchers [48].

Although only 11 of the approximately 33 currently recognized species of *Erythronium* were included in this study, the sampling was designed to reflect the major geographically structured lineages previously proposed within the genus, rather than to exhaustively represent species-level diversity. The present dataset includes representative taxa from Eurasia, Eastern North America, and Western North America, while acknowledging that taxon coverage within the species-rich Western North

American lineage remains incomplete. Accordingly, the results presented here are intended to provide a plastome-based framework for examining major infrageneric relationships within *Erythronium* in a well-defined intergeneric context, rather than a comprehensive species-level phylogeny.

This study aims to fill these gaps by leveraging plastome-scale data to produce a robust phylogenetic framework, infer divergence times, and reconstruct ancestral areas. Our ultimate goal is to elucidate the biogeographic origin of *Erythronium*, clarify its interspecific relationships, and provide new insights into the evolutionary dynamics of temperate monocot lineages. These results are expected to contribute to the taxonomic resolution of the genus and improve our understanding of historical migration and diversification processes in Northern Hemisphere floras.

Table 1 List of samples used for DNA extraction

Taxa	Genus	Collection site	Source/ Voucher
Erythronium albidum	*Erythronium*	Turkey Run Trailhead, McLean, VA 22,101, United States	YG240306
Erythronium americanum	*Erythronium*	Turkey Run Trailhead, McLean, VA 22,101, United States	YG240301
Erythronium californicum	*Erythronium*	Route de Nimes, Saint-Gilles 30,800 Saint-Gilles, France	Château Pérouse Plant ID : 10,814
Erythronium caucasicum	*Erythronium*	Route de Nimes, Saint-Gilles 30,800 Saint-Gilles, France	Château Pérouse Plant ID : 10,830
Erythronium citrinum	*Erythronium*	Bot. Tuinen, Domeinbeheer, Arboretum Bokrijk, B-3600 GENK, BEL	Utrecht University Botanic Gardens 2021BL00265
Erythronium dens-canis	*Erythronium*	Jardin Botanique, CH-1007 LAUSANNE, CHE	Utrecht University Botanic Gardens 1998BL01374
Erythronium grandiflorum	*Erythronium*	U.S.A., Colorado, Grand, Arapaho National Forest	T. G. Lammers 14,570
Erythronium multiscapideum	*Erythronium*	Route de Nimes, Saint-Gilles 30,800 Saint-Gilles, France	Château Pérouse Plant ID : 12,822
Erythronium revolutum	*Erythronium*	Budapestlaan 17, 3584 CD Utrecht, The Netherlands	Utrecht University Botanic Gardens 2009BL01500
Erythronium tuolumnense	*Erythronium*	Botanical Garden, S-413 19 Göteborg, SWE	Utrecht University Botanic Gardens 2005BL00880

Materials and methods

Taxon sampling and DNA extraction

In this study, we collected 10 species of the genus *Erythronium* from natural habitats and botanical gardens located outside of Korea. Fresh leaf materials were immediately desiccated with silica gel. Voucher samples were prepared for most of the samples. Some samples were stored in the Gachon University Herbarium (GCU), and each sample was assigned a specific accession number. The others were stored in collaborating herbaria in each country. DNA extraction followed a modified CTAB method (2× concentration), as described by Doyle and Doyle [13]. The DNA concentration and purity were assessed via a Biospec-nano spectrophotometer (Shimadzu), and the integrity was verified via 1.5% agarose gel electrophoresis.

Next-generation sequencing, plastome assembly, and annotation

We conducted NGS on the DNA extracted from the 10 taxa of the genus *Erythronium* via the Illumina MiSeq platform (Illumina, Seoul, Korea) (Table 1). The raw reads were initially assembled *de novo* via the GetOrganelle toolkit [19]. While most taxa yielded complete plastomes through this approach, some species resulted in incomplete assemblies. For those cases, we employed *Erythronium japonicum* Decne. as a reference plastome and applied both the 'Map to Reference' and *de novo* assembly methods in Geneious Prime 2022.0.2 [21] to improve completeness and accuracy. The quality of all final assemblies was evaluated by examining coverage depth and mapping consistency. Gene annotation was conducted with GeSeq [41], and all tRNA genes were revalidated via tRNAscan-SE [7]. Finally, complete plastomes were visualized via OGDraw [17].

Comparative plastome analysis

For comparative analyses, we used a total of 11 *Erythronium* plastomes, including the 10 newly sequenced plastomes in this study. A total of 11 *Erythronium* species were selected for plastome sequencing and phylogenetic analysis. Although this represents a subset of the approximately 33 known species in the genus, most *Erythronium* species occur in Western North America and exhibit relatively low morphological and genetic divergence. Accordingly, the selected taxa were chosen to encompass the known geographical and genetic breadth of the genus, thereby ensuring representative coverage in downstream analyses. Structural comparisons were conducted with respect to genome size, gene content, gene order, and GC content. All sequences were aligned

via the MUSCLE algorithm [15], and sequence similarity and divergence were visualized via the mVISTA platform in LAGAN mode [5]. The plastome of *Amana edulis* (Miq.) Honda was used as a reference genome due to its close phylogenetic relationship. Furthermore, boundaries between the IR/LSC and IR/SSC regions were examined across taxa and visualized via IRScope [3], enabling comparative evaluation of plastome structural variation within the genus.

In addition, we assessed the nucleotide diversity (*Pi*) of protein-coding sequences, transfer RNA genes, and ribosomal RNA genes across *Erythronium* plastomes. Previously published plastome sequences were incorporated to evaluate the extent of polymorphism within the genus. *Pi* values were calculated via DnaSP v6.12.03 [36], with a window size of 100 bp and a step size of 25 bp applied in the sliding window analysis of sequence divergence.

Phylogenomic analysis

Phylogenetic analyses were conducted via 48 plastome sequences from the family Liliaceae, 38 of which were obtained from the NCBI GenBank database. In total, 11 taxa were designated outgroup species for the phylogenetic tree. A total of 78 protein-coding sequences were extracted from each plastome and aligned via the MUSCLE algorithm implemented in Geneious Prime 2022.0.2. All alignments were manually checked and adjusted where necessary. For maximum likelihood (ML) analysis, the concatenated dataset was partitioned by gene (78 partitions). To assess the effect of partitioning strategy, an additional ML analysis was conducted under an unpartitioned scheme, and the resulting topologies and bootstrap support values were compared (Table S1). Bayesian inference (BI) analysis was performed on the concatenated dataset under a single substitution model.

We used MEGA 11 to select the optimal substitution model on the basis of Akaike's information criterion (AIC) [39]. ML analysis was performed via IQ-TREE Web Server with the GTR + I + G substitution model [42], which was selected via jModelTest v2.1.7 under the Akaike information criterion (AIC). The robustness of the internal branches was evaluated with 10,000 ultrafast bootstrap replicates. BI analysis was conducted in MrBayes v3.2.7 with two independent runs [35], each consisting of four chains and set to run for 5,000,000 generations. Trees were sampled every 1,000 generations, and the first 25% of the samples were discarded as burn-in. Posterior probabilities (PPs) were calculated on the basis of the remaining trees, and consensus trees were visualized via FigTree v1.4.4 [31]. Accordingly, the relatively broad outgroup sampling adopted in this study reflects methodological requirements for divergence-time estimation and ancestral area reconstruction, rather than an effort to address deep phylogenetic relationships within Liliales.

Molecular dating analysis

We performed molecular dating analysis via BEAST v1.10.4 to estimate the divergence times within the genus *Erythronium* [38]. A total of 78 plastid protein-coding sequences were aligned and concatenated, and the alignment was used to generate an XML input file in BEAUti. For divergence-time estimation, the dataset was partitioned by gene, while substitution model parameters were linked among partitions, and branch-specific rate variation was modeled using an uncorrelated lognormal relaxed clock.

We applied the GTR + I + G substitution model, an uncorrelated lognormal relaxed clock, and a Yule speciation prior [14, 16]. Three fossil calibration points were applied as lognormal priors for divergence time estimation, following previously established calibrations in Liliales. The stem node of Ripogonaceae was calibrated at 51 million years ago (Mya) on the basis of leaf microfossils of *Ripogonum tasmanicum* Conran, R.J.Carp. & G.J.Jord. Carpenter et al., [6], Conran [11], . The stem node of *Luzuriaga* Ruiz & Pav. was constrained to 23.2 Mya via fossil leaves of *Luzuriaga peterbannisteri* Conran, Bannister, Mildenh., & D.E.Lee [12]. For the crown node of *Smilax* L., a mean age of 46.5 Mya was applied, which was derived from fossil records spanning the late Paleocene to mid-Eocene [44, 45]. All fossil priors were modeled as lognormal distributions with a mean of 1.0, standard deviation of 1.25, and an offset age corresponding to each fossil calibration point. The root age of the tree was constrained to 90 Mya on the basis of recent molecular clock estimates for Liliales [25]. The MCMC analysis was run for 100 million generations with sampling every 1,000 generations, and the first 10% of the trees were discarded as burn-in. The resulting trees were summarized via TreeAnnotator to obtain a maximum clade credibility (MCC) tree.

To assess the robustness of divergence-time estimates to prior assumptions, we conducted an additional BEAST analysis in which the fixed root-age constraint was replaced with a Normal prior. In the primary analysis, the root age was constrained to 90 Ma, whereas in the alternative analysis a Normal prior was applied to the root age (mean = 90 Ma, SD = 15 Ma). All other model settings, including fossil calibrations, substitution model, relaxed-clock model, tree prior, and MCMC configuration, were kept identical between analyses. Node age estimates from both analyses were compared and are summarized in Table S2.

Ancestral area reconstruction

The eight biogeographic areas were defined a priori based on the geographic distributions of the sampled taxa, floristic regions, and broad-scale climatic patterns, with area boundaries conservatively delineated using specimen collection localities, published distribution records, and the Köppen Climate Classification [23]. Although *Erythronium* species occur in only five of the eight defined areas (Eurasia, Eastern North America, Western North America, East Asia, and Northeast Asia), all eight areas were retained because outgroup taxa are distributed in the remaining regions (Africa, India, and Oceania), ensuring representation of each area in the analysis. For *Erythronium*, each species was assigned to a single biogeographic area, reflecting their geographically discrete and non-overlapping distributions.

To reconstruct ancestral distribution areas and infer patterns of geographic diversification within *Erythronium*, we employed the Bayesian Binary Method (BBM) implemented in RASP v4.3 [46]. BBM was selected because it accounts for phylogenetic uncertainty by integrating over the posterior distribution of trees and provides marginal probabilities for ancestral ranges at each node, offering a robust probabilistic framework well-suited to the objectives of this study. Species ranges were coded based on their entire known distributions across the eight defined areas. All post-burn-in trees from the BEAST analysis were used as input for the BBM analysis, which was run for 500,000 generations with two simultaneous runs of 10 Markov chains each. The analysis employed a fixed-state frequency (Jukes–Cantor) model with equal among-site rate variation. The "Compute Condense" option in RASP was used to generate a consensus tree, which served as the basis for ancestral area reconstruction, with the maximum number of ancestral areas per node set to four.

Results

Plastome features

We newly assembled 10 plastomes of *Erythronium* species. The length of the assembled plastomes varied between 150,647 and 151,827 bp (Fig. 1). The average GC content was 36.6%, which is relatively low compared with that of other members of the Liliaceae. The average total length of the plastomes was 151,035 bp, which is likewise shorter than those of other Liliaceae species (Table 2). Each *Erythronium* plastome analyzed in this study comprised 131 genes in total. Among these genes, 18 were duplicated within the inverted repeat (IR) regions. The remaining 113 unique genes consisted of 78 protein-coding sequences (CDSs), 30 tRNA genes, and four rRNA genes. Among the plastid protein-coding sequences, *inf*A was consistently pseudogenized across all *Erythronium* species, and domain analysis revealed extensive loss of functional regions; this gene was therefore excluded from subsequent analyses, resulting in a final dataset of 78 plastid protein-coding genes used in this study (Table 3).

Plastome assembly validation and reference bias mitigation

All newly assembled plastomes exhibited complete coverage and high sequencing depth, with mean depth exceeding 50× across all taxa (Table S3). Read remapping confirmed uniform coverage across the LSC, SSC, and IR regions, providing clear support for IR boundary positions. In addition, continuous read coverage across sequence termini verified successful circularization of all plastomes and the absence of assembly breakpoints.

Although mean sequencing depth varied among taxa, plastome assemblies were generated primarily using a *de novo* approach implemented in GetOrganelle, thereby minimizing potential reference bias. A reference plastome (*E. japonicum*) was used only at the initial contig identification stage to facilitate plastid sequence recovery and was not involved in genome extension or scaffolding. Final assemblies were independently validated by remapping raw reads to the assembled plastomes.

Comparative chloroplast genome structure and nucleotide diversity

In addition to minor differences in noncoding regions, the plastome sequences of *Erythronium* taxa presented very little overall variation (Fig. 2). No major differences were detected at the boundaries of the quadripartite plastome structure among the species. A notable exception was found in *E. japonicum*, where the start position of the *ndh*F gene was slightly shifted downstream compared with that in other species (Fig. 3).

We calculated *Pi* values for each protein-coding sequence (Fig. 4). The mean *Pi* across all CDS regions was 0.00349. The *ycf*1 gene presented the highest *Pi* value (0.0137), whereas 32 genes presented *Pi* values of 0, indicating extremely low sequence divergence among species. The rRNA and tRNA genes exhibited negligible sequence variation across species, providing limited resolution (Fig. 5).

Phylogenomic analysis

A phylogenomic analysis was conducted via a concatenated dataset comprising 78 plastid protein-coding sequences, totaling 70,324 bp. Three phylogenetic trees were constructed via maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference (BI) methods, and the topology commonly supported across analyses is presented. The ML and BI analyses produced congruent topologies, with most clades showing strong

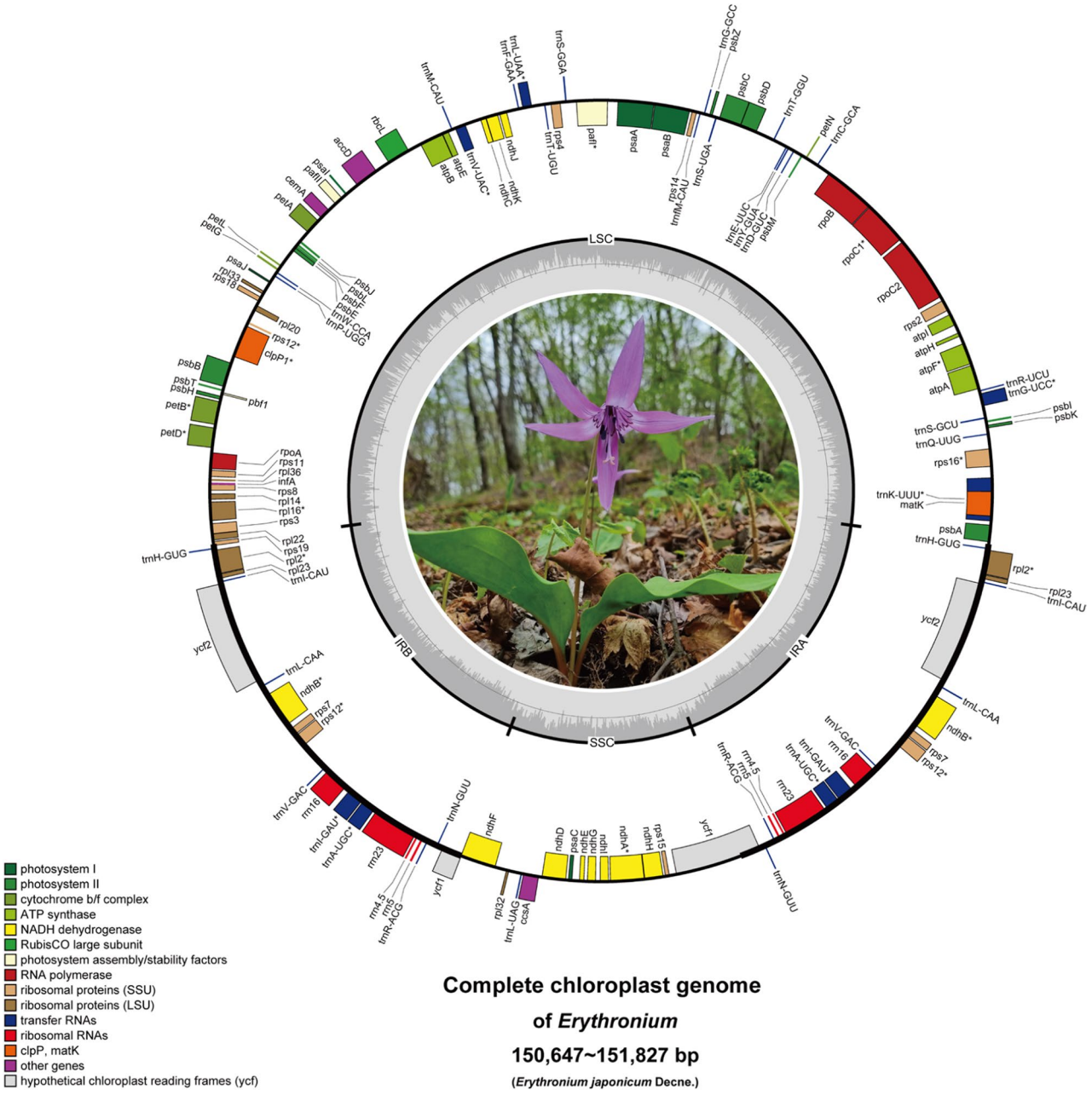

Fig. 1 Representative plastome map of *Erythronium*. Genes are color-coded by functional group. The transcription of genes occurs in a clockwise direction for those positioned on the inner side of the circular map and in a counterclockwise direction for those on the outer side. The GC content is depicted by a central gray circle

support values—100% bootstrap support (ML-BS) and a Bayesian posterior probability (PP) of 1.00.

In contrast, the MP analysis revealed slightly different topologies in certain regions; specifically, the phylogenetic positions of the genera *Tricyrtis* Wall. and *Calochortus* Pursh were reversed compared with those of the ML and BI trees. Aside from this discrepancy, the MP tree largely shared the same topology with the ML and BI analyses and generally presented high bootstrap support values (MP-BS).

Across all analyses, the genus *Erythronium* was clearly resolved into three well-supported clades corresponding to Eurasia, Eastern North America, and Western North America. *Amana* was identified as the sister group to *Erythronium* in all three analyses, with strong support values. This analysis also confirmed that the three clades of *Erythronium* exhibit geographically discontinuous divergence patterns. Comparison between gene-partitioned and unpartitioned ML analyses recovered

Table 2 Summary of plastome features in this study

Taxa	Subfamily	Family	GC content	Length (bp)	GenBank accession number
Erythronium albidum	Lilioideae	Liliaceae	36.5%	151,599	PX122074
Erythronium americanum	Lilioideae	Liliaceae	36.5%	151,114	PX122075
Erythronium californicum	Lilioideae	Liliaceae	36.7%	150,715	PX122076
Erythronium caucasicum	Lilioideae	Liliaceae	36.7%	151,312	PX122077
Erythronium citrinum	Lilioideae	Liliaceae	36.7%	150,695	PX122078
Erythronium dens-canis	Lilioideae	Liliaceae	36.6%	151,333	PX122079
Erythronium grandiflorum	Lilioideae	Liliaceae	36.7%	150,647	PX122080
Erythronium japonicum	Lilioideae	Liliaceae	36.6%	151,827	PX122081
Erythronium multiscapideum	Lilioideae	Liliaceae	36.7%	150,727	PX122082
Erythronium revolutum	Lilioideae	Liliaceae	36.7%	150,651	PX122083
Erythronium tuolumnense	Lilioideae	Liliaceae	36.7%	150,761	PX122084
Amana edulis	Lilioideae	Liliaceae	37.6%	151,459	OL351567.1
Amana erythronioides	Lilioideae	Liliaceae	37.6%	151,699	KY401421.2
Tulipa alberti	Lilioideae	Liliaceae	36.6%	152,382	OR458821.1
Tulipa korolkowii	Lilioideae	Liliaceae	36.6%	152,155	PP693078.1
Tulipa sylvestris	Lilioideae	Liliaceae	36.7%	151,940	MT261172.1
Tulipa uniflora	Lilioideae	Liliaceae	36.5%	152,254	PP861170.1
Gagea triflora	Lilioideae	Liliaceae	37.0%	150,345	MT261157.1
Lilium amabile	Lilioideae	Liliaceae	37.0%	152,567	MT261159.1
Lilium bulbiferum	Lilioideae	Liliaceae	37.0%	152,686	MW465412.1
Lilium formosanum	Lilioideae	Liliaceae	37.0%	152,653	MK753243.1
Lilium philadelphicum	Lilioideae	Liliaceae	37.1%	152,173	KY940847.1
Fritillaria camschatcensis	Lilioideae	Liliaceae	36.9%	151,569	MW794313.1
Fritillaria cirrhosa	Lilioideae	Liliaceae	37.0%	151,083	MT806755.1
Fritillaria meleagroides	Lilioideae	Liliaceae	37.0%	151,846	MF947710.1
Cardiocrinum cathayanum	Lilioideae	Liliaceae	37.1%	152,502	MK104128.1
Notholirion bulbuliferum	Lilioideae	Liliaceae	37.1%	153,019	MW890006.1
Notholirion macrophyllum	Lilioideae	Liliaceae	37.1%	152,143	MH011354.1
Clintonia udensis	Medeoloideae	Liliaceae	37.0%	156,214	MT261153.1
Medeola virginiana	Medeoloideae	Liliaceae	37.0%	153,914	MK673752.1
Tricyrtis formosana	Streptopoideae	Liliaceae	37.3%	156,018	MK673751.1
Tricyrtis macropoda	Streptopoideae	Liliaceae	37.4%	155,778	MT261173.1
Scoliopus bigelovii	Streptopoideae	Liliaceae	37.2%	154,689	MK673747.1
Prosartes lanuginosa	Streptopoideae	Liliaceae	37.0%	158,265	MK673749.1
Streptopus obtusatus	Streptopoideae	Liliaceae	37.1%	157,614	MK673750.1
Streptopus ovalis	Streptopoideae	Liliaceae	37.1%	157,359	MT261171.1
Calochortus venustus	Calochortoideae	Liliaceae	37.4%	155,688	MT261150.1
Calochortus uniflorus	Calochortoideae	Liliaceae	37.4%	155,794	MK673754.1
Ripogonum album	-	Ripogonaceae	37.6%	160,751	OQ848582.1
Ripogonum elseyanum	-	Ripogonaceae	37.6%	160,292	OQ848585.1
Lapageria rosea	-	Philesiaceae	37.5%	160,054	OQ848581.1
Smilax glabra	-	Smilacaceae	37.1%	157,889	MZ566572.1
Smilax nipponica	-	Smilacaceae	37.1%	158,178	MT261170.1
Smilax riparia	-	Smilacaceae	37.0%	158,570	OL351263.1
Luzuriaga radicans	-	Luzuriageae	38.1%	157,885	KM233640.1
Paris thibetica	-	Melanthiaceae	37.2%	162,708	KY247143.1
Veratrum oxysepalum	-	Melanthiaceae	37.7%	153,705	MW147219.1
Bomarea ovallei	-	Alstroemeriaceae	38.1%	155,018	MW345247

Table 3 Gene composition within the plastomes of *Erythronium* species

Groups of genes		Names of genes	No.
RNA genes	Ribosomal RNAs	*rrn*4.5$^{\times 2}$, *rrn*5$^{\times 2}$, *rrn*16$^{\times 2}$, *rrn*23$^{\times 2}$	8
	Transfer RNAs	*trn*K-UUU^{a}, *trn*Q-UUG, *trn*S-GCU, *trn*G-UCC^{a}, *trn*R-UCU, *trn*C-GCA, *trn*D-GUC, *trn*Y-GUA, *trn*E-UUC, *trn*T-GGU, *trn*S-UGA, *trn*G-GCC, *trnf*M-CAU, *trn*S-GGA, *trn*T-UGU, *trn*L-UAA^{a}, *trn*F-GAA, *trn*V-UAC^{a}, *trn*M-CAU, *trn*W-CCA, *trn*P-UGG, *trn*H-GUG^{X2}, *trn*I-CAU^{X2}, *trn*L-CAA^{X2}, *trn*V-GAC^{X2}, *trn*I-GAU^{aX2}, *trn*A-UGC^{aX2}, *trn*R-ACG^{X2}, *trn*N-GUU^{X2}, *trn*L-UAG	38
Protein genes	Photosystem I	*psa*A, *psa*B, *psa*C, *psa*I, *psa*J	5
	Photosystem II	*psb*A, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*H, *psb*I, *psb*J, *psb*K, *psb*L, *psb*M, *psb*T, *psb*Z	14
	Photosystem I assembly	*paf*I^{b}, *paf*II	2
	Photosystem biogenesis	*pbf*1	1
	Cytochrome	*pet*A, *pet*B^{a}, *pet*D^{a}, *pet*G, *pet*L, *pet*N	6
	ATP synthases	*atp*A, *atp*B, *atp*E, *atp*F^{a}, *atp*H, *atp*I	6
	Large unit of Rubisco	*rbc*L	1
	NADH dehydrogenase	*ndh*A^{a}, *ndh*B^{aX2}, *ndh*C, *ndh*D, *ndh*E, *ndh*F, *ndh*G, *ndh*H, *ndh*I, *ndh*J, *ndh*K	12
	ATP-dependent protease subunit *P*	*clp*P1^{b}	1
	Envelope membrane protein	*cem*A	1
Ribosomal proteins	Large units of ribosome	*rpl*2^{aX2}, *rpl*14, *rpl*16^{a}, *rpl*20, *rpl*22, *rpl*23$^{\times 2}$, *rpl*32, *rpl*33, *rpl*36	11
	Small units of ribosome	*rps*2, *rps*3, *rps*4, *rps*7$^{\times 2}$, *rps*8, *rps*11, *rps*12$^{\times 2}$, *rps*14, *rps*15, *rps*16^{a}, *rps*18, *rps*19	14
Transcription translation	RNA polymerase	*rpo*A, *rpo*B, *rpo*C1^{a}, *rpo*C2	4
	Initiation factor	*infA*$^{\psi}$	1
	Miscellaneous protein	*acc*D, *ccs*A, *mat*K	3
	Hypothetical proteins and conserved reading frames	*ycf*1, *ycf*2$^{\times 2}$	3
Total			131

a: gene with one intron

b: gene with two introns

X2: duplicated gene

ψ: pseudogene

identical topologies, with only minor differences in bootstrap support values among focal nodes (Table S1).

Divergence time estimation

Divergence time estimation was conducted via the same dataset employed in the phylogenomic analysis, which was based on BEAST inference. The crown node of *Erythronium* was estimated to have diverged approximately 16.7 Mya, with a 95% highest posterior density (HPD) interval of 9.3–26.4 Mya (Fig. 6; Table 4). Subsequent divergences of the three major clades were estimated as follows: the Eurasian clade at 8.7 Mya (95% HPD: 3.7–18.0 Mya), the Eastern North American clade at 5.8 Mya (95% HPD: 1.1–13.4 Mya), and the Western North American clade at 3.4 Mya (95% HPD: 1.1–9.3 Mya).

To evaluate the sensitivity of divergence-time estimates to prior assumptions, we conducted an additional analysis in which the fixed root-age constraint was replaced with a Normal prior (mean = 90 Ma, SD = 15 Ma). Under this alternative setting, mean node ages shifted for several focal nodes, particularly the crown of Erythronium and the Western North American clade; however, 95% HPD intervals showed considerable overlap between the two analyses (Table S2). These results indicate that although absolute age estimates are influenced by root-age prior specification, the inferred temporal framework and relative divergence structure remain broadly consistent.

Ancestral area reconstruction

The ancestral area reconstruction inferred from the BBM analysis conducted via the RASP program is shown in Fig. 7. The analysis included species distributed in East Asia, West Asia, Europe, Eastern North America, and Western North America. The ancestral area of *Erythronium* was inferred to be East Asia (B). Notably, *Erythronium* species found in Eastern North America were again confirmed to have diverged from the Eurasian clade—rather than from the geographically closer Western North American clade—which is consistent with previous phylogenetic analyses.

Discussion

Plastome features of *Erythronium* L.

This study presents, for the first time, the complete plastomes of ten previously unreported species within the genus *Erythronium* L. The cp. genome sizes ranged from 150,647 bp in *E. grandiflorum* to 151,827 bp in *E. japonicum*, with an average GC content of 36.6%. A 261–267 bp deletion (corresponding to 87–89 codons, depending on the reference) was identified within the *ycf*2 gene of *Erythronium* plastomes. Due to this deletion represents an in-frame deletion, it does not disrupt the open reading frame of *ycf*2; however, its functional consequences cannot be directly evaluated based on

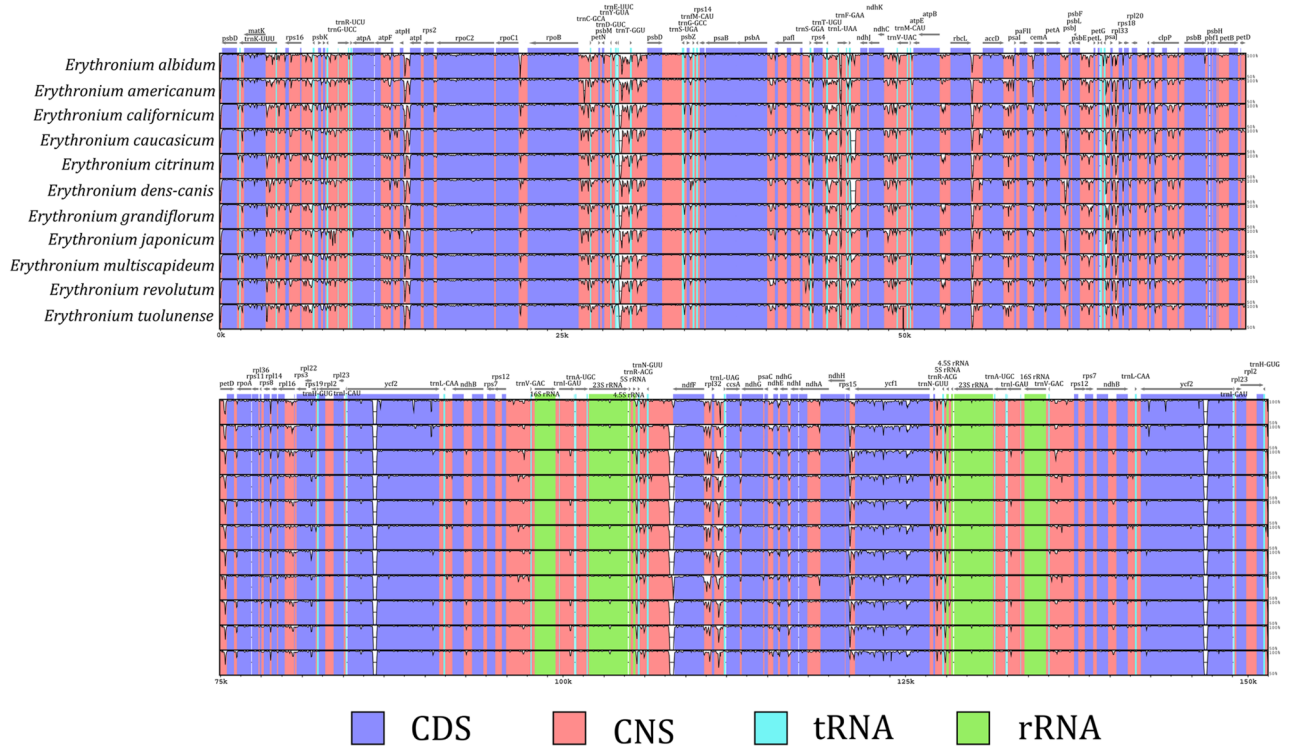

Fig. 2 Comparative plastome alignment of 11 *Erythronium* species via mVISTA. The alignment was performed using the plastome of *Amana edulis* as the reference. The vertical axis of the graph represents the sequence similarity among species. *CDS (coding sequences), *CNS (noncoding sequences)

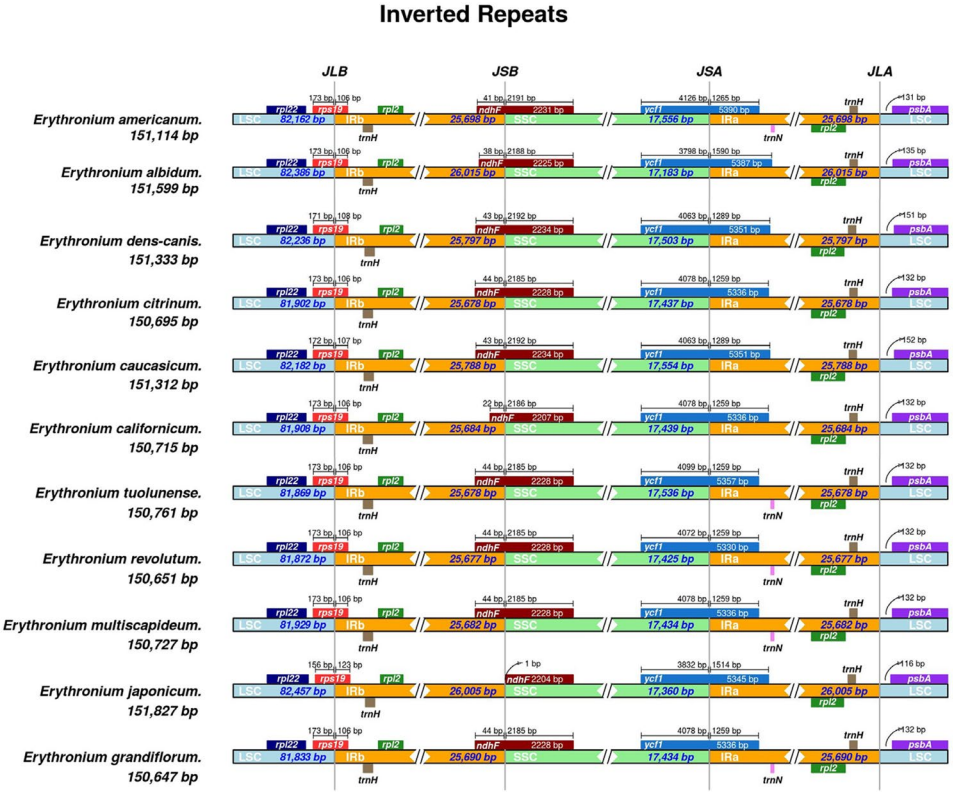

Fig. 3 Comparisons of LSC, SSC, and IR region boundaries between 11 *Erythronium* plastomes

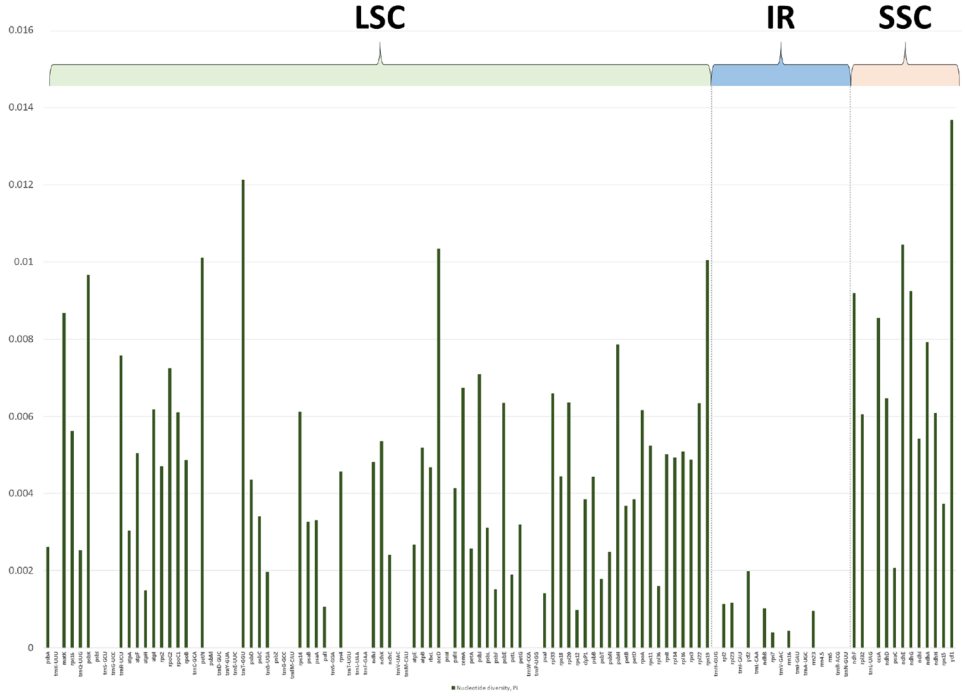

Fig. 4 Nucleotide diversity (Pi) values of protein-coding sequences in 11 *Erythronium* species

sequence data alone. Structural changes in the *ycf*2 gene are frequently observed among monocot lineages, often involving partial deletions or rearrangements. For example, Choi and Lee [8] reported partial deletions in the *ycf*2 gene across multiple species of *Iris* Tourn. ex L., with these deletions located outside of functionally conserved domains. The *inf*A gene, which encodes translation initiation factor 1 (IF-1) involved in chloroplast protein synthesis, was pseudogenized in all analyzed *Erythronium* species. This pattern is consistent with other members of the Liliaceae family and has been widely reported in angiosperms, where it is often associated with gene transfer to the nuclear genome [26].

The plastomes of *Erythronium* retain the typical quadripartite structure found in most land plants, featuring a large single-copy (LSC) region and a small single-copy (SSC) region separated by two inverted repeats (IRs). No major structural rearrangements or boundary shifts were found among the species analyzed, suggesting a conserved genome architecture across the genus. However, compared with other species, *E. japonicum* presented a slight shift in the start position of the *ndh*F gene. This variation is unlikely to affect gene expression or protein integrity and appears to be structurally neutral. Notably, the region surrounding the *ndh*F start site in *E. japonicum* presented a relatively high AT content, which may underlie the observed shift. Previous studies have reported that AT-rich regions in plastid genomes are prone to structural instability and are often associated with insertions, deletions, or inversions [30]. Our findings are consistent with this pattern and provide a possible mechanistic explanation for the localized structural variability. To evaluate genetic variation within the genus, we conducted a sliding window-based analysis of (*Pi*) across all plastid protein-coding sequences. The results revealed that overall nucleotide variability was low. The highest *Pi* value, recorded in the *ycf*1 gene, was only 0.014, while the IR regions presented the lowest values, reinforcing their known structural and evolutionary stability.

Compared with other genera within Liliaceae, *Erythronium* presented a relatively low level of plastome sequence divergence. For example, in *Tulipa* L., the *ycf*1 gene presented a *Pi* value of 0.0188 [2], whereas in *Lilium* Tourn. ex L., the value was approximately 0.019 [37]. These comparative values underscore the limited nucleotide variability observed in *Erythronium* and suggest that this genus may be more evolutionarily conserved at the plastome level than its close relatives are.

In summary, the newly generated plastome data for *Erythronium* reveal a high degree of structural conservation, low nucleotide diversity, and consistent evolutionary patterns within the genus.

Phylogenetic analysis

Molecular phylogenetic studies on the genus *Erythronium* have thus far been limited in scope, primarily relying on a small number of samples and gene regions, and only a few studies have attempted to clarify its phylogenetic structure. Allen et al. [1], analyzed 16 species of

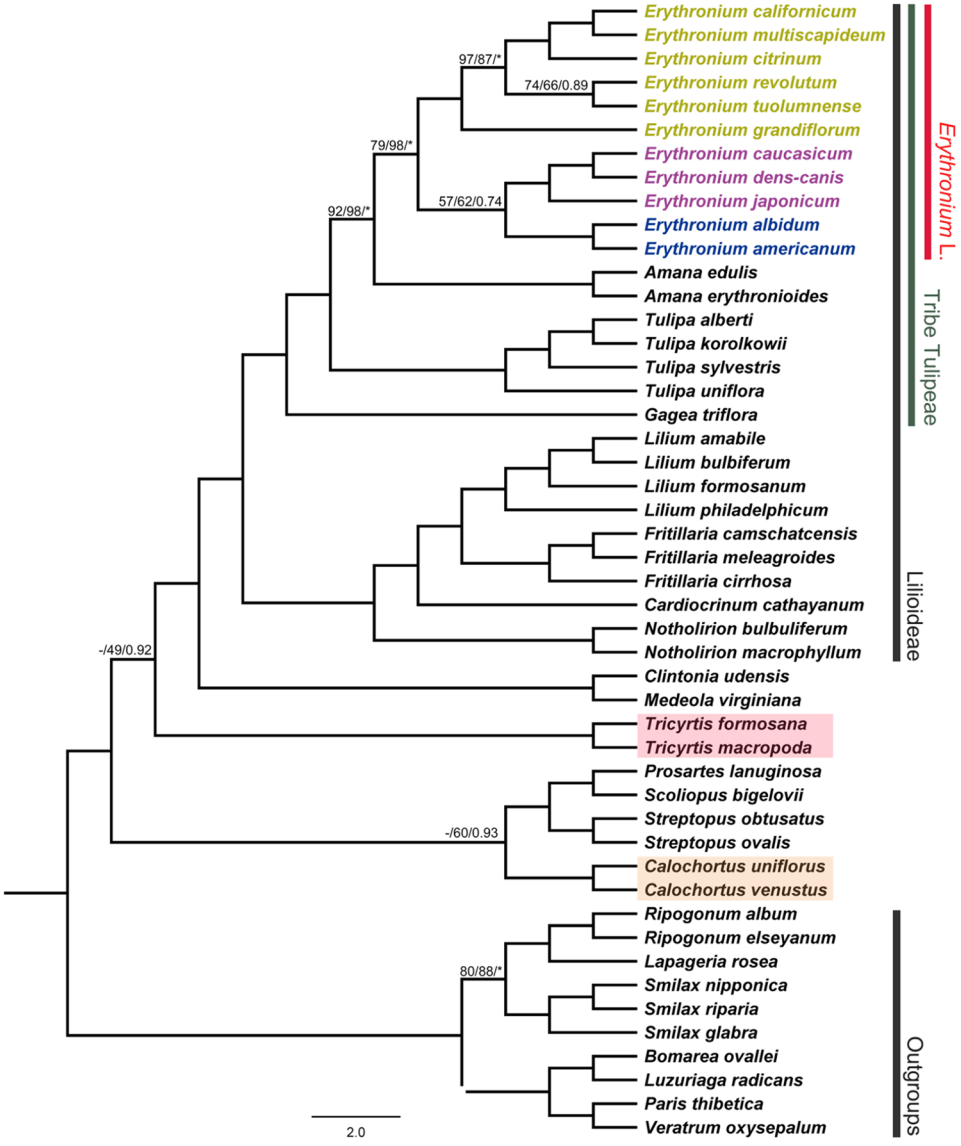

Fig. 5 Maximum likelihood tree of 48 taxa inferred from 78 plastid protein-coding sequences. The numbers indicate support (maximum parsimony bootstrap (MP-BS)/maximum likelihood bootstrap (ML-BS)/posterior probability (PP)). Only support under MP-BS = 100%/ML-BS = 100%/PP = 1.00 is shown. The dashes "-" indicate incongruence between the MP and ML/BI trees. "*" indicates that the support value is 100% or 1.00

Erythronium via the ITS and plastidial *mat*K regions and reported that the genus was divided into three geographically distinct clades: Eastern North America, Western North America, and Eurasia. Clennett et al., [10], through analysis of *rps*16, *trn*K, and ITS sequences, suggested that *Erythronium* and *Amana* share a close phylogenetic relationship, with *Tulipa* forming a sister group to the two genera. Although *Amana* is morphologically distinct from *Erythronium*—lacking foliar mottling and consistently exhibiting a trilobed stigma—phylogenetic analyses revealed its position with low support and instability, highlighting the need for further studies to clarify its systematic placement.

While these previous studies have contributed to elucidating the major taxonomic structure of *Erythronium*, some lineages—particularly within the Western North American group—remain unresolved in terms of species boundaries and interrelationships. To address these limitations, the present study performed phylogenetic analyses using 78 plastid protein-coding sequences from the complete plastome. To clarify the phylogenetic placement of *Erythronium* within the family Liliaceae, a total of 48 species—including 11 outgroup taxa—were included in the phylogenetic tree reconstruction. The results confirmed a distinct division into three clades—Eurasia, Eastern North America, and Western North America—which is consistent with the branching

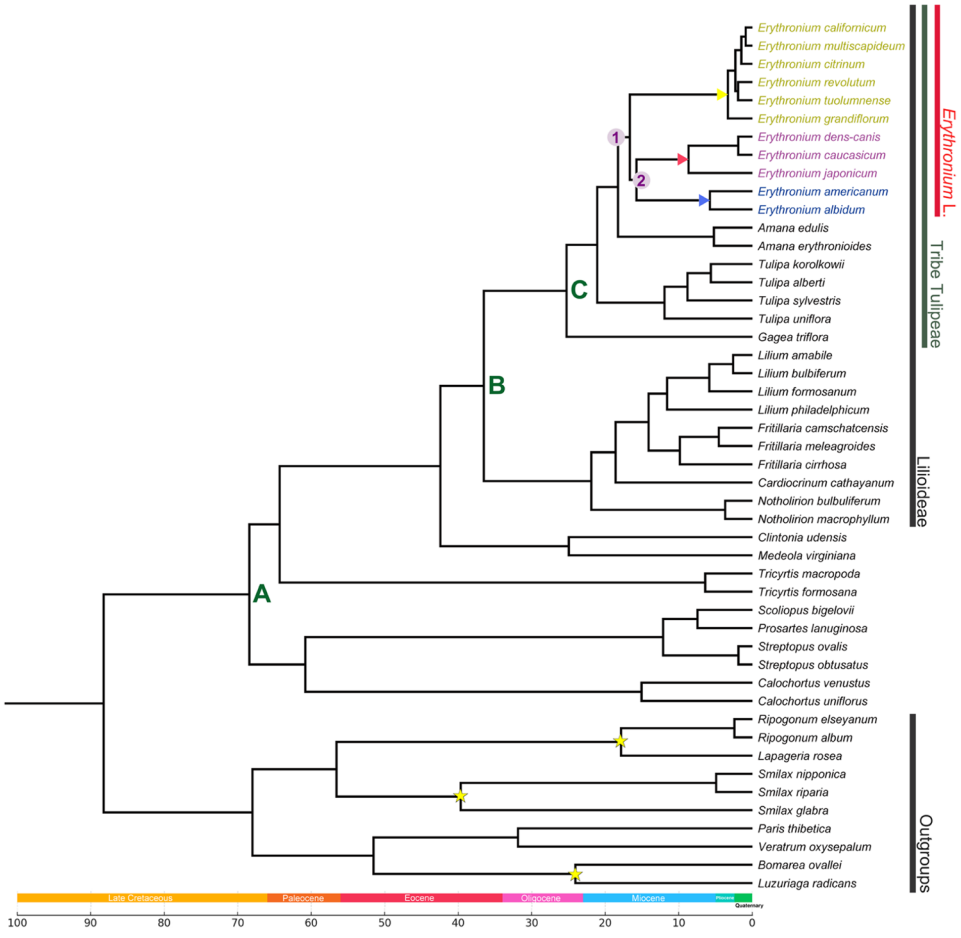

Fig. 6 Chronogram showing the divergence times estimated in BEAST on the basis of the combined 78 plastid protein-coding sequences. The fossil calibration points are marked with yellow stars

Table 4 Posterior age distributions of major nodes in this study

Node	Node description	Mean age (Mya)	95% HPD
A	Family Liliaceae	68.5	45.0-88.4
B	Subfamily Lilioideae	36.6	21.9–59.0
C	Tribe Tulipeae	25.3	14.8–58.6
1	Dispersal event of *Erythronium*	16.7	9.3–26.4
2		15.8	8.8–25.4
Blue	Eastern North American	5.8	1.1–13.4
Red	Eurasian clade	8.7	3.7–18.0
Yellow	Western North American	3.6	1.1–9.3

patterns previously reported. Notably, *Amana* was consistently recovered as the closest sister group to *Erythronium*, as supported by strong statistical values across all the analytical methods, thus reinforcing their close phylogenetic affinity.

The major plastome clades recovered in this study correspond to several well-established morphological traits within *Erythronium*. Species belonging to the Western North American clade lack an elaiosome, a character that is known to distinguish these taxa from Eurasian and Eastern North American species. In contrast, species of the Eurasian clade typically exhibit strongly reflexed tepals at anthesis, a floral trait that differentiates them from the other major lineages.

These morphology–clade associations indicate that the phylogenetic structure inferred from plastome data is reflected not only in geographic patterns but also in phenotypic differentiation across the genus. While morphological characters in *Erythronium* can be variable, particularly among Western North American taxa, the correspondence observed here provides a useful interpretative context for the major lineages identified in this study.

Phylogenetic trees generated via MP, ML, BI showed congruent topologies, supporting a stable evolutionary structure. While the ML and BI analyses yielded consistent topologies with strong support values, the MP analysis produced a different topology, notably altering the phylogenetic placement of the genus *Tricyrtis*. Such incongruences among phylogenetic methods have also been reported in previous studies, for example, in

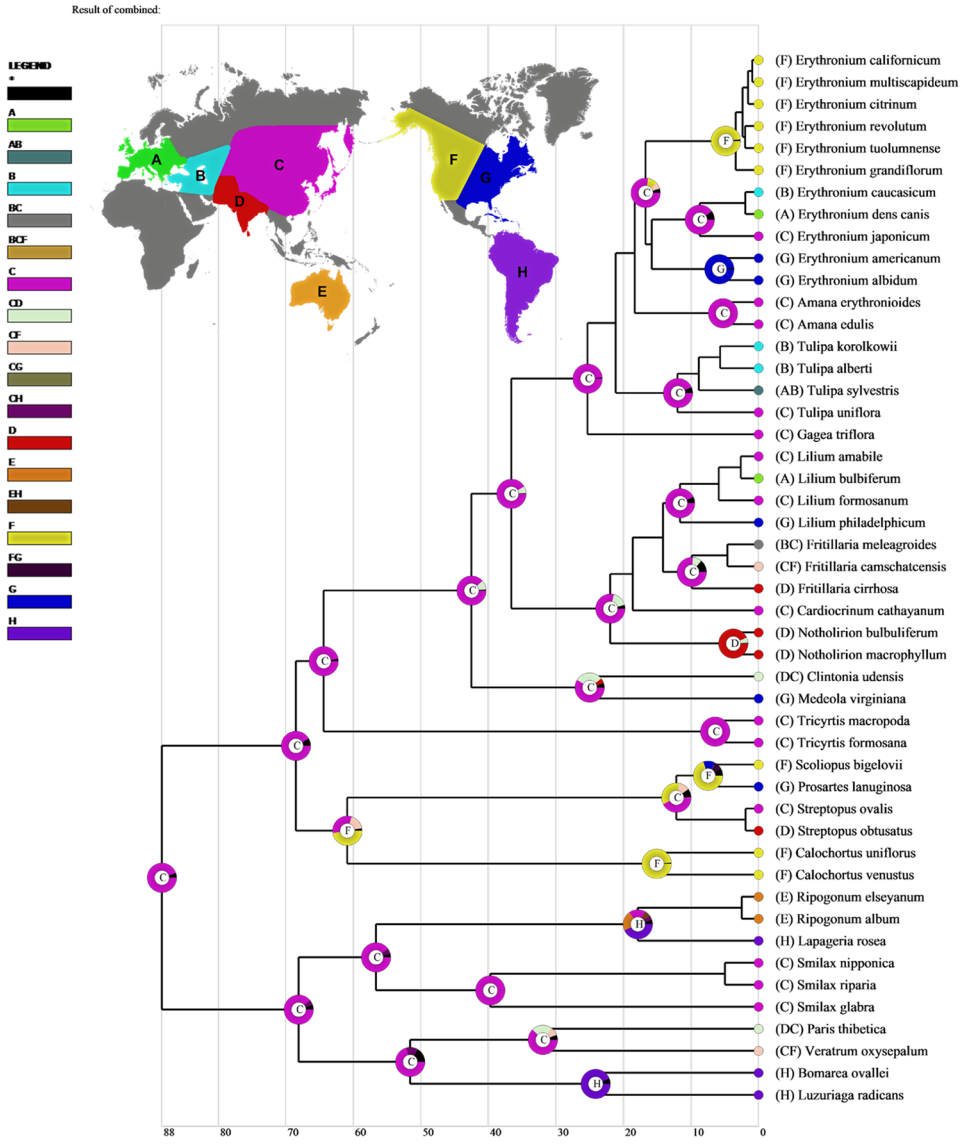

Fig. 7 Summary of the Bayesian binary method (BBM) model of ancestral area reconstruction based on a reduced BEAST combined-gene chronogram. The biogeographical regions used in the BBM analysis were coded as follows: A, Europe; B, Western Asia; C, East Asia; D, South Asia; E, Australia; F, Western North America; G, Eastern North America; H, South America

Commelinaceae, where MP and ML/BI analyses resulted in topological differences [20]. The placement of *Tricyrtis* has been the subject of ongoing taxonomic debate, with various studies proposing alternative positions within Liliaceae, underscoring the need for further investigation [22, 28].

Divergence time and biogeographical origin

The results of this study confirmed that *Erythronium* is divided into three well-supported clades—Eastern North America, Western North America, and Eurasia—each exhibiting a geographically disjunct and mutually exclusive distribution. Such disjunct patterns are common not only in plant taxa but also across a wide range of other organisms. In particular, the morphological and genetic similarities between taxa in East Asia and Eastern North America, known as the "Asa Gray disjunction", have long constituted a classic model system in historical biogeography. Tiffney [40], described a series of climatic and vegetational shifts throughout the Cenozoic era that likely contributed to such patterns. During the early Eocene, warm and humid global climates allowed tropical and subtropical vegetation to extend across much of the Northern Hemisphere. As global temperatures gradually cooled from the late Eocene through the Oligocene, tropical vegetation retreated and was replaced by expanding temperate deciduous forests. Throughout the Miocene, plant exchanges continued via the Bering

land bridge, and there was substantial diversification of plant assemblages adapted to cooler climates. The rapid climate cooling that followed the Pliocene, along with repeated glacial–interglacial cycles, likely induced large-scale shifts in plant distributions, range fragmentations, and extinctions across the Northern Hemisphere. In this context, the relatively cooler global temperatures prevailing from the Miocene onward would have favored the expansion of cold-adapted plant lineages such as *Erythronium*. Conversely, the drastic temperature declines after the Pliocene may have caused fragmentation or extinction of lineages that were once broadly distributed across the northern temperate zone.

According to our divergence time analysis, *Erythronium* crown age was estimated approximately 16.7 Mya, with East Asia inferred as the most likely ancestral area. The three major clades—Eurasian, Eastern North American, and Western North American—diversified during the late Miocene to early Pliocene, with mean age estimates of 8.7, 5.8, and 3.4 Mya, respectively. Although mean estimates suggest a sequential pattern of divergence, the 95% HPD intervals overlap considerably, and therefore the precise temporal ordering of these events should be interpreted with caution. These results suggest that diversification within *Erythronium* largely occurred during the late Miocene to early Pliocene, a period characterized by progressive global cooling and major ecological shifts. This timing is broadly consistent with patterns observed in monocot evolution, in which diversification rates increased in several lineages during the late Miocene under changing climatic conditions. For example, in a phylogeographic study of *Gagea* Salisb., Peterson et al., [29] reported that the genus originated in Southwestern Asia during the early Miocene and experienced rapid diversification during the middle to late Miocene, accompanied by expansion into the Mediterranean and East Asian regions. These parallels suggest that the late Miocene may have represented an important interval for diversification in temperate-adapted monocots such as *Erythronium*.

In light of the phylogenetic and biogeographic patterns revealed in this study, the geographically disjunct diversification of *Erythronium* into three major clades—Eurasia, Eastern North America, and Western North America—appears to be the result of multiple interacting factors, including historical dispersal events, paleoclimatic fluctuations, and long-term geographic isolation.

First, the initial intercontinental dispersal of *Erythronium* from East Asia to North America was likely facilitated by the Bering land bridge during the Miocene. This land connection serves as a major floristic corridor, particularly for temperate taxa, and is widely invoked to explain the classic East Asia–Eastern North America disjunction pattern [40, 43].

Second, the estimated divergence times of the three clades—occurring primarily between the late Miocene and early Pliocene—coincide with a period of progressive global cooling. Such climatic changes are known to have caused habitat fragmentation, range shifts, and extinction events, promoting allopatric speciation across the Northern Hemisphere [47]. Notably, the diversification trajectory of *Erythronium* mirrors that of other Liliaceae members, such as *Gagea*, which also experienced accelerated lineage diversification during this time in response to the same environmental pressures [29].

Third, within North America, the divergence between the Eastern and Western clades may have been reinforced by the presence of major physiographic barriers such as the Rocky Mountains. These mountains have historically restricted gene flow and promoted biotic differentiation in numerous plant groups [32] and may have similarly contributed to the regional isolation and independent evolution of *Erythronium* lineages.

Collectively, these findings support the hypothesis that the disjunct phylogeographic structure of *Erythronium* is the result of a complex interplay between Miocene-era dispersal, post-Miocene climatic changes, and continental-scale geographic isolation.

Limitations and implications of plastome-based phylogenetic inference

Although plastome-scale data provide strong phylogenetic signal and have proven effective for resolving deep and intermediate-level relationships in angiosperms, plastomes are generally maternally inherited and thus represent a single parental lineage. Consequently, plastome phylogenies do not necessarily reflect the species tree and may be influenced by processes such as chloroplast capture, introgression, or historical hybridization.

These limitations are particularly relevant for *Erythronium*, as several species—especially in Western North America—exhibit overlapping distributions, extensive morphological similarity, and previously reported hybridization. Accordingly, the plastome-based relationships recovered in this study should be interpreted as representing chloroplast lineage history rather than definitive species-level evolutionary relationships.

The present study did not include nuclear loci or phylogenomic target-capture datasets that could independently corroborate the plastome topology. The absence of nuclear evidence implies that several alternative evolutionary scenarios cannot be distinguished with the current dataset. Specifically, apparent plastome monophyly of geographically defined lineages may reflect chloroplast capture rather than organismal relationships, recent divergence and incomplete lineage sorting may contribute to discordance between plastid and nuclear genealogies, and ongoing or historical gene flow—particularly

within the species-rich Western North American lineage—may obscure species boundaries and produce non-bifurcating evolutionary histories.

Therefore, species-level relationships within each major clade, the directionality and number of intercontinental dispersal events, and any interpretations that implicitly assume a strictly bifurcating species tree should be regarded as provisional when based solely on plastome data. In addition, divergence-time estimates derived from a single non-recombining locus system represent the timing of chloroplast lineage splits, which may predate or postdate species divergences under introgression or chloroplast capture, and should be interpreted accordingly.

In this context, while our results robustly support the sister-group relationship between *Erythronium* and *Amana*, as well as the presence of three major geographically structured plastome lineages within *Erythronium*, fully resolving species-level relationships and testing hypotheses of reticulate evolution will require integrative analyses incorporating low-copy nuclear genes or phylogenomic target-enrichment approaches in future studies.

Conclusion

This study conducted an integrative analysis of the plastome-based evolutionary history, phylogenetic structure, and biogeographical dispersal patterns of the genus *Erythronium* on the basis of chloroplast genome data. By incorporating ten newly sequenced chloroplast genomes, we re-evaluated the phylogenetic relationships and divergence times within the genus. The assembled plastomes exhibited a typical quadripartite structure, with several genes displaying small deletions contributing to structural variation among plastomes. The pseudogenization of *inf*A, which was observed consistently across the sampled taxa, corresponds to a well-documented pattern in Liliaceae. Overall, genetic variation was relatively low, which may account for the reduced support values observed in certain phylogenetic analyses.

Phylogenomic analyses based on 78 plastid protein-coding sequences revealed three well-supported and geographically distinct clades corresponding to Eurasia, Eastern North America, and Western North America. *Amana* was identified as the closest sister group to *Erythronium*, with strong support via all the analytical methods. While the ML and BI analyses produced congruent topologies with high support values, the MP analysis yielded topological differences in certain taxa, such as *Tricyrtis*, a genus previously noted for its unstable taxonomic placement. These inconsistencies may result from low nucleotide variation and methodological sensitivity, as observed in related studies. Molecular dating analysis via BEAST v1.10.4 estimated a crown age of approximately 16.7 Mya for *Erythronium*, with East Asia inferred as the most likely ancestral area. Subsequent divergence of the three major plastome lineages occurred during the late Miocene to early Pliocene, with mean age estimates of 8.7, 5.8, and 3.4 Mya for the Eurasian, Eastern North American, and Western North American clades, respectively. This temporal framework is broadly consistent with diversification patterns reported in several monocot lineages during periods of global cooling and ecological change. Ancestral area reconstruction further supported East Asia as the most likely ancestral region, with intercontinental dispersal to North America likely associated with the availability of the Bering land bridge.

The current geographic disjunction of *Erythronium* is best explained by a combination of historical intercontinental dispersal, post-Miocene climatic shifts, and regional geographic barriers such as the Rocky Mountains. Collectively, these findings provide new insights into the evolutionary dynamics of temperate-adapted monocots and highlight the value of complete plastome data in resolving phylogenetic and biogeographic patterns in complex plant groups.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-026-08677-6.

Supplementary Material 1: Table S1 Comparison of bootstrap support values(BS) between gene-partitioned and unpartitioned maximum likelihood (ML) analyses. Table S2 Sensitivity analysis of divergence-time estimates comparing primary and alternative root-age prior settings. Table S3 Summary of plastome assembly quality and validation metrics. Fig S1. Representative plastome map of *E. albidum*. Genes are color-coded by functional group. The transcription of genes occurs in a clockwise direction for those positioned on the inner side of the circular map and in a counterclockwise direction for those on the outer side. The GC content is depicted by a central gray circle. Fig S2. Representative plastome map of *E. americanum*. Genes are color-coded by functional group. The transcription of genes occurs in a clockwise direction for those positioned on the inner side of the circular map and in a counterclockwise direction for those on the outer side. The GC content is depicted by a central gray circle. Fig S3. Representative plastome map of *E. californicum*. Genes are color-coded by functional group. The transcription of genes occurs in a clockwise direction for those positioned on the inner side of the circular map and in a counterclockwise direction for those on the outer side. The GC content is depicted by a central gray circle. Fig S4. Representative plastome map of *E. caucasicum*. Genes are color-coded by functional group. The transcription of genes occurs in a clockwise direction for those positioned on the inner side of the circular map and in a counterclockwise direction for those on the outer side. The GC content is depicted by a central gray circle. Fig S5. Representative plastome map of *E. citrinum*. Genes are color-coded by functional group. The transcription of genes occurs in a clockwise direction for those positioned on the inner side of the circular map and in a counterclockwise direction for those on the outer side. The GC content is depicted by a central gray circle. Fig S6. Representative plastome map of *E. dens-canis*. Genes are color-coded by functional group. The transcription of genes occurs in a clockwise direction for those positioned on the inner side of the circular map and in a counterclockwise direction for those on the outer side. The GC content is depicted by a central gray circle. Fig S7. Representative plastome map of *E. grandiflorum*. Genes are color-coded by functional group. The transcription of genes occurs in a clockwise direction for those positioned on the inner side of the circular map and in a counterclockwise direction for those on the outer side. The GC content is depicted by a central gray circle. Fig S8. Representative plastome map of *E. multiscapideum*. Genes are color-coded by functional group. The

transcription of genes occurs in a clockwise direction for those positioned on the inner side of the circular map and in a counterclockwise direction for those on the outer side. The GC content is depicted by a central gray circle. Fig S9. Representative plastome map of *E. revolutum*. Genes are color-coded by functional group. The transcription of genes occurs in a clockwise direction for those positioned on the inner side of the circular map and in a counterclockwise direction for those on the outer side. The GC content is depicted by a central gray circle. Fig S10. Representative plastome map of *E. tuolumnense.* Genes are color-coded by functional group. The transcription of genes occurs in a clockwise direction for those positioned on the inner side of the circular map and in a counterclockwise direction for those on the outer side. The GC content is depicted by a central gray circle. Fig S11. The Maximum Parsimony bootstrap consensus tree of 48 taxa inferred from 78 CDS. Numbers above the node indicate maximum parsimony bootstrap support values (MP-BS). Only support under MP-BS = 100% is shown.

Acknowledgements
The authors thank Sue Zmarzty (Royal Botanic Gardens, Kew), Dr. A. Stefanaki (Utrecht University Botanic Gardens), Jan-Willem Vos (SCCV Château Pérouse), Margeaux Apple (Cambridge University Botanic Garden), and Eduardo Fernández Pascual for help in collecting plant material.

Authors' contributions
YGK designed the study, performed data collection, data processing, and analysis, and wrote the manuscript. JJ contributed to the conceptualization and revised the manuscript. JHK supervised the research, provided critical comments on the manuscript, and approved the final version for publication.

Funding
This research was supported by a grant from the National Research Foundation of Korea (NRF), funded by (NRF-2017R1D1A1BO6029326) and the Gachon University research fund of 2023. (GCU- 202307920001).

Data availability
The plastome sequences newly generated in this study have been deposited in the GenBank database under accession numbers PX122074-PX122084. All other data generated or analysed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate
The study including plant samples complies with relevant institutional, national, and international guidelines and legislation. No specific permits were required for plant collection. The study did not require ethical approval or consent, as no endangered or protected plant species were involved.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 9 October 2025 / Accepted: 27 March 2026
Published online: 14 April 2026

References

1. Allen GA, Soltis DE, Soltis PS. Phylogeny and biogeography of Erythronium (Liliaceae) inferred from chloroplast matK and nuclear rDNA ITS sequences. Systematic Botany. 2003;28(3):512–23.
2. Almerekova S, Yermagambetova M, Ivaschenko A, Turuspekov Y, Abugalieva S. Comparative Analysis of Plastome Sequences of Seven *Tulipa* L.(Liliaceae Juss.) Species from Section Kolpakowskianae Raamsd. Ex Zonn and Veldk. Int J Mol Sci. 2024;25(14):7874.
3. Amiryousefi A, Hyvönen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics. 2018;34(17):3030–1.
4. Applegate EI. The genus *Erythronium*: a taxonomic and distributional study of the Western North American species. Madroño. 1935;3(2):58–113.
5. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, Sidow A, Batzoglou S, Program NCS. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res. 2003;13(4):721–31.
6. Carpenter RJ, Wilf P, Conran JG, Cúneo NR. Early Eocene Ripogonum. Paleontologia Electronica. 2014;17(3):1–39.
7. Chan PP, Lowe TM. tRNAscan-SE: searching for tRNA genes in genomic sequences. Springer; 2019.
8. Choi T-Y, Lee S-R. Complete plastid genome of Iris orchioides and comparative analysis with 19 Iris plastomes. PLoS ONE. 2024;19(4):e0301346.
9. Clennett JC. A taxonomic revision of Erythronium L.(Liliaceae). The Open University]; 2006.
10. Clennett JC, Chase MW, Forest F, Maurin O, Wilkin P. Phylogenetic systematics of *Erythronium* (Liliaceae): morphological and molecular analyses. Bot J Linn Soc. 2012;170(4):504–28.
11. Conran JG, Carpenter RJ, Jordan GJ. Early eocene Ripogonum (Liliales: Ripogonaceae) leaf macrofossils from southern Australia. Aust Syst Bot. 2009;22(3):219–28.
12. Conran JG, Bannister JM, Mildenhall DC, Lee DE, Chacón J, Renner SS. Leaf fossils of Luzuriaga and a monocot flower with in situ pollen of Liliacidites contortus Mildenh. & Bannister sp. nov.(Alstroemeriaceae) from the Early Miocene. Am J Bot. 2014;101(1):141–55.
13. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical bulletin. 1987.
14. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol. 2006;4(5):e88.
15. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.
16. Gernhard T, Hartmann K, Steel M. Stochastic properties of generalized Yule models, with biodiversity applications. J Math Biol. 2008;57:713–35.
17. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47(W1):W59–64.
18. Harper RM. *Erythronium* albidum in Alabama, and some of its relatives. Castanea. 1945;10(1):1–7.
19. Jin J-J, Yu W-B, Yang J-B, Song Y, DePamphilis CW, Yi T-S, Li D-Z. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21:1–31.
20. Jung J, Pellegrini MO, Kim J-H. Historical biogeography and plastome evolution of Commelinaceae Mirb.(Commelinales) corroborate the East Gondwanan origins. BMC Plant Biol. 2025;25(1):533.
21. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647–9.
22. Kim JS, Kim J-H. Updated molecular phylogenetic analysis, dating and biogeographical history of the lily family (Liliaceae: Liliales). Bot J Linn Soc. 2018;187(4):579–93.
23. Köppen WP. Grundriss der klimakunde. W. de Gruyter; 1931.
24. Lin R-C. Evolution and Genetics of Floral Color Polymorphisms in Clarkia gracilis ssp. sonomensis and Erythronium umbilicatum Duke University]. 2020.
25. Mennes CB, Lam VK, Rudall PJ, Lyon SP, Graham SW, Smets EF, Merckx VS. Ancient Gondwana break-up explains the distribution of the mycoheterotrophic family Corsiaceae (Liliales). J Biogeogr. 2015;42(6):1123–36.
26. Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, Kavanagh TA, Hibberd JM, Gray JC, Morden CW. Many parallel losses of infA from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell. 2001;13(3):645–58.
27. Parks CR, Hardin JW. Yellow *Erythronium*s of the Eastern United States. Brittonia. 1963;15(3):245–59.
28. Patterson TB, Givnish TJ. Phylogeny, concerted convergence, and phylogenetic niche conservatism in the core Liliales: insights from rbcL and ndhF sequence data. Evolution. 2002;56(2):233–52.
29. Peterson A, Harpke D, Peterson J, Harpke A, Peruzzi L. A pre-Miocene Irano-Turanian cradle: Origin and diversification of the species-rich monocot genus Gagea (Liliaceae). Ecol Evol. 2019;9(10):5870–90.
30. Qin Q, Dong Y, Chen J, Wang B, Peng Y, Zhang X, Wang X, Zeng J, Zhong G, Zhang S. Comparative analysis of chloroplast genomes reveals molecular evolution and phylogenetic relationships within the Papilionoideae of Fabaceae. BMC Plant Biol. 2025;25(1):1–20.

31. Rambaut, A. 2009. FigTree. Tree Figure Drawing Tool. Institute of Evolutionary Biology, University of Edinburgh. Available from: http://tree.bio.ed.ac.uk/software/figtree/
32. Raven PH, Axelrod DI. Origin and relationships of the California flora. Volume 72. Univ of California; 1978.
33. Ravi V, Khurana J, Tyagi A, Khurana P. An update on chloroplast genomes. Plant Syst Evol. 2008;271:101–22.
34. Roccaforte K, Russo SE, Pilson D. Hybridization and reproductive isolation between diploid *Erythronium* mesochoreum and its tetraploid congener E. albidum (Liliaceae). Evolution. 2015;69(6):1375–89.
35. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61(3):539–42.
36. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DnaSP 6: DNA sequence polymorphism analysis of large datasets. Mol Biol Evol. 2017;34(12):3299–302.
37. Sheikh-Assadi M, Naderi R, Kafi M, Fatahi R, Salami SA, Shariati V. Complete chloroplast genome of *Lilium* ledebourii (Baker) Boiss and its comparative analysis: lights into selective pressure and adaptive evolution. Sci Rep. 2022;12(1):9375.
38. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 2018;4(1):vey016.
39. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol. 2021;38(7):3022–7.
40. Tiffney BH. Perspectives on the origin of the floristic similarity between Eastern Asia and Eastern North America. J Arnold Arboretum. 1985;66(1):73–94.
41. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq–versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45(W1):W6–11.
42. Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. 2016;44(W1):W232–5.
43. Wen J. Evolution of Eastern Asian and Eastern North American disjunct distributions in flowering plants. Annu Rev Ecol Syst. 1999;30(1):421–55.
44. Wilde V, Frankenhäuser H. The middle eocene plant taphocoenosis from Eckfeld (Eifel, Germany). Rev Paleobotany Palynology. 1998;101(1–4):7–28.
45. Wilf P. Late Paleocene–early Eocene climate changes in southWestern Wyoming: Paleobotanical analysis. Geol Soc Am Bull. 2000;112(2):292–307.
46. Yu Y, Harris AJ, Blair C, He X. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. Mol Phylogenet Evol. 2015;87:46–9.
47. Zachos J, Pagani M, Sloan L, Thomas E, Billups K. Trends, rhythms, and aberrations in global climate 65 Ma to present. Science. 2001;292(5517):686–93.
48. Zhang N, Zeng L, Shan H, Ma H. Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. New Phytol. 2012;195(4):923–37.

Publisher's Note