BMC Plant Biology

RESEARCH Open Access

Phylogenomics of East Asian lineage within subgenus *Anguinum* (*Allium*, Amaryllidaceae): insights into its taxonomic puzzles and phylogenetic conflicts

Rui-Yu Cheng[1], Yuan Wang[1], Bo-Ni Song[1], Jin-Bo Tan[1], Jia-Hui Sun[2], Song-Dong Zhou[1], Xing-Jin He[1] and Deng-Feng Xie[1*]

Abstract

Phylogenomic data enriched with informative loci have significantly improved phylogenetic resolution and facilitated the elucidation of evolutionary mechanisms underlying phylogenetic discordance. Species of the East Asian lineage (EAL) within the *Allium* subgenus *Anguinum* are widespread in the Himalaya–Hengduan Mountains (HHMs), which exhibit long-standing taxonomic ambiguity and phylogenetic discordance, calling for investigation. In this study, we collected 102 samples, including 45 transcriptomes and 57 plastid genomes, covering multiple populations of all currently recognized taxa within the EAL and relatives. A total of 2,186 low-copy nuclear genes (LCGs) and 163 plastid sequences (including 111 genes and 52 intergenic regions) were employed for phylogenetic analyses. Our results revealed that the EAL is a monophyletic taxon but exhibits a polytomous phylogeny, it further divides into four sublineages in the LCG-based tree and two sublineages in the plastid-based tree, which display distinct geographical distribution patterns. Samples of *A. ovalifolium* var. *leuconeurum*, *A. ovalifolium* var. *cordifolium* and *A. funckiifolium* from the northwest Sichuan Basin and Qinling-Daba Mountains clustered within the *A. ovalifolium* samples of these regions, while *A. nanodes* is entirely embedded within the HHMs populations of *A. prattii* and *A. ovalifolium*. Extensive phylogenetic conflicts were detected within EAL, and the ancestral area reconstruction indicates that the EAL originated in the Hengduan Mountains (HDMs). Morphological and phylogenetic evidence confirmed the varietal status of *A. ovalifolium* var. *leuconeurum* and *A. ovalifolium* var. *cordifolium*, while also proposing the reclassification of *A. funckiifolium* as *A. ovalifolium* var. *funckiifolium*. The observed polytomous phylogeny within EAL is likely attributed to rapid radiations triggered by geological events and climatic fluctuations during the late Pliocene and Pleistocene, coupled with recurrent isolation–contact dynamics, which resulted in the retention of ancestral polymorphisms and historical gene flow. Widespread phylogenetic discordance in the EAL is mainly due to incomplete lineage sorting (ILS), with hybridization also playing key roles. This study not only reveals the underlying causes of taxonomic controversies within the EAL but

*Correspondence:
Deng-Feng Xie
df_xie2017@163.com

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

also provides critical insights into the unique phylogenetic patterns and evolutionary mechanisms shaping plant lineages in the HHMs biodiversity hotspot.

Keywords East Asian lineage of *Anguinum*, Phylogenome, Incomplete lineage sorting, Phylogenetic discordance, Reticulate evolution

Introduction

Allium L., one of the largest monocotyledon genera in the Northern Hemisphere, comprises more than 1,000 species [1, 2] and is the sole member of the monotypic tribe Allieae within subfamily Allioideae (Amaryllidaceae) [3, 4]. Species of this genus are distributed predominantly across temperate and subtropical regions, including seasonally dry areas like the Irano-Turanian zone. The genus exhibits two primary diversity centers: one spanning Southwest/Central Asia to the Mediterranean, and another in North America [5–8], and encompasses economically critical crops like garlic, onion, and leek [9]. Taxonomically, *Allium* is divided into 15 subgenera, 72 sections and three evolutionary lineages, which have been confirmed by many studies [2, 6, 7, 10–13]. Despite revisions that integrated morphological and molecular markers (*rbcL*, *ITS*, *rps16*) [6, 10], challenges persist, including polyphyly in subgenera (e.g., *Allium*, *Cepa*), unresolved relationships within lineages, subgenera, or sections [9, 10, 14–16], and unclear causes for phylogenetic conflicts [15–17].

The East Asian lineage (EAL) is one of the two lineages of the subg. *Anguinum* (G. Don ex W.D.J. Koch) N. Friesen [6, 10], species of which are characterized by distinctive broad leaves and reticulated bulb tunics, features that serve as key diagnostic characters differentiating it from other *Allium* taxa [1, 6, 18]. Since the establishment of sect. *Anguinum*, the number of species within this taxon has undergone continuous modifications due to accumulating morphological and molecular evidence, as well as the ongoing discovery of new species (Fig. S1) [1, 5, 18–21]. According to the most recent research by Yang et al. [20], the subgenus currently comprises 10 recognized species and several infraspecific variants, exhibiting a broad transcontinental distribution spanning southwestern Europe, East Asia, and northeastern North America [1]. In China, this subgenus is represented by eight species and two varieties. *Allium prattii* C.H.Wright and *A. ovalifolium* Hand.-Mazz have wide distributions across the Himalaya–Hengduan Mountains (HHMs), whereas *A. microdictyon* Prokh., *A. ochotense* Prokh., and *A. listera* Stearn predominantly occur in central and northeastern China. In contrast, *A. funckiifolium* Handel-Mazzetti, *A. ovalifolium* var. *leuconeurum* J.M. Xu, *A. ovalifolium* var. *cordifolium* (J.M. Xu) J.M. Xu, and *A. nanodes* Airy Shaw exhibit restricted distributions confined to the Hengduan Mountains (HDMs) and adjacent areas. The European endemic *A. victorialis* L. is distributed across the Caucasus, and the North American endemic *A. tricoccum* Solander occurs throughout the United States and Canada. Additionally, the Korean island endemic *Allium ulleungense* H.J. Choi & N. Friesen is restricted to Ulleungdo Island [1, 5, 12, 22–24]. Based on phylogenetic trees inferred from *ITS* and chloroplast fragments, the subgenus is divided into two major clades: the East Asian Lineage (EAL) and the Eurasian-North American Lineage (ENAL) [1, 6]. The EAL is represented by *Allium prattii* and includes two additional species and two varieties (*A. ovalifolium*, *A. nanodes*, *A. ovalifolium* var. *leuconeurum*, and *A. ovalifolium* var. *cordifolium*). *Allium prattii* is characterized by purple tepals, broad linear leaves with a cuneate base, and is widely distributed in alpine meadows and shrublands at elevations above 3,000 m in the HHMs region. *A. ovalifolium* is distinguished by white flowers, ovate leaves with a shallow cordate or rounded base. It is widely distributed along forest margins and hillsides at elevations of 1,500–3,200 m in the HDMs and central China. This species includes two varieties: *A. ovalifolium* var. *leuconeurum*, distinguished by its white leaf veins, and *A. ovalifolium* var. *cordifolium*, which differs in having a distinctly cordate leaf base. These two varieties are narrowly distributed and co-occur on forested slopes around 2,500 m elevation in the northwestern part of the Sichuan Basin. *Allium nanodes* is a narrow endemic characterized by falcate leaves and extremely short pedicels, and is confined to scree slopes at elevations above 4,000 m in the HDMs. Notably, although *A. funckiifolium* lacks molecular data due to its long absence from collection, it has generally been considered part of the East Asian Lineage in previous studies, based on morphological features and distribution described by Xu et al. [18]. This species is unique in having a single leaf and is restricted to bamboo forests and forest understories at around 2,000 m elevation along the border between Hubei and Chongqing provinces in China. In addition to the species of the EAL, three other species of this subgenus occur in China: *A. listera*, *A. microdictyon*, and *A. ochotense*, all of which are widely distributed in forested mountain slopes at elevations of 600–2,000 m in central and northeastern China [7, 18, 24–26].

Early research on the EAL primarily centered on morphological and cytological analyses. Pioneering work demonstrated significant taxonomic value through distinct adaxial-abaxial leaf epidermal differentiation among its five constituent species [27]. Subsequent investigations

highlighted striking morphological homogeneity within the EAL, particularly in foliar traits [22, 23]. Notably, population-level studies of *A. prattii* revealed geographically structured leaf morphological variation indicative of environmental adaptation [28]. Cytologically, while all *Anguinum* species share a conserved chromosome base number ($n = 8$), they exhibit remarkable ploidy diversity spanning diploid (2n) to pentaploid (5n) cytotypes [29–33]. Molecular phylogenetic analyses revealed that the subg. *Anguinum* is sister to subg. *Caloscordum* (Herb.) R.M. Fritsch [6, 8, 10]. The position of the EAL was further confirmed by Yang et al. [20],which is sister to the ENAL,while only 3 species of the ENAL distributed in China. The phylogenetic position of *A. funckiifolium* has remained unclear due to its long history of being uncollected, and previous studies have placed it withinthe EAL based on morphological characteristics [1, 20].

Although the monophyly of subg. *Anguinum* and the sister relationship between these two alliances are well-supported, species delimitation within each alliance, particularly in the EAL, remains contentious. Analyses of ITS and cpDNA markers initially revealed that *A. prattii* is sister to *A. nanodes* and also showed that *A. ovalifolium* clusters with its varieties *A. ovalifolium* var. *leuconeurum* and *A. ovalifolium* var. *cordifolium* [1, 23]. However, these relationships proved unstable, collapsing into a mosaic-like topological complexity. In addition, the morphological characteristics and phylogenetic position of *A. funckiifolium* have remained unclear due to its long history of being uncollected. Divergence time inferred from *ITS* and cpDNA markers, as well as the complete plastome, consistently indicate that the EAL represents an early-diverging lineage [1, 20]. The ancestral taxa of this lineage gave rise to a series of species now distributed across the HHMs region and surrounding areas within a relatively short timespan of several million years [1, 20]. Notably, as an early-diverging lineage, the EAL not only exhibits unstable interspecific relationships, but also displays widespread phylogenetic incongruence between nuclear and chloroplast gene trees [1, 34], potentially attributable to interspecific gene flow within the lineage. Current studies are constrained by incomplete taxon sampling, limited integration of morphological data, and the intrinsic limitations of relying solely on the single or several molecular markers. Consequently, these studies lack direct evidence for the causes of nuclear–plastid discordance within the EAL, and most conclusions remain speculative. Future research should aim to collect more samples of the EAL, integrate multidimensional datasets, and apply phylogenomic approaches to resolve species relationships and explore the underlying drivers of phylogenetic incongruences within the EAL.

Phylogenomic data possess numerous informative characters and have been widely used for phylogenetic analyses in recent years [35–43]. Phylogenomic approaches also have been efficiently used for explaining the causes of phylogenetic conflicts, and similar data and methods also have been used for phylogenetic reconstruction and evolutionary analyses in species of *Allium* [15, 16, 34, 44]. Here, we first investigate the monophyly of the EAL based on 57 complete chloroplast genomes, which covered all 10 recognized species within subg. *Anguinum* and two additional varieties. Based on this phylogenetic framework, we further incorporated 45 transcriptomes and 45 plastomes from multiple populations of all EAL species and related taxa to perform comprehensive phylogenetic analyses, where the transcriptome data were generated specifically to recover single-copy nuclear genes (SCGs) and low-copy nuclear genes (LCGs). Morphological data were also integrated to (i) reconstruct the species phylogeny of the EAL within subgenus *Anguinum* and analyze their relationships, and (ii) investigate the mechanisms underlying the observed phylogenetic conflicts within the lineage.

Materials and methods

Sampling and morphological anatomy

Over the last five years, we conducted extensive field surveys and collected plant materials from multiple populations of species in subg. *Anguinum* (Table S1). All samples were collected from the wild without the need for specific permission. Voucher specimens were deposited at the Herbarium of Sichuan University (SZ), and specimen numbers are provided in Table S1. All plant materials were identified taxonomically by Dr. Deng-Feng Xie, encompassing multiple individuals of all target species within the EAL. To provide an overview of the analytical framework used in this study, we summarized the entire workflow in a flowchart (Fig. 1), including sampling, plastome assembly, phylogenetic reconstruction, and conflict analysis. Freshly collected leaf specimens were flash-frozen in liquid nitrogen and maintained at −80 °C until RNA isolation and transcriptome sequencing. Meanwhile, the corresponding leaf samples were also used for DNA extraction and whole genome resequencing. Floral tissue samples from various plant organs were fixed in FAA (formalin-acetic acid–alcohol) solution to facilitate comprehensive morphological and anatomical examinations. In this study, a total of 45 transcriptome profiles were generated, including 42 newly sequenced samples and 3 obtained from NCBI (Table S2). In addition, 57 complete chloroplast genome datasets were generated, including 45 newly sequenced samples and 12 publicly available datasets from NCBI (Table S3). We selected *Allium neriniflorum* as the outgroup. To visualize the morphological differences among the EAL species, we collected plant images for each species in the field and captured detailed microscopic images of their

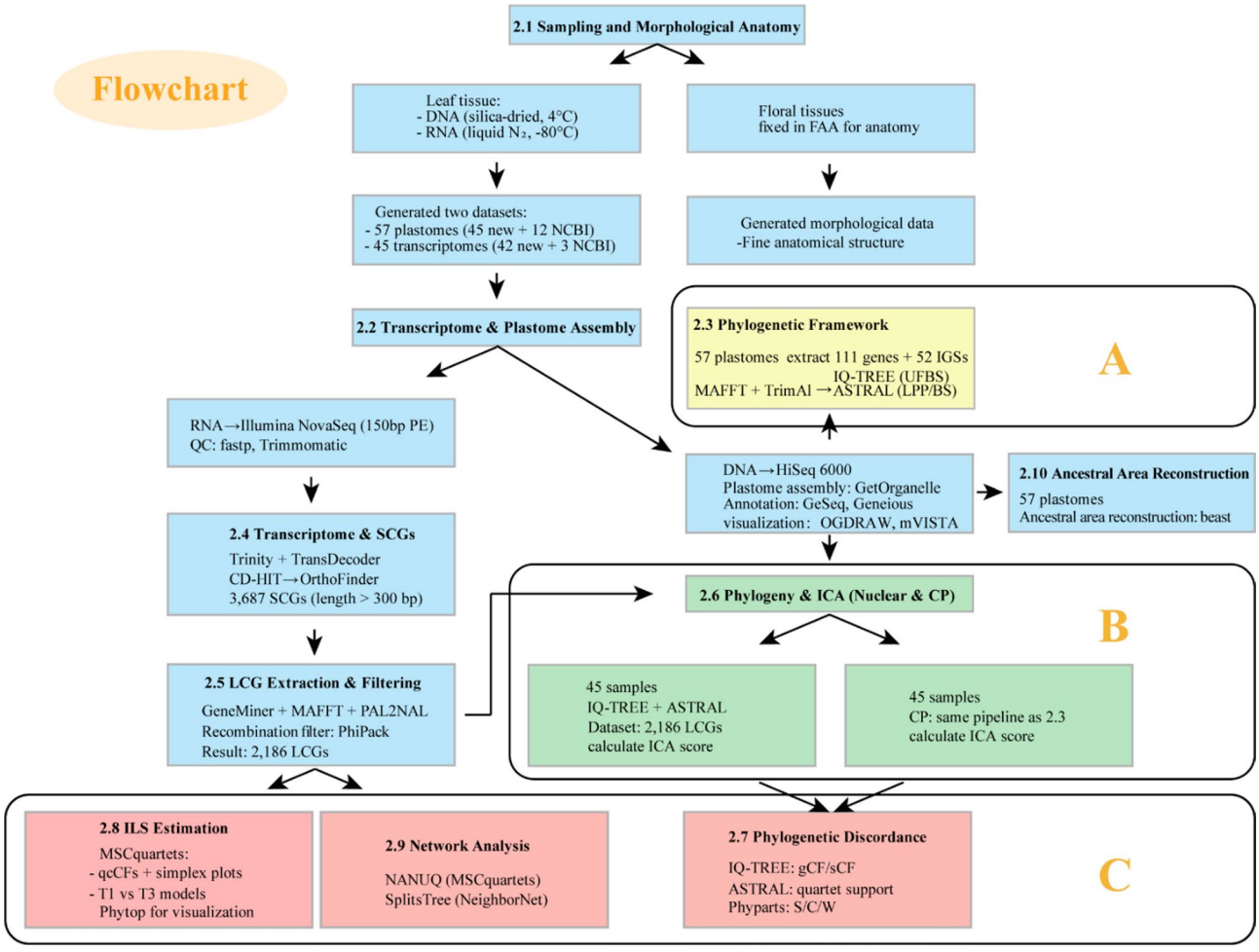

Fig. 1 Research workflow diagram. (**A**) Phylogenetic framework reconstruction of subg. *Anguinum* (highlighted in yellow); (**B**) Phylogenetic analysis of the EAL (highlighted in green); (**C**) Investigation of phylogenetic conflicts and their sources within the EAL (highlighted in red)

flower and leaf morphology using a stereomicroscope (Nikon, SMZ25).

Transcriptome sequencing, whole-genome resequencing and plastome assembly

Transcriptomic sequencing started with RNA isolation using the RNAprep Pure Plant Kit (Tiangen Biotech), followed by rigorous quality control assessments via a NanoPhotometer spectrophotometer (IMPLEN) and an Agilent 2100 Bioanalyzer. Sequencing libraries were constructed with the NEB Next Ultra™ RNA Library Prep Kit (Illumina) and subjected to 150 bp paired-end sequencing on an Illumina NovaSeq 6000 platform. Raw sequencing data underwent preprocessing with fastP [45] and Trimmomatic v0.36 [46] to eliminate adapter contamination and low-quality sequences, generating refined reads for transcriptomic assembly and analytical workflows. For the whole-genome resequencing, total genomic DNA was extracted from fresh or silica gel-dried materials using the CTAB method [47], followed by whole-genome resequencing on the Illumina HiSeq 6000 platform (Novogene, Beijing, China). After quality assessment and filtering using fastp, high-quality reads were assembled into plastomes using GetOrganelle v1.7.7.1 [48], which were subsequently annotated using GeSeq [49]. Finally, we manually verified and refined the annotated plastomes using Geneious v11.0.5 [50]. The annotated plastid genomes of 36 individuals from the EAL were used to construct visual genome maps and perform comparative analyses using the online programs OGDRAW and mVISTA [51, 52]. All raw sequencing data of transcriptomes have been securely archived in the Genome Sequence Archive (GSA) at the National Genomics Data Center (NGDC), with detailed accession identifiers cataloged in Table S2. Additionally, all assembled plastome genome sequences have been deposited in the NCBI database, and their corresponding accession numbers are listed in Table S3.

Phylogenetic framework construction of subgenus *Anguinum*

A total of 57 assembled complete chloroplast genomes were used for the phylogenetic reconstruction of subg. *Anguinum*. We extracted both genes and intergenic spacers (IGSs) from the plastomes, identifying 111 genes and 52 IGSs in total for subsequent tree-building. Sequence alignment was performed using MAFFT v7.505 (default parameters), followed by trimming with TrimAl v1.2 [53] under the '-gt 0.9' threshold. The concatenated supermatrix of 163 chloroplast sequences was generated using PhyloSuite [54]. Phylogenetic reconstruction employed two complementary approaches: (1) concatenated maximum likelihood (ML) analysis and (2) coalescent-based species tree inference, to ensure topological robustness through mutual validation. The maximum likelihood (ML) method was used in concatenation analysis employing IQ-TREE v1.6.8 [55]. The best partitioning scheme for each codon position in each gene was determined using ModelFinder [56], and branch support was assessed with 1,000 Ultra-Fast bootstrap (UFBS) replicates (parameter '-bb 1,000'). The coalescent-based approach was further used to estimate the species tree. Gene trees were generated using IQ-TREE v1.6.8, with 1000 ultrafast bootstraps and the best-fitting model identified with ModelFinder. ASTRAL v 5.6.3 was used to infer the species trees [57]. In ASTRAL-III, the best ML gene trees were used to infer the species tree and two methods were used to infer the species with LPP and BS values: (i) comprehensive gene tree integration applying local posterior probabilities (LPP) for nodal confidence evaluation under parameter configuration '-t 3'; and (ii) species tree inference utilizing optimized maximum likelihood gene trees with multilocus branch support assessed through 1,000 bootstrap replicates (parameters: -i -b -r) [58].

Transcriptome assembly and single-copy genes identification

To establish a phylogenetically robust dataset, we developed a multi-stage bioinformatic workflow that integrated systematic filtering and curation protocols. The analysis pipeline began with *de novo* transcriptome assembly for all 10 samples in subg. *Anguinum* using Trinity v2.8.4 [59], including two individuals from *A. prattii*, two from *A. ovalifolium*, and one representative per specimen for all other species (Table S4). Candidate coding sequences (CDSs) were predicted via TransDecoder (http://transdecoder.sourceforge.net/), followed by redundancy reduction through CD-HIT 4.6.1 [60] using a stringent similarity threshold (-c 0.98). Orthogroup inference was then performed using OrthoFinder v2.5.2 [61] to perform the orthogroup search. Only the single-copy orthologues with length > 300 bp were retained, which resulted in a final set of 3,687 phylogenetically informative SCGs and was used for subsequent sequence capture.

Assembly of captured sequence, alignment, and filtering

The software GeneMiner [62] was used to assemble low-copy genes (LCGs) from the capture sequenced quality-filtered reads. The sequences of the above identified 3,687 SCGs from each species were used as target input file (references) for GeneMiner, with the extraction procedure running following the parameters as: "-rtfa refs (3,687 SCGs) -min 300 -max 5,000 -limit_count auto -t 4 -bn 20 -o outfile". The MAFFT v 7.429 [63] was then used to align amino acid (AA) sequences, and the corresponding codon alignments were converted from the AA alignments using PAL2NAL v 14.0 [64]. To ensure data quality, we removed any aligned loci with more than 20% missing data, as well as individual DNA sequences that were shorter than 300 bp or contained more than 50% gaps. Moreover, we detected potential genetic recombination in the alignments, as it can bias the inference of the species tree when using coalescent methods [65]. We used PhiPack v 1.1 [66] to calculate the pairwise homoplasy index (Φ) for recombination, considering an alignment with a *P*-value of less than 0.05 as significant. These significant alignments were removed, and the remaining rigorous data were used for subsequent analyses, which resulted in 2,186 LCGs shared by all 45 samples.

Phylogenetic reconstruction of EAL and ICA scores calculation

Given that our study focuses on the EAL, we refer to species from another lineage simply as 'The ENAL' in the subsequent analyses. The above 2,186 LCGs dataset from 45 samples was used to reconstruct the nuclear phylogeny of species in the EAL. Concatenation-based maximum likelihood (ML) and coalescent-based species tree inference were used for phylogenetic reconstruction, applying standardized protocols consistent with the phylogenetic framework reconstruction (see Sect. 2.3). The concatenated ML topology was reconstructed via IQ-TREE v1.6.8 with ultra-fast bootstrap support (UFBS), while coalescent-based species trees were generated using ASTRAL-III version 5.6.3 (BS and LPP support values). For the 45 plastome datasets corresponding to the transcriptomic samples, a total of 111 genes and 52 intergenic spacers (IGSs) were identified and extracted. Subsequent analytical steps also followed a standardized workflow identical to that used for phylogenetic framework construction (see Sect. 2.3).

We used phyparts v0.01 [67, 68] to compare gene tree and species tree topologies based on both the transcriptomic and plastome datasets, to quantify phylogenetic concordance. The Internode Certainty All (ICA) values were calculated to evaluate nodal support heterogeneity.

ICA values near 1 indicate strong internodal consensus, values near 0 suggest equiprobable bipartition conflicts, and values less than 0 signify the presence of conflicting bipartitions with certain frequencies.

Phylogenetic discordance analysis

To thoroughly assess phylogenetic discordances among gene trees as well as between nuclear and plastid species trees, the study employed several approaches to evaluate phylogenetic discordances based on multiple phylogenetic trees derived from Sect. 2.6. Using nuclear and plastid gene datasets, the following phylogenetic trees were inferred: (i) a concatenated maximum likelihood (ML) tree constructed in IQ-TREE (with UFBS support values) and (ii) a species tree inferred via ASTRAL-III version 5.6.3 (with BS and LPP support values).

First, the gene concordance factor (gCF) and site concordance factor (sCF) were calculated using IQ-TREE version 2.1.3 [55]. The gCF measures the proportion of gene trees that include a specific branch, whereas the sCF quantifies the proportion of alignment sites that support the same branch [69]. For further details on the methodology and implementation, refer to the official IQ-TREE website (https://www.iqtree.org). Secondly, we applied ASTRAL-III with the '-t 2' parameter, which computes three quartet support values (q1, q2, q3) to measure the consistency between gene trees and the inferred species tree. These three values show quartet support for the main topology, the first alternative, and the second alternative, respectively. To further evaluate the phylogenetic discordance, phyparts v0.01 was used to quantify node support on the species tree. This approach provides three values for evaluating internal nodes: (i) Concordance (S); (ii) Top Conflict (C), and (iii) Other Conflict (W), providing a detailed view of phylogenetic conflicts within the dataset. The analysis was conducted using the phypartspiecharts.py script (available at https://github.com/mossmatters/phyloscripts/blob/master/phypartspiecharts).

Incomplete lineage sorting estimation

To investigate whether the discordance between nuclear and plastid phylogenies could be attributed to incomplete lineage sorting (ILS), we utilized a method that summarizes quartets of gene trees [70]. This approach involves computing the quartet count concordance factors (qcCFs) and deriving *P*-values to evaluate the compatibility of gene trees with a species tree [71]. Using the function quartetTreeTestInd from the R package 'MSCquartets' [70], we calculated qcCFs for all 2,186 LCG trees derived from 45 individuals representing 6 species and 2 varieties. These qcCFs were then used to generate simplex plots within MSCquartets, employing two models: one with a known species tree (model T1) and another without a species tree hypothesis (model T3). For the T1 model, we incorporated the 45-taxon species tree inferred from ASTRAL-III. In the simplex plots, each quartet count vector was represented by color-coded symbols, with red and blue indicating rejection or failure to reject the test at specified alpha levels (set at 0.01, 0.001, 1e-04, and 1e-06). Symbols closer to the centroid suggest ILS, while those nearer to the vertices imply alternative mechanisms, such as introgression.

To further investigate and visualize signals of ILS and introgression/hybridization (IH) across lineages, we employed the software Phytop [72]. This tool provides a reliable quantification of ILS and IH levels among lineages and effectively visualizes these patterns based on ASTRAL-generated gene trees. Each gene tree was reconstructed using IQ-TREE v1.6.8, following the procedure outlined in Sect. 2.4. The species tree was then inferred using ASTRAL-III v5.6.3 with the parameter '-t 2'. Finally, the inferred species tree was used as input in Phytop to compute and visualize ILS and IH values.

Network analysis

To investigate the hypothesis that reticulate evolution, caused by hybridization or introgression, contributes to observed discordances, we further inferred a hybridization network employing NeighborNet using the Quartet distance (NANUQ) method [73]. Empirical quartet counts were calculated from 2,186 LCG trees of 45 taxa, and the network quartet distances between taxa was determined using the NANUQ function in the R package MSCquartets. The analysis was conducted under the 'T3 model' with the parameter "$\alpha = 0.01$" [70]. Subsequently, a split graph was generated for visualization using the NeighborNet algorithm in SplitsTree4 version 4.15.1 [74], based on the quartet distances between taxa.

Ancestral area reconstruction

To explore the relationship between the phylogenetic structure and conflicts of the EAL and its biogeographic history, we used six regions based on the paleogeographic and climatic evidence, geographic distribution and floral composition for ancestral area reconstruction [1, 18, 22, 23, 75]: (A) eastern Himalayas, (B) Hengduan Mountains, (C) Taihang Mountains, (D) Northeast Asia, (E) North America, and (F) Europe. We conducted ancestral state reconstruction using Bayesian discrete trait analysis implemented in BEAST v1.10.4 [76], based on the plastid dataset from 57 samples. Sampling localities were coded as discrete geographic states, with six regions defined according to population distribution. The analysis was performed under the GTR substitution model with empirical base frequencies and gamma-distributed rate heterogeneity across sites using four rate categories (GTR + Gamma). A strict molecular clock was assumed. The discrete trait analysis followed the Bayesian

stochastic search variable selection (BSSVS) framework to identify well-supported migration pathways between regions. Markov chain Monte Carlo (MCMC) chains were run for 250 million generations, sampling every 250,000 steps, with the first 10% discarded as burn-in. Convergence and effective sample sizes (ESS > 200) were assessed in Tracer v1.7.11 [76]. The maximum clade credibility (MCC) tree, including node information (inferred ancestral distributions), was summarized using TreeAnnotator and visualized in FigTree [76, 77]. Divergence time estimates (Mya) for subg. *Anguinum*, as well as for the EAL and ENAL, are labeled at the corresponding nodes based on the results of Yang et al. [20].

Results

Phylogenetic framework of subgenus *Anguinum*

We first confirmed the phylogenetic framework of subg. *Anguinum* using 57 plastomes, which cover all recognized species within the subgenus. The phylogenetic trees constructed using both concatenation and coalescent-based methods yielded consistent topological structures, which consistently support that subg. *Anguinum* can be stably divided into two lineages (Fig. S2, UFBS/BS = 100, LPP = 1.00): the EAL and the ENAL. The EAL includes *A. funckiifolium*, *A. ovalifolium*, *A. prattii*, *A. nanodes*, *A. ovalifolium* var. *leuconeurum*, and *A. ovalifolium* var. *cordifolium*, while the ENAL comprises *A. victorialis*, *A. listera*, *A. tricoccum*, *A. microdictyon*, *A. ochotense*, and *A. ulleungense*. This well-supported result provides a robust and clear foundation for our subsequent analyses of the phylogeny and phylogenetic conflicts within the EAL.

Morphological analysis

Through field investigations, literature review and detailed anatomical experiments, we collected and summarized the morphological characteristics of all species and varieties of the EAL. In general, species of the EAL have cylindrical to conical bulbs; the outer skin is fibrous and exhibits a distinct reticulate pattern. The leaves are usually 2 in number, arranged oppositely, rarely 1 or 3, linear to ovate in shape, and often tapering at the base to form a petiole. The outer perianth segments are narrower than the inner ones and boat-shaped. The filaments are entire; the ovary tapers at the base into a short stalk, with one ovule per locule. Among them, *A. ovalifolium* is characterized by having two leaves that are lanceolate-oblong to ovate-oblong in shape, with a rounded to shallow cordate base. Its perianth segments have obtuse or notched apices, often bearing irregular small teeth (Fig. 2, A1-F1). *A. funckiifolium* is morphologically very similar to *A. ovalifolium*, but it can be distinguished by having only a single leaf and a scape that forms a Y-shaped structure with the petiole (Fig. 2, A6-F6). The two varieties of *A. ovalifolium*, namely *A. ovalifolium* var. *cordifolium* and *A. ovalifolium* var. *leuconeurum*, closely resemble the typical form of *A. ovalifolium*. However, *A. ovalifolium* var. *cordifolium* is distinguished by its deeply cordate leaf base and often prominently undulate margins (Fig. 2, A3-F3),

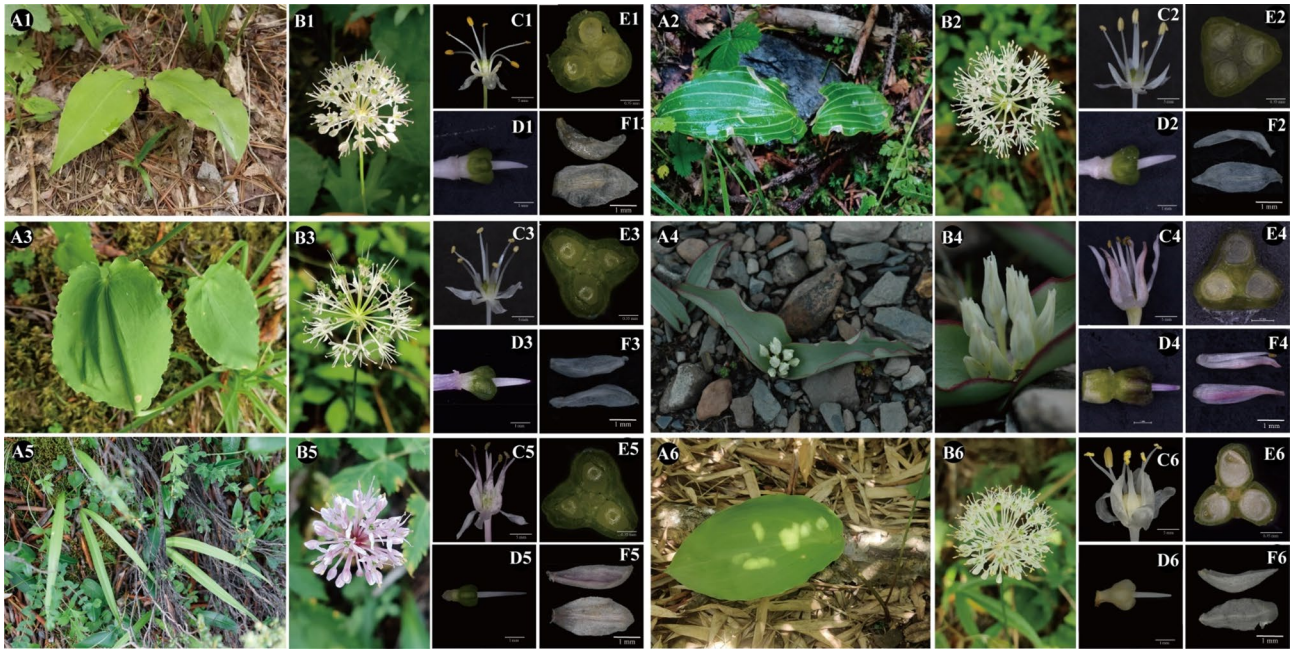

Fig. 2 Morphological comparison of species of the EAL. **A1**-**F1** *Allium ovalifolium*; **A2**-**F2** *Allium ovalifolium var. leuconeurum*; **A3**-**F3** *Allium ovalifolium var. cordifolium*; **A4**-**F4** *Allium nanodes*; **A5**-**F5** *Allium prattii*; **A6**-**F6** *Allium funckiifolium*. **A1**-**A6** Overall plant morphology; **B1**-**B6** Inflorescence morphology; **C1**-**C6** Flower morphology; **D1**-**D6** Ovary and stigma morphology; **E1**-**E6** Cross-section of the ovary; **F1**-**F6** Morphology of the inner and outer tepals (upper row: outer tepals; lower row: inner tepals)

while *A. ovalifolium* var. *leuconeurum* is easily identified by its white-veined leaves (Fig. 2, A2-F2). *A. prattii* features 2 linear or linear-lanceolate leaves with a cuneate base that slightly extends along the petiole, and its perianth segments are typically pink to purplish-red (Fig. 2, A5-F5). *A. nanodes* is unique within the genus, bearing 2 leaves that often curve backward in a sickle-like manner and a very short scape, setting it apart from other species (Fig. 2, A4-F4). Furthermore, we observed intermediate morphological forms between species in certain populations, including transitional forms between *A. prattii* and *A. nanodes* (Fig. S3B) and between *A. ovalifolium* and *A. prattii* (Fig. S3E).

Transcriptome dataset and phylogenetic analyses

For the 45 transcriptome samples, the number of valid reads (clean reads) ranged from 43,045,294 to 130,018,252 and the GC content ranged from 39.99% to 43.82% (Table S2). After rigorous screening, 2,186 LCGs were finally identified from 45 transcriptomes and used for phylogenetic analysis. Our analyses consistently support that subg. *Anguinum* is polytomous and splits into two lineages (the EAL and the ENAL) with full support in multiple methods (UFBS/BS = 100, LPP = 1.00) (Figs. 3 and S4). In all phylogenetic trees generated, *A. nanodes*, *A. prattii*, *A. ovalifolium*, *A. ovalifolium* var. *cordifolium*, *A. ovalifolium* var. *leuconeurum*, and *A. funckiifolium* form the EAL (LPP = 1.00; UFBS/BS = 100 at the node), while the remaining species constitute the ENAL (LPP = 1.00; UFBS/BS = 100 at the node) (Figs. 3 and S4). Furthermore, individuals within the EAL were clustered into four sublineages, which exhibited distinct patterns of geographical distribution (Fig. 3). The sublineage I mainly distributed in the HDMs, within it, all individuals of *A. nanodes* (1–4) form a monophyletic clade, which is sister to a group of intermingled individuals of *A. prattii* (1–4) and *A. ovalifolium* (1–3). Sublineage II distributed in the Northwestern HDMs, it is a monophyletic branch composed entirely of individuals of *A. prattii* (5–11). Sublineage III is a nearly monophyletic branch formed by *A. ovalifolium* (4, 5) and its two varieties, *A. ovalifolium* var. *cordifolium* (1–4) and *A. ovalifolium* var. *leuconeurum* (1, 2). All samples of this sublineage exhibit a clustered distribution pattern along the northwestern margin of the Sichuan Basin, located in the northeastern section of the HDMs. However, sublineage IV is a complex branch formed by individuals of multiple species, it distributed in Qinling-Daba mountains (QDMs) and included two individuals of *A. prattii* (12, 13), three individuals of *A. ovalifolium* (8–10), and two individuals of *A.*

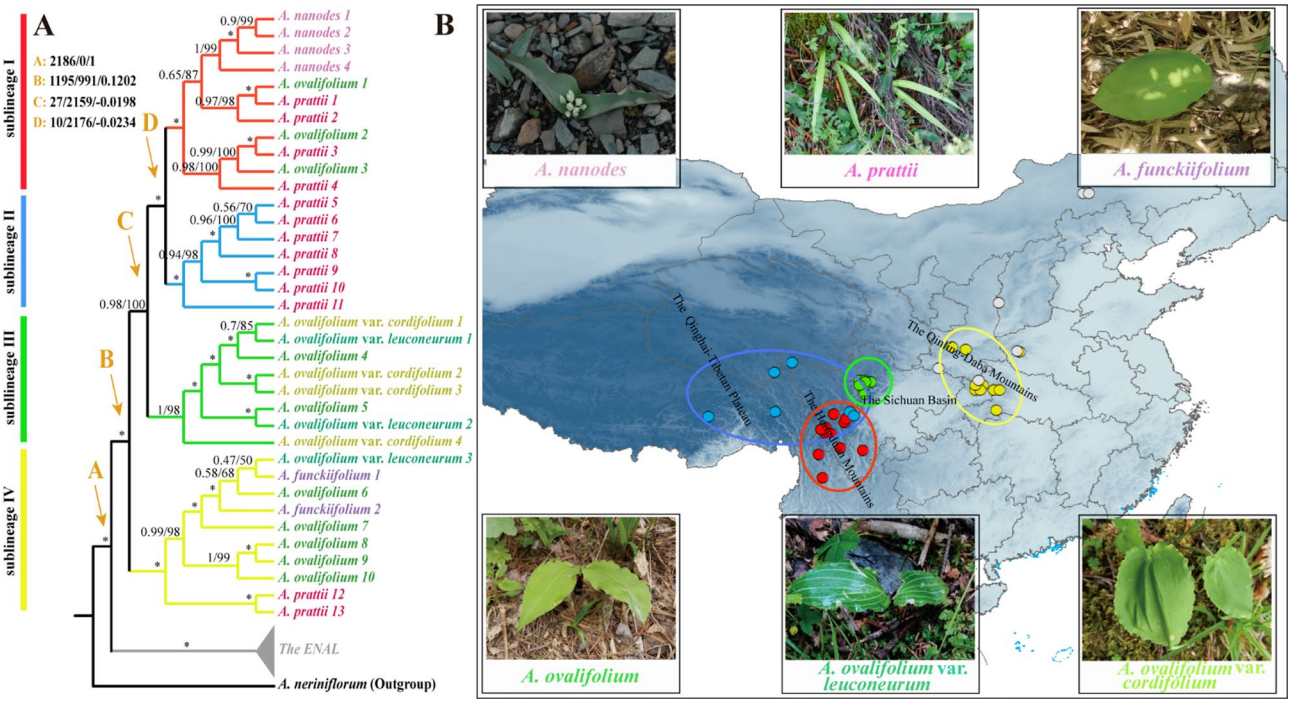

Fig. 3 Species tree inferred from 2,186 single-copy genes (LCGs) from 45 samples using ASTRAL-III, along with the distribution of individual sampling locations. (**A**) Support values from different methods [BS/ASTRAL-III and LPP/ASTRAL-III] are indicated above the branches. An asterisk (*) indicates maximum support (e.g., 100%). Different sublineages within the EAL are labeled with distinct colors on the branches, and different species are marked with distinct colors for their species names, with all individuals of the same species marked in the same color. The figure provides the number of gene trees in concordance/conflict with the displayed node, as well as the "internode certainty all" (ICA) scores for each node (A to E). (**B**) The sampling locations of all individuals are plotted based on their geographic coordinates, with individuals within each sublineage labeled and circled in the corresponding branch color

funckiifolium, which showed nested relationship with *A. ovalifolium* (6, 7) and its variety *A. ovalifolium* var. *leuconeurum* (3) (Figs. 3 and S4). Phylogenies inferred from concatenated and coalescent methods exhibited slight phylogenetic discordance, however, all four sublineages were consistently resolved across both analytical frameworks (Fig. 4). Regarding ICA scores, substantial conflicts among gene trees were observed at major nodes. All LCGs supported the division of the EAL and the ENAL (Fig. 3, node A). Among the 2,186 LCGs, the divergence node of the EAL was supported by 1,195 genes (ICA = 0.1202). Within the lineage, pronounced gene tree discordance was observed at critical branch nodes, specifically at the ancestral nodes of sublineages I and II (node D, ICA = −0.0234) and the common ancestor of sublineages I–III (node C, ICA = −0.0198). These nodes exhibited low phylogenetic support, with fewer than 30 LCGs recovering the inferred relationships.

Plastid features and phylogenetic analyses

For the plastid genome analysis, the 57 plastid genomes ranged in length from 153,124 bp to 154,610 bp (Table S3). Among them, the 36 individuals from the EAL were further processed for genome visualization and comparative analysis (Figs. S5 and S6). The results indicated that the EAL species exhibit highly conserved chloroplast genome structures, with no obvious differences among species (Fig. S5). Notably, the mVISTA analysis revealed that multiple populations of *A. prattii*, *A. ovalifolium*, and *A. nanodes* share strikingly similar patterns in several regions, including the intergenic spacers between *trnS-GGA* and *rps4*, *psbE* and *petL*, and *ccsA* and *ndhD*, all of which show distinct deletion events (Fig. S6). To capture as much information as possible about all reliable variant sites in the plastid genomes, we extracted and rigorously screened 111 genes and 52 intergenic regions (IGSs) for phylogenetic analysis. Phylogenomic analysis based on the dataset comprising 45 individual samples yielded a topological structure consistent with that of the 57-sample dataset, with the two major lineages robustly resolved under both concatenation and coalescent-based methods. Compared to the LCGs dataset, the EAL exhibited greater phylogenetic complexity at the plastid level, with more pronounced nested relationships among species (Fig. S7). Within the EAL, all individuals formed two well-defined sublineages (sublineages i and ii) (Figs. 4 and S7), which were consistent with the geographic locations. Individuals of sublineage i main distributed in the

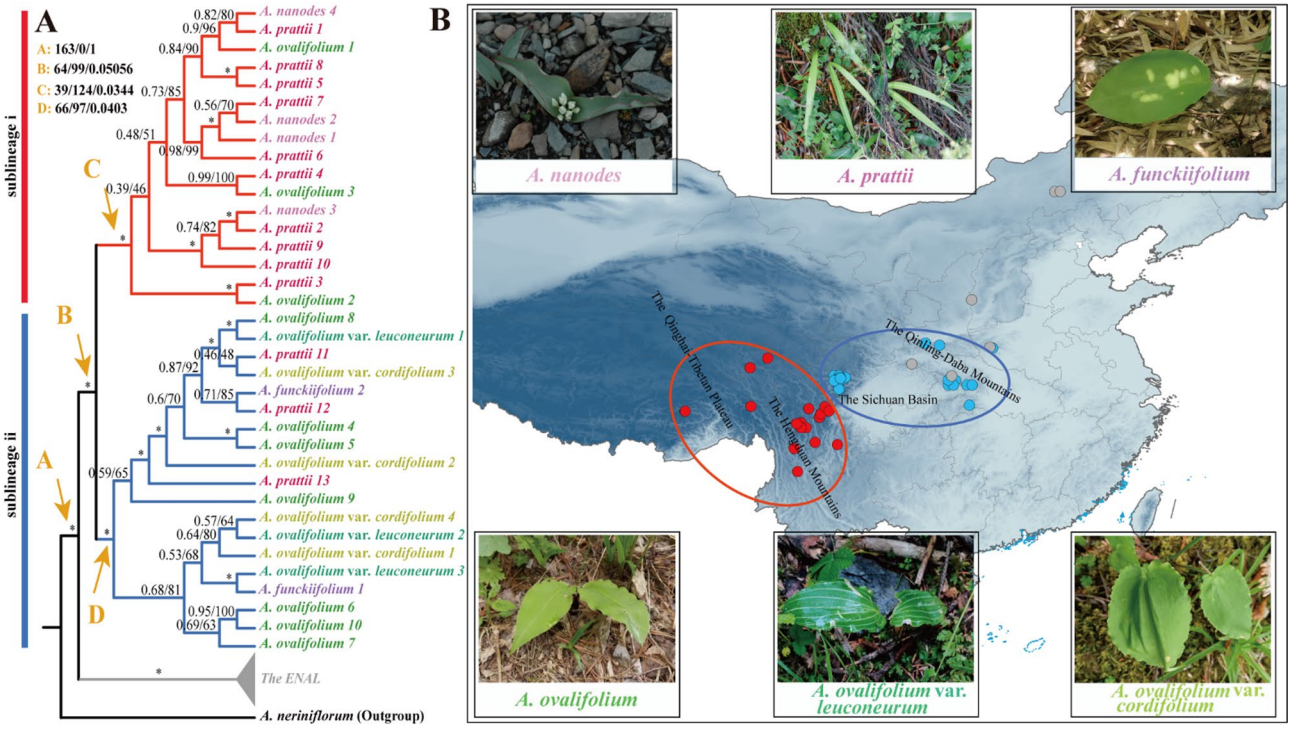

Fig. 4 Species tree inferred from 163 plastid sequences (including 111 genes and 52 intergenic regions) from 45 samples using ASTRAL-III, along with the distribution of individual sampling locations. (**A**) Support values from different methods [BS/ASTRAL-III and LPP/ASTRAL-III] are indicated above the branches. An asterisk (*) indicates maximum support (e.g., 100%). Different sublineages within the EAL are labeled with distinct colors on the branches, and different species are marked with distinct colors for their species names, with all individuals of the same species marked in the same color. The figure provides the number of gene trees in concordance/conflict with the displayed node, as well as the "internode certainty all" (ICA) scores for each node (A to E). (**B**) The sampling locations of all individuals are plotted based on their geographic coordinates, with individuals within each sublineage labeled and circled in the corresponding branch color

HDMs, and composed by *A. nanodes* (1–4), *A. prattii* (1–10), and *A. ovalifolium* (1–3), which exhibited nested relationships. Samples of sublineage ii mainly distributed in northwestern adjacent of the Sichuan Basin and QDMs, including *A. funckiifolium* (1, 2), *A. prattii* (11–13), *A. ovalifolium* (4–10), and its two varieties, *A. ovalifolium* var. *cordifolium* (1–4) and *A. ovalifolium* var. *leuconeurum* (1–3). All samples of this group clustered with each other and displayed complex relationships. In addition, phylogenetic topologies inferred from concatenated and coalescent approaches detected extensive conflicts within the EAL (Fig. S7). Regarding ICA scores, significant conflicts among gene trees were detected at major nodes. Among the 163 plastid sequences, only 64 sequences (ICA = 0.05056) supported the divergence node of the EAL (Fig. S7B, node B). Notably, more pronounced conflicts were observed within the lineage, with only 39 sequences (ICA = 0.0344) supporting the divergence of sublineage i (Fig. S7C, node D) and 66 sequences (ICA = 0.0403) supporting the divergence of sublineage ii (Fig. S7B, node E).

Phylogenetic discordance analyses

To elucidate the underlying causes of the observed incongruences between gene trees as well as between nuclear and plastid phylogenies in the EAL, we conducted comprehensive analyses utilizing transcriptome data and plastid sequences obtained from the 45 samples. In the phylogenetic trees constructed based on LCGs (Fig. S4) and plastid sequences (Fig. S7), extensive phylogenetic conflicts were detected within the internal branches of the EAL. In the LCG tree, sublineage I and sublineage II formed sister clades, encompassing multiple populations including *A. nanodes*, *A. prattii*, and *A. ovalifolium*. In contrast, these individuals were clustered and nested within a single clade (sublineage i) in the plastid phylogeny. For example, in the LCG tree, multiple individuals of *A. nanodes* (1–4) and *A. prattii* (5–11) formed distinct clades, whereas in the plastid tree, they were intermingled and failed to form independent clades. A similar pattern was observed for sublineage III and sublineage IV in the LCG tree, which included multiple populations of *A. funckiifolium*, *A. prattii*, *A. ovalifolium*, and its two varieties, while these individuals were intermingled and nested within a large clade (sublineage ii) in the plastid tree (Figs. 3, 4 and 5).

Quartet-based analysis of the LCGs-based tree and plastid tree consistently showed a huge proportion of loci did not support the divergent nodes, indicating a notable level of phylogenetic inconsistency (Fig. 5). High

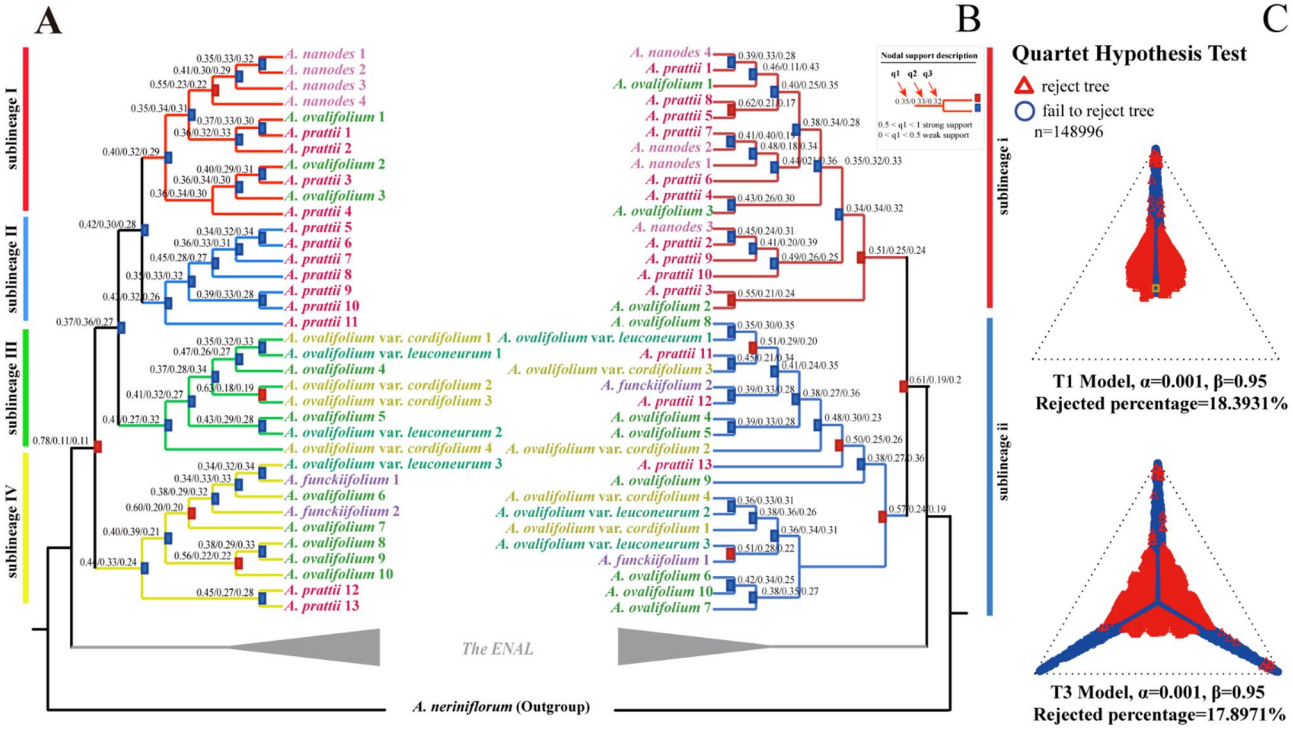

Fig. 5 Cytonuclear and genealogical discordances along with the detection of ILS. Species trees are inferred from (**A**) 2,186 LCGs and (**B**) 163 plastid sequences. Values q1/q2/q3 of each node are shown above the branches. In the heatmap coloring of nodes and branches, red represents high support strength for the topology of the node, indicating high reliability of the node, while blue represents low support strength for the topology of the node, indicating low reliability of the node. Different species are marked with distinct colors for their species names. (**C**) Simplex plots of qcCFs under the MSC model of ILS with T1 and T3 models (α = 0.001, other levels are provided in Fig. S11) for the 2,186 LCG trees. Red triangles in the plot indicate rejection of the MSC on the species tree, blue circles indicate probable ILS

proportion of main topology (q1 > 0.6) was detected at the divergence nodes of the EAL (Fig. 5, node A, q1/q2/q3 = 0.78/0.11/0.11). Given the divergence nodes of sublineages, significant gene tree conflicts were detected, particularly in the phylogeny reconstructed using the LCG-based dataset, where the main topology constituted less than 50% (q1 < 0.50). In the quartet analysis using PhyParts, the divergence nodes of the two major lineages received consistent support (concordant LCG/plastid sequences = 2186/0). The divergence node of the EAL also received majority support (concordant LCG/plastid sequences = 1195/991). In contrast, nearly all other nodes were marked in red, indicating severe phylogenetic conflicts at these positions (Fig. S8). Furthermore, we evaluated the phylogenetic discordance through calculating the gene concordance factor (gCF) and site concordance factor (sCF) values. In the LCG tree, the gCF values and sCF values ranged from 0.27 to 69.4% (mean = 15.61%) and 32.87–78.26% (mean = 44.70%), respectively (Fig. S9). In the plastid tree, the gCF and sCF values ranged from 0.61 to 46.01% (mean = 11.13%) and 27.91–100% (mean = 72.05%), respectively (Fig. S10). Significant correlations ($P < 0.001$) were observed between both concordance factors (gCF and sCF) and bootstrap support (BS) or branch length, indicating that low BS or branch length values corresponded to the lowest gCF and sCF values in both the LCG and plastid datasets (Fig. 6). Additionally, while more nodes in the LCG and plastid trees exhibited full BS support, their sCF values were lower than their gCF values. A positive correlation was also detected between BS and branch length in both the LCG and plastid trees (Fig. 6D, H).

Intensity of ILS and reticulate evolution

We employed the "T1" and "T3" models in MSCquartets analysis to evaluate the intensity of ILS within the EAL based on 2,186 LCG trees. The results detected a total of 148,996 exhaustive quartets across all taxa. As the rejection thresholds varied from 0.01 to 1e-06, the percentage of rejected trees ranged from 12.07% to 24.19% under the T1 model, while under the T3 model, this proportion ranged from 11.87% to 23.52% (Figs. 5C and S11). In the simplex plots, a large number of quartet concordance factors (qcCFs) were plotted near the T1 or T3 model lines, with blue circles tightly clustered near the vertices. These results indicate that ILS is the primary driver of phylogenetic discordance, while additional mechanisms such as hybridization or introgression also explain some of the phylogenetic conflicts. Phytop was utilized to detect ILS across 12 key nodes within the multi-population species of the EAL. Significant ILS signals (ILS-i) were detected at these 12 nodes, ranging from 32.4% (Fig. 7, node A) to 98.8% (node K). High ILS signals (ILS-i > 60.0%) were detected at all nodes except for node A. Among them, five nodes showed particularly strong ILS signals: nodes I, L, J, and G within sublineage I, and node K within sublineage IV. The ILS values for these nodes, from lowest to highest, were 89.1%, 89.8%, 90.6%, 98.0%, and 98.8%, respectively. Among these five nodes, nodes I, J, and

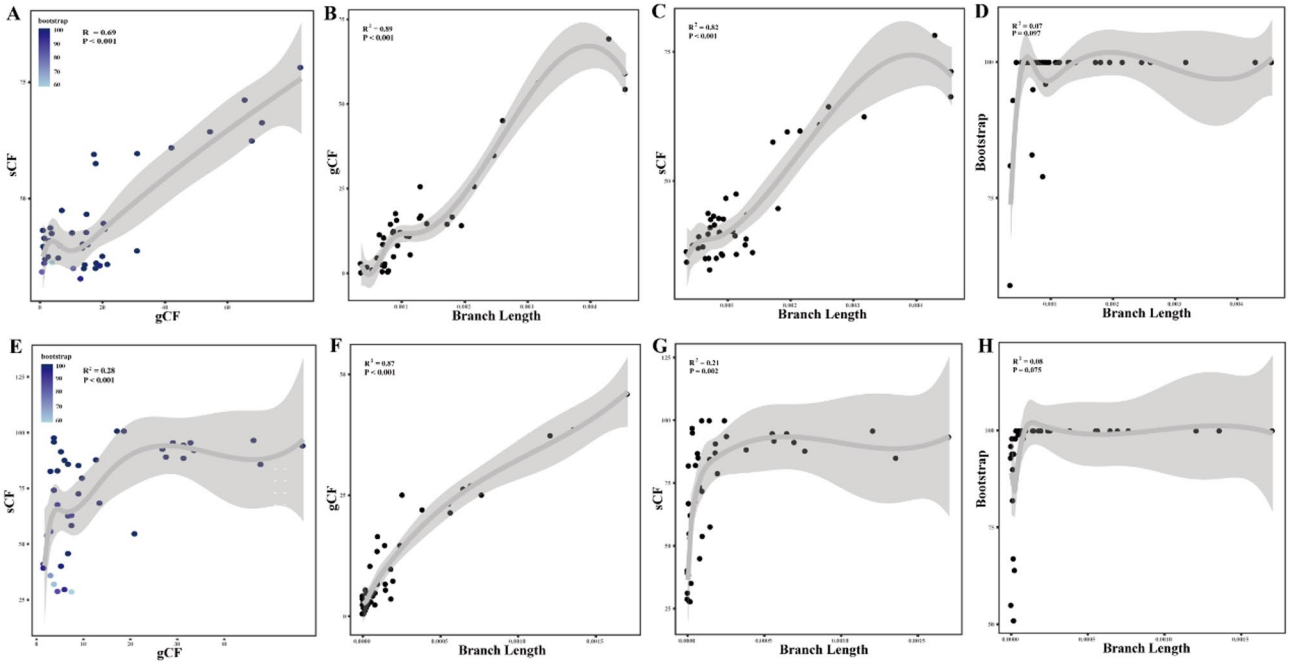

Fig. 6 Connections between two concordance factors (gCF and sCF) and bootstrap (BS) or branch length across all nodes of the phylogenetic tree are examined using (**A**-**D**) the LCGs datasets and (**E**-**H**) plastid sequences datasets

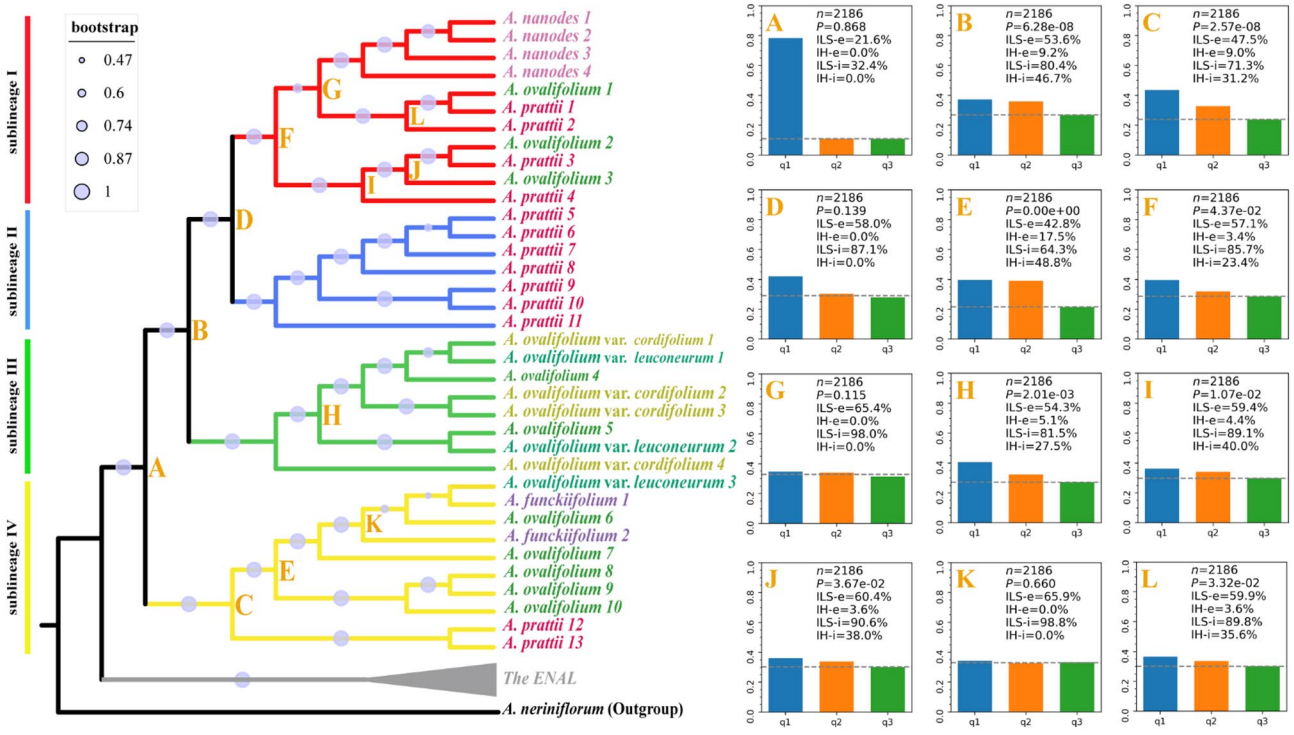

Fig. 7 Detection of the ILS and reticulate evolution signals from species of the EAL. (**A**) Species tree inferred from the 2,186 LCGs of the EAL. Different sublineages within the EAL are labeled with distinct colors on the branches, and different species are marked with distinct colors for their species names, with all individuals of the same species labeled in the same color. Node support values (BS) are represented by circles of different sizes. (**B**) Signals of incomplete lineage sorting (ILS-i) and introgression/hybridization (IH-i) across the nodes (**A**-**L**) are identified and visualized, corresponding to the nodes in the phylogenetic tree

L represent divergence points among partially nested populations of *A. prattii* and *A. ovalifolium*; node G corresponds to the divergence between certain populations of *A. prattii* and *A. ovalifolium* and all populations of *A. nanodes*; while node K represents the divergence between some populations of *A. ovalifolium* (including its varieties) and all populations of *A. funckiifolium*. Additionally, certain introgression/hybridization signals (IH-i) were detected at 8 key nodes (B, C, E, F, H, I, J and L), ranging from 23.4% (Fig. 7, node F) to 48.8% (node E). Significant introgression/hybridization signals were observed at node B (IH-i = 46.7%) and node E (IH-i = 48.8%). Specifically, node B represents the common ancestral node of sublineages I–III and includes all population individuals distributed in the HDMs and surrounding areas, while node E corresponds to the divergence point between *A. ovalifolium* and *A. funckiifolium* within sublineage IV (Fig. 7). These results indicate that ILS is the primary driver of phylogenetic conflicts, while introgression/hybridization also plays an important role. We employed the NANUQ network method to determine whether reticulate evolution has contributed to the phylogenetic conflicts. The analysis results depicted a reticulated topology and revealed a complex network structure (Fig. S12).

Reconstruction of ancestral area

To further validate our hypothesis regarding the underlying mechanisms of the nested phylogenetic relationships within the EAL, we performed ancestral area reconstruction based on 57 plastome samples. By integrating our results with the divergence time estimates from Yang et al., based on plastid genomes, we found that the EAL represents an early-diverging lineage within subg. *Anguinum*, originating around 3.05 Ma (95% HPD: 1.22–5.34 Ma). Its most recent common ancestor likely originated in the HDMs (region B, Fig. S13). The origin of subg. *Anguinum* itself likely began with an early divergence event in the HDMs at approximately 4.85 Ma (95% HPD: 2.08–8.32 Ma). Subsequently, the ancestral lineage of the EAL expanded outward from the HDMs, spreading westward to the Himalayas and the Qinghai–Tibet Plateau, and eastward to the QDMs. Further diversification in these regions gave rise to the various lineages observed within the EAL (Fig. S13).

Discussion

Phylogenetic relationships of the EAL

Phylogenomic data and methods have been demonstrated to be highly effective in exploring species delimitation, phylogenetic reconstruction, and evolutionary

mechanisms [34, 41–43, 78–80]. We here confirmed that the subg. *Anguinum* can be divided into two lineages (EAL and ENL) based on extensive samples from multiple populations. This division is strongly supported (UFBS/BS = 100, LPP = 1.0; Figs. 3 and 4 and S2) and is consistent with previous studies [1, 6, 7, 20]. The EAL further divided four and two sublineages in the LCGs-based tree and the plastid data tree, respectively, which exhibited strong concordance with their geographic distributions (Figs. 3 and 4). Such high concordance between phylogenetic lineages and their geographic distribution may typically be attributed to geographic isolation or adaptive evolution [81, 82]. Given the complex geographic and environmental conditions of species in the EAL [1, 18], we speculate that populations among sublineages are separated by geographical barriers, and gene flow among sublineages is restricted. However, gene flow between species within each sublineage is frequent, which was supported by our results (nodes B, C, E, F, H, I and L in Fig. 7), such gene introgression may obscure species boundaries, resulting in embedded and reticulate phylogenetic topologies of the EAL, as demonstrated by our results (Fig. S12). Similar scenario has been reported in other taxa, such as *Quercus*, *Liriodendron*, and *Picea* [83–85].

In addition, through synthesizing previous literature [18, 75], conducting systematic field surveys, and performing detailed morphological examinations and molecular analyses, we conclusively validated the taxonomic status of two long-disputed varieties of *A. ovalifolium* (*A. ovalifolium* var. *cordifolium* and *A. ovalifolium* var. *leuconeurum*). Specifically, these two varieties differ from the typical *A. ovalifolium* only in leaf characteristics: *A. ovalifolium* var. *cordifolium* has deeply cordate leaf bases (Fig. 2A3), while *A. ovalifolium* var. *leuconeurum* has white leaf veins (Fig. 2A2), whereas the typical *A. ovalifolium* has green veins and rounded or shallowly cordate bases (Fig. 2A1). Ecologically and phenologically, the two varieties are narrow endemics distributed only in the northwestern Sichuan Basin, with ranges overlapping with the widespread *A. ovalifolium*, and they share similar habitats and identical flowering/fruiting periods. Our phylogenetic results further revealed that the two varieties of *A. ovalifolium* do not form distinct monophyletic groups but are nested within *A. ovalifolium* individuals from the same region. All these lines of evidence confirm the taxonomic status of *A. ovalifolium* and its varieties (Fig. 3, sublineage III).

We also discovered a new distribution site of *A. ovalifolium* var. *leuconeurum* in the QDMs, which is sympatric with *A. funckiifolium* and *A. ovalifolium*. The *A. funckiifolium* is distributed only in the QDMs and has long remained uncollected and understudied due to its narrow distribution and unclear record [1, 5, 6, 16, 19, 20, 86]. Our phylogenetic results showed that the two populations of *A. funckiifolium* were phylogenetically embedded within the *A. ovalifolium* and *A. ovalifolium* var. *leuconeurum*, rather than forming a distinct monophyletic clade (Fig. 3, sublineage IV). *A. funckiifolium* differs from *A. ovalifolium* in its solitary larger leaf (16.5–22.8 cm long × 11.3–15.7 cm wide), contrasting with the latter's two smaller leaves (8–15 cm long × 3–7 cm wide), as described in earlier taxonomic treatments [18, 75]. However, based on our field investigations and morphological studies, we observed that both leaf number and size are variable traits in these two species, insufficient to distinguish them taxonomically. Furthermore, the distribution range of *A. funckiifolium* is entirely encompassed by that of *A. ovalifolium*, with which it shares similar habitats and phenology (flowering and fruiting periods). Thus, integrating previous literature with our phylogenetic results, morphological analyses, and observations of habitat and phenology, we propose to classify *A. funckiifolium* as a variety of *A. ovalifolium*, namely *A. ovalifolium* var. *funckiifolium*.

A total of 2,186 orthologous genes were here used for phylogenetic analyses, which effectively eliminated concerns regarding insufficient informative sites in this dataset. Both concatenation and coalescent approaches yielded highly supported species relationships (Figs. 3 and S4). However, our phylogenetic tree detected polytomous phylogenetic results, especially for *A. prattii*, *A. ovalifolium* and *A. nanodes* (Fig. 3, sublineage I). Such nested phylogenetic pattern may be attributed to the following evolutionary processes. Based on previous studies and our field investigations [16, 18, 75], although the three species exhibit markedly different habitats: *A. prattii* grows in shrublands on mountain slopes above 3,000 m elevation, *A. ovalifolium* inhabits shaded understory environments at lower altitudes, and *A. nanodes* is restricted to extreme high-altitude scree slopes. They are sympatrically distributed in the HDMs region, with some populations showing transitional and indistinguishable morphologies (Figs. 3 and S3). Extensive studies have identified HDMs as a biodiversity hotspot characterized by complex topography and unique climatic conditions, and frequent hybridization events have been documented in many taxa, such as *Rhododendron*, *Swertia*, and other *Allium* species [16, 87–92]. Previous research indicates that subg. *Anguinum* originated approximately 4.85 Mya (95% HPD, 2.08–8.32 Ma), representing a recently diverged lineage, with the most recent common ancestor of the EAL emerging around 3.05 Mya (95% HPD, 1.22–5.34 Ma), coinciding with the further uplift of the Qinghai–Tibetan Plateau and the intense orogenic activities in the HDMs during the Late Pliocene [16, 20, 93]. Furthermore, the relatively short branch lengths observed in the phylogenetic tree provide

additional evidence for rapid radiative evolution within this subgenus [16, 20]. Recent studies have demonstrated that rapid evolutionary radiation consistently maintains higher levels of ancestral polymorphism retention compared to gradual speciation events [42, 94–96]. As a stochastic evolutionary process, such retention of ancestral polymorphisms obscures species boundaries and generates nested phylogenetic patterns [16, 97]. In addition, the results of our ancestral area reconstruction indicate that the EAL most likely originated in the HDMs, which is consistent with the inferred origin of subg. *Anguinum* (Fig. S13). This suggests that the species within the EAL may have diversified in situ from their ancestral lineage. Overall, considering the pervasive reticulate evolution observed in subg. *Anguinum* (Fig. S12), the short divergence time of the EAL, the result of ancestral distribution reconstruction, the distinct habitats occupied by *A. prattii*, *A. ovalifolium*, and *A. nanodes*, and their intermixed morphological traits, we hypothesize that the blurred species boundaries and mutually embedded phylogenetic patterns among these three species may be linked to historical geological events and climatic fluctuations in the HDMs region. The climatic fluctuations triggered by the uplift of the Qinghai–Tibet Plateau during the Late Pliocene (3.6–2.588 Mya), along with intense orogenic movements in the HDMs that led to habitat fragmentation, likely drove rapid divergence of ancestral populations of the EAL within their respective microhabitats [98]. The subsequent Pleistocene period (2.588–0.126 Mya), characterized by alternating glacial and interglacial cycles, may have further caused cyclical range expansions and contractions with isolation and contact dynamics of the EAL species [98, 99]. These dynamics likely facilitate recurrent gene flow in contact zones, particularly between *A. prattii* and *A. ovalifolium*. The rapid radiation followed by this cyclical "isolation-contact" dynamic ultimately resulted in the retention of ancestral polymorphisms, and the establishment of complex reticulate relationships through frequent genetic exchange, as evidenced by our MSCquartets analyses (Fig. 7: node F, IH-i = 23.4%, ILS-i = 85.7%; node I, IH-i = 40%, ILS-i = 89.1%; node J, IH-i = 38%, ILS-i = 90.6%; node L, IH-i = 35.6%, ILS-i = 89.8%).

In summary, through synthesizing historical literature, field surveys, phylogenetic reconstructions, and morphological assessments, our results confirm the taxonomic status of *A. ovalifolium* var. *cordifolium* and *A. ovalifolium* var. *leuconeurum* and propose the taxonomic revision of *A. funckiifolium* to *A. ovalifolium* var. *funckiifolium* comb. nov. In addition, our phylogenetic reconstruction of the EAL of *Allium* subg. *Anguinum* revealed a complex pattern of nested relationships. Combined with previous estimates of divergence times, we speculate that this interspecific nesting may be closely related to the uplift of the Qinghai–Tibetan Plateau and climatic fluctuations during the late Pliocene. Such scenario has been documented in other taxa, such as sclerophyllous oaks and *Pedicularis* [100, 101], as well as in other *Allium* species like *sect. Sikkimensia* [97]. Future research should expand taxon sampling and incorporate more comprehensive genomic data to clarify the species relationships and evolutionary dynamics of the EAL and relatives.

Causes underlying phylogenetic discordances and current chanllenges for the EAL

Phylogenetic discordance is very common in eukaryotes, and it can be attributed to various factors, among which ILS and introgression or hybridization (reticulate evolution) are the prevalent causes [102–108]. In this study, significant discordances were detected not only among individual gene trees (Figs. S4 and S7) but also between the phylogenetic trees inferred from 2,186 LCGs and 163 plastid-sequences datasets (Figs. 5 and S8). The qcCFs statistics confirm that ILS is the main cause of phylogenetic conflicts, although other mechanisms (e.g., introgression or hybridization) should also be considered as possible sources of phylogenetic discordance, with α-values ranging from 0.01 to 1e-06 (Figs. 5C and S11). ILS is prevalent in recently and rapidly diverging lineages, such as those undergoing adaptive radiation, due to the short time intervals between speciation events [42, 94–96]. Previous studies have revealed that the origin of subg. *Anguinum* dates back to approximately 4.85 million years ago (Mya), with the most recent common ancestor of the EAL emerging around 3.05 Ma [20]. Additionally, the species within this lineage exhibit relatively short branch lengths in the phylogeny of the larger *Allium* system [16, 20]. All the evidence suggests that the species in this subgenus have undergone rapid radiation, and the retention of certain ancestral polymorphisms may contribute to phylogenetic conflicts [109]. The short branch lengths but fully supported nodes (BS = 100) (Fig. 6D, H), and a significant positive correlation between branch lengths and two concordance factors (gCF and sCF) (Fig. 6, $P < 0.001$), further supporting a relationship between ILS (as indicated by branch lengths) and phylogenetic conflicts [34, 42, 43]. Moreover, strong ILS signals were detected at nearly all key nodes within the EAL (Fig. 7: nodes B–L; ILS-i values ranging from 32.4 to 98.8%), further confirming that incomplete lineage sorting (ILS) is the primary cause of phylogenetic discordance.

In addition to ILS, reticulate evolution (introgression or hybridization) also plays a significant role in shaping the phylogenetic lineage structure of eukaryotes, particularly in plants [92, 105, 110]. We provide direct evidence demonstrating the contribution of hybridization or introgression to phylogenetic conflicts in the EAL (Figs. 5 and 7).

By comparing the phylogenetic trees inferred from LCGs and plastid data under quartet concordance factor analysis, we found that phylogenetic conflicts are widespread in the EAL (Figs. 5 and S8). Meanwhile, we also detected significant signals of introgression/hybridization (IH-i) at several key nodes within the EAL (Fig. 7), such as the ancestral node of the lineage (node B, IH-i = 46.7%) and the divergence nodes of multiple sublineages, particularly numerous nodes within sublineage I (node F, IH-i = 23.4%; node I, IH-i = 40%; node J, IH-i = 38%; node L, IH-i = 35.6%). The coexistence of phylogenetic conflicts and introgression/hybridization signals at specific nodes in the phylogenetic tree may indicate that hybridization or introgression has a non-negligible contribution to the discordance in phylogenetic relationships at these positions. The network topology inferred from NANUQ analysis also corroborated the existence of reticulate evolution (Fig. S12).

Under the influences of ILS and hybridization or introgression, interspecific boundaries within the EAL have become increasingly blurred, leading to more complex species relationships, a pattern consistent with our findings here (Figs. 5 and S8). The species within this lineage are primarily distributed in the HDMs and adjacent regions, an area characterized by exceptionally complex geological and climatic conditions [87–90, 93]. While substantial geographic barriers may isolate species into distinct regions, such isolation is often transient, and the speciation model accompanied by gene flow has gained broad academic recognition [92, 111]. Our field investigations have revealed coexisting sympatric-allopatric distribution patterns among the EAL species, while molecular analyses demonstrate significant interspecific ILS and hybrid introgression signals, along with phylogenetically nested species relationships. We therefore propose that these species have undergone a complex evolutionary process: on one hand, all species retain substantial ancestral genetic information with incomplete lineage sorting over short timescales; on the other hand, frequent interspecific hybridization and introgression driven by geological and climatic changes in the HDMs and adjacent areas have shaped species relationships with distinct geographic distribution patterns (as illustrated in Fig. 3). Meanwhile, local adaptation through niche differentiation and natural selection continuously drives species-specific evolution [112], particularly evident in widespread species *A. prattii* and *A. ovalifolium*. These multifaceted factors not only influence genetic variation patterns but also obscure species boundaries within this subgenus [113, 114].

Overall, our study reconstructed a comprehensive phylogeny of the EAL and, for the first time, used multiple phylogenomic datasets to elucidate the underlying causes of its phylogenetic discordances and nested relationships. Compared to previous studies [1, 5, 6, 10, 17, 77–79], We have undertaken a more extensive sampling of the EAL by including multiple populations per species and, for the first time, incorporated all recognized species within this lineage. Our study focused on the EAL, most species of which are endemic to China, historically been marked by unclear phylogenetic relationships and long-standing taxonomic controversies. The results suggest that the EAL exhibits complex species relationships potentially involving multiple speciation modes and different evolutionary stages, representing a typical species complex. However, even with the aid of high-throughput sequencing data, resolving the species relationships within the EAL and across the entire subg. *Anguinum* remains highly challenging. Future studies should conduct comprehensive field surveys across its distribution range while collecting population-level morphological, cytological, and molecular data to further elucidate species relationships and speciation mechanisms.

Conclusion

In summary, this study systematically investigated the complex phylogenetic relationships, causes of phylogenetic discordance, and evolutionary history of the East Asian lineage of *Allium* subg. *Anguinum* by integrating phylogenomic, morphological, and ecological evidence. Phylogenetic analyses strongly support that all species of the East Asian lineage form a distinct clade, which is sister to the Eurasian–North American lineage. Due to the combined effects of partially retained ancestral polymorphisms and reticulate evolutionary processes, species within the East Asian lineage exhibit pronounced phylogenetic nesting and blurred species boundaries. The rapid radiation of this lineage was likely driven by multiple factors, including climatic fluctuations triggered by the further uplift of the Qinghai–Tibetan Plateau during the late Pliocene and habitat fragmentation caused by orogenic activity in the HDMs. Following diversification, dramatic climatic oscillations during Pleistocene glacial–interglacial cycles likely promoted a repeated "isolation–contact" speciation model in glacial refugia, resulting in the phylogenetic nesting and morphological intergradation among *A. prattii*, *A. nanodes*, and *A. ovalifolium*. Multiple lines of evidence confirmed the taxonomic status of two varieties of *A. ovalifolium* (var. *cordifolium* and var. *leuconeurum*) and proposed the reclassification of *A. funckiifolium* as a new variety (*A. ovalifolium* var. *funckiifolium*). Furthermore, the widespread phylogenetic discordance within the EAL was found to be mainly driven by incomplete lineage sorting (ILS) and frequent hybridization or introgression events. This study not only enhances our understanding of the evolutionary dynamics of this lineage but also provides a typical case for exploring the evolutionary mechanisms of complex

species groups in the biodiversity hotspot of the HDMs. Future research should incorporate denser population sampling, whole-genome data, and ecological niche modeling to further elucidate the evolutionary processes underlying speciation in this lineage.

Abbreviations

EAL	East Asian lineage
ENAL	Eurasian-North American lineage
LCGs	low-copy nuclear genes
SCGs	single-copy nuclear genes
ILS	incomplete lineage sorting
HHMs	Himalayas and Hengduan Mountains
HDMs	Hengduan Mountains
QDMs	Qinling-Daba mountains

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06982-0.

Supplementary Material 1.

Acknowledgements
We are grateful to Xiao Yu-Lin from Sichuan University for providing information on the distribution of some materials.

Authors' contributions
RC: Investigation, Methodology, Data curation, Writing – original draft, Formal analysis. YW: Methodology, Data curation. BS: Methodology, Software. JT: Investigation, Methodology. JS: Investigation, Data curation. SZ: Investigation, Methodology. XH: Supervision, Methodology. DX: Supervision, Methodology, Data curation, Writing – review & editing.

Funding
This work was supported by the Sichuan Science and Technology Program (Grant Number 2025ZNSFSC1129), the National Natural Science Foundation of China (Grant Nos. 32100180), the Key project at central government level: the ability establishment of sustainable use for valuable Chinese medicine resources (2060302).

Data availability
The newly generated raw transcriptome sequencing data from this study have been deposited in the Genome Sequence Archive (GSA) at the National Genomics Data Center (NGDC), under BioProject accession number PRJCA038020. The associated BioSamples are registered under accession numbers SAMC5325376–SAMC5325417, and the raw sequencing data are available under GSA accession number CRA027140.The assembly and annotated chloroplast genomes were deposited in NCBI with accession numbers PV817864-PV817907. Voucher specimens were identifed by Deng-Feng Xie and deposited in SZ (Herbarium, Sichuan University, Sichuan Province) with deposition numbers (SZ00240503–SZ00240547).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Key Laboratory of Bio-Resources and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, Sichuan 610065, People's Republic of China
[2]State Key Laboratory for Quality Ensurance and Sustainable Use of Dao-di Herbs, National Resource Center for Chinese Materia Medica, Academy of Chinese Medical Sciences, Beijing 100700, People's Republic of China

Received: 22 April 2025 / Accepted: 27 June 2025
Published online: 22 July 2025

References

1. Herden T, Hanelt P, Friesen N. Phylogeny of *Allium* L. subgenus *Anguinum* (G. Don. Ex W.D.J. Koch) N. Friesen (Amaryllidaceae). Mol Phylogenet Evol. 2016;95:79–93.
2. Govaerts R, Kington S, Friesen N, Fritsch R, Snijman DA, Marcucci R, Silverstone-Sopkin PA, Brullo S. World checklist of Amaryllidaceae. Royal Botanic Gardens, Kew; 2021 [cited 2024 Feb 28]. Available from: http://apps.kew.org/wcsp/.
3. Chase MW, Rev Ea LJL. A phylogenetic classification of the land plants to accompany APG III. Biol J Linn Soc. 2009;161(2):122–7.
4. Chase MW, Christenhusz MJM, Fay MF, Byng JW, Judd WS, Soltis DE, Mabberley DJ, Sennikov AN, Soltis PS, Stevens PF. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. Biol J Linn Soc. 2016;181(1):1–20.
5. Choi HJ. A partial revision of *Allium* (Amaryllidaceae) in Korea and northeastern China. Bot J Linn Soc. 2011;167(2):153–211.
6. Friesen N, Fritsch RM, Blattner FR. Phylogeny and new intrageneric classification of *Allium* (Alliaceae) based on nuclear ribosomal DNA *ITS* sequences. Aliso. 2006;22(1):372–95.
7. Fritsch RM, Friesen N. Evolution, domestication and taxonomy. In: Rabinowitch HD, Currah L, editors. Allium Crop Science. Wallingford, UK: CABI; 2002. p. 5–30.
8. Nguyen NH, Driscoll HE, Specht CD. A molecular phylogeny of the wild onions (*Allium*; Alliaceae) with a focus on the Western North American center of diversity. Mol Phylogenet Evol. 2008;47(3):1157–72.
9. Xie DF, Tan JB, Yu Y, Gui LJ, Su DM, Zhou SD, He XJ. Insights into phylogeny, age and evolution of *Allium* (Amaryllidaceae) based on the whole plastome sequences. Ann Botany. 2020;125(7):1039–55.
10. Li QQ, Zhou SD, He XJ, Yu Y, Zhang YC, Wei XQ. Phylogeny and biogeography of *Allium* (Amaryllidaceae: Allieae) based on nuclear ribosomal internal transcribed spacer and Chloroplast *rps16* sequences, focusing on the inclusion of species endemic to China. Ann Botany. 2010;106(5):709–33.
11. Fritsch RM. Checklist of *Allium* L. (Amaryllidaceae) species in the old world and their taxonomic and nomenclatural status. Stapfia. 2012;97:125–41.
12. Fritsch RM, Blattner F, Gurushidze M. New classification of *Allium* L. Subg. *Melanocrommyum* (WEBB & BERTHEL.) ROUY (Alliaceae) based on molecular and morphological characters. Phyton; Ann Rei Botanicae. 2010;49:145–320.
13. Fritsch RM, Abbasi M. New taxa and other contributions to the taxonomy of *Allium* L. (Alliaceae) in Iran. Rostaniha. 2008;9:1–73.
14. Li QQ, Zhou SD, Huang DQ, He XJ, Wei XQ. Molecular phylogeny, divergence time estimates and historical biogeography within one of the world's largest monocot genera. AoB Plants. 2016;8(4 Suppl):plw041.
15. He XJ. Integrating high-volume molecular and morphological data into the evolutionary studies of *Allium*. Plant Divers. 2023;46(1):1–2.
16. Zhang ZZ, Liu G, Li M. Incomplete lineage sorting and gene flow within *Allium* (Amayllidaceae). Mol Phylogenet Evol. 2024;195:108054.
17. Han TS, Zheng QJ, Onstein R, Rojas Andrés B, Hauenschild F, Muellner-Riehl A, Xing YW. Polyploidy promotes species diversification of *Allium* through ecological shifts. New Phytol. 2019;225(1):571–83.
18. Xu JM, Kamelin RV. *Allium* L. In: Wu ZY, Raven PH, Hong DY, editors. Flora of China, Vol. 24. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press; 2000. p. 165–202.
19. Choi HJ, Yang S, Yang JC, Friesen N. *Allium ulleungense* (Amaryllidaceae), a new species endemic to Ulleungdo island, Korea. Korean J Plant Taxonomy. 2019;49(4):294–9.
20. Yang JY, Kim SH, Gil HY, Choi HJ, Kim SC. New insights into the phylogenetic relationships among wild onions (*Allium*, Amaryllidaceae), with special emphasis on the subgenera *Anguinum* and *Rhizirideum*, as revealed by plastomes. Front Plant Sci. 2023;14:1124277.
21. Prokhanov JI. To the knowledge of crop onions and Garlics in China and Japan. Bull Appl Bot Genet Plant Breed. 1930;24:123–88.
22. Lu Y. Studies on Karyotypes and Molecular systematics of the genus *Allium* L. subgenus *Anguinum*. Master's thesis. Chengdu: Sichuan University; 2015.

23. Jin FY. Phylogeny and ecological adaptation of the genus *Allium* L. subgen. *Anguinum* in China. Master's thesis. Chengdu: Sichuan University; 2019.
24. Stearn WT. *Allium* in the Balkans and Asia minor: notes on their ecology and distribution. Annals Goulandris Nat History Museum. 1980;4:83–98.
25. Vvedenskii AI. On the ecology and geography of *Allium* species in central Asia. Botanicheskii Zhurnal. 1986;71:624–31.
26. Zoku T. Ecological studies of wild *Allium* species in Japan. J Japanese Bot. 1984;59:65–72.
27. Zhang CY. Systematic taxonomic studies on *Allium* in Southwest China. Master's thesis. Chengdu: Sichuan University; 2009.
28. Zhang SS, Tang ZJ, Chen K, Ma XJ, He XJ, Xie DF. Investigating the variation in leaf traits within the *Allium Prattii* C.H. Wright population and its environmental adaptations. Plants. 2025;14(4):541.
29. Friesen NV. Karyotype investigations in species of *Allium* L. from Siberia and the Far East. Botanicheskii Zhurnal. 1986;71:1579–86.
30. Jing HC, Stiekema WJ, van Lammeren AAM. Chromosome behavior during mitosis and interphase in *Allium cepa* L. as investigated by in situ hybridization with a satellite DNA sequence. Genome. 1999;42:1216–21.
31. Xue H, Li J, Zhang X, Lin Z. Karyotype analysis of five wild *Allium* species from China. Acta Phytotaxonomica Sinica. 2000;38:333–41.
32. Tang Y, Li X, Liu QW. Cytological studies on ten species of *Allium* (Liliaceae) from China. J Syst Evol. 2005;43:77–86.
33. Probatova NS, Sokolovskaya AP, Rudyka EG. Chromosome numbers in some species of *Allium* (Alliaceae) from the Russian Far East. Botanicheskii Zhurnal. 2006;91:935–41.
34. Zhang ZZ, Liu G, Li MJ. Phylotranscriptomic discordance is best explained by incomplete lineage sorting within *Allium* subgenus *Cyathophora* and thus hemiplasy accounts for interspecific trait transition. Plant Divers. 2024;46(1):28–38.
35. Lin HY, Hao YJ, Li JH, Fu CX, Soltis P, Zhao YP. Phylogenomic conflict resulting from ancient introgression following species diversification in *Stewartia* s.l. (Theaceae). Mol Phylogenet Evol. 2019;135(8):1–11.
36. Smith S, Walker-Hale N, Walker J, Brown J. Phylogenetic conflicts, combinability, and deep phylogenomics in plants. Syst Biol. 2019;69(3):579–92.
37. Meleshko O, Martin M, Schröck C, Lamkowski P, Schmutz J, Healey A, Piatkowski B, Shaw A, Weston D, Flatberg K, et al. Extensive genome-wide phylogenetic discordance is due to incomplete lineage sorting and not ongoing introgression in a rapidly radiated bryophyte genus. Mol Biol Evol. 2021;38(9):2750–66.
38. Wang G, Zhang XT, Herre E, McKey D, Machado C, Yu WB, Cannon C, Arnold M, Pereira R, Ming R, et al. Genomic evidence of prevalent hybridization throughout the evolutionary history of the fig-wasp pollination mutualism. Nat Commun. 2021;12(1):718.
39. Cheng L, Han QW, Chen F, Li MG, Balbuena T, Zhao YY. Phylogenomics as an effective approach to untangle cross-species hybridization event: A case study in the family Nymphaeaceae. Front Genet. 2022;13:1031705.
40. Zhang L, Zhu XX, Zhao YY, Guo J, Zhang TK, Huang WC, Jie H, Hu Y, Huang CH, Ma H. Phylotranscriptomics resolves the phylogeny of Pooideae and uncovers factors for their adaptive evolution. Mol Biol Evol. 2022;39(2):msac026.
41. Liu J, Lindstrom A, Gong YQ, Dong SS, Liu YS, Zhang SZ, Xun G. Eco-evolutionary evidence for the global diversity pattern of *Cycas* (Cycadaceae). J Integr Plant Biol. 2024;66(6):1170–91.
42. Xie DF, Li J, Sun JH, Cheng RY, Wang Y, Song BN, He XJ, Zhou SD. Peering through the hedge: multiple datasets yield insights into the phylogenetic relationships and incongruences in the tribe lilieae (Liliaceae). Mol Phylogenet Evol. 2024;200(17):108182.
43. Xu L, Song Z, Li T, Jin Z, Zhang B, Du S, Liao S, Zhong X, Chen Y. New insights into the phylogeny and infrageneric taxonomy of *Saussurea* based on hybrid capture phylogenomics (Hyb-Seq). Plant Divers. 2025;47(1):21–33.
44. Li M, Zheng Z, Liu J, Yang Y, Ren G, Ru D, Zhang S, Du X, Ma T, Milne R, et al. Evolutionary origin of a tetraploid *Allium* species on the Qinghai-Tibet plateau. Mol Ecol. 2021;30(22):5780–95.
45. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinf (Oxford England). 2018;34(17):i884–90.
46. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinf (Oxford England). 2014;30(15):2114–20.
47. Doyle J. A rapid DNA isolation procedure for small amounts of fresh leaf tissue. Phytochem Bull. 1987;19:11–5.
48. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. GetOrganelle: a fast and versatile toolkit for accurate de Novo assembly of organelle genomes. Genome Biol. 2020;21(1):241.
49. 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.
50. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinf (Oxford England). 2012;28(12):1647–9.
51. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004;32(Web Server issue):W273–9.
52. 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(Web Server issue):W575–81.
53. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinf (Oxford England). 2009;25(15):1972–3.
54. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 2020;20(1):348–55.
55. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–74.
56. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–9.
57. Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics. 2018;19(Suppl 6):153.
58. Sayyari E, Mirarab S. Fast coalescent-based computation of local branch support from quartet frequencies. Mol Biol Evol. 2016;33(7):1654–68.
59. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. De Novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. Nat Protoc. 2013;8(8):1494–512.
60. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinf (Oxford England). 2012;28(23):3150–2.
61. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015;16(1):157.
62. Xie P, Guo Y, Teng Y, Zhou W, Yu Y. GeneMiner: A tool for extracting phylogenetic markers from next-generation sequencing data. Mol Ecol Resour. 2024;24(3):e13924.
63. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
64. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 2006;34(suppl2):W609–12.
65. Morales-Briones DF, Kadereit G, Tefarikis DT, Moore MJ, Smith SA, Brockington SF, Timoneda A, Yim WC, Cushman JC, Yang Y. Disentangling sources of gene tree discordance in phylogenomic data sets: testing ancient hybridizations in Amaranthaceae s.l. Syst Biol. 2021;70(2):219–35.
66. Bruen TC, Philippe H, Bryant D. A simple and robust statistical test for detecting the presence of recombination. Genetics. 2006;172(4):2665–81.
67. Smith SA, Moore MJ, Brown JW, Yang Y. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. BMC Evol Biol. 2015;15(1):150.
68. Salichos L, Stamatakis A, Rokas A. Novel information theory-based measures for quantifying incongruence among phylogenetic trees. Mol Biol Evol. 2014;31(5):1261–71.
69. Minh BQ, Hahn MW, Lanfear R. New methods to calculate concordance factors for phylogenomic datasets. Mol Biol Evol. 2020;37(9):2727–33.
70. Rhodes JA, Baños H, Mitchell JD, Allman ES. MSCquartets 1.0: quartet methods for species trees and networks under the multispecies coalescent model in R. Bioinf (Oxford England). 2021;37(12):1766–8.
71. Allman ES, Mitchell JD, Rhodes JA. Gene tree discord, simplex plots, and statistical tests under the coalescent. Syst Biol. 2022;71(4):929–42.
72. Shang HY, Jia KH, Li NW, Zhou MJ, Yang H, Tian XL, Ma YP, Zhang RG. Phytop: a tool for visualizing and recognizing signals of incomplete lineage sorting

and hybridization using species trees output from ASTRAL. Hortic Res. 2025;12(3):uhae330.
73. Allman ES, Baños H, Rhodes JA. NANUQ: a method for inferring species networks from gene trees under the coalescent model. Algorithms Mol Biology: AMB. 2019;14:24.
74. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006;23(2):254–67.
75. Xu JM. Flora reipublicae popularis sinicae. Volume 14. Beijing: Science; 1980.
76. 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.
77. Rambaut A. FigTree, version 1.4.2 [software]. Edinburgh: Institute of Evolutionary Biology, University of Edinburgh; 2014 [cited 2024 Feb 28]. Available from: http://tree.bio.ed.ac.uk/software/figtree/.
78. Feng S, Bai M, Rivas-González I, Li C, Liu S, Tong Y, Yang H, Chen G, Xie D, Sears KE, et al. Incomplete lineage sorting and phenotypic evolution in marsupials. Cell. 2022;185(10):1646–60.
79. Guo C, Luo Y, Gao LM, Yi TS, Li HT, Yang JB, Li DZ. Phylogenomics and the flowering plant tree of life. J Integr Plant Biol. 2023;65(2):299–323.
80. Yang L, Harris AJ, Wen F, Li Z, Feng C, Kong H, Kang M. Phylogenomic analyses reveal an allopolyploid origin of core Didymocarpinae (Gesneriaceae) followed by rapid radiation. Syst Biol. 2023;72(5):1064–83.
81. Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, Rissler L, Victoriano PF, Yoder AD. Phylogeography's past, present, and future: 10 years after avise, 2000. Mol Phylogenet Evol. 2010;54(1):291–301.
82. Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS. Comparative phylogeography of unglaciated Eastern North America. Mol Ecol. 2006;15(14):4261–93.
83. Du FK, Wang T, Wang Y, Ueno S, de Lafontaine G. Contrasted patterns of local adaptation to climate change across the range of an evergreen oak, *Quercus aquifolioides*. Evol Appl. 2020;13(9):2377–91.
84. Shen Y, Xia H, Tu Z, Zong Y, Yang L, Li H. Genetic divergence and local adaptation of Liriodendron driven by heterogeneous environments. Mol Ecol. 2022;31(3):916–33.
85. Feng S, Xi E, Wan W, Ru D. Genomic signals of local adaptation in *Picea crassifolia*. BMC Plant Biol. 2023;23(1):534.
86. Feng B, Wang XF, Chen SS, Zhang Y, Su XJ, Song SY. Transcriptome analysis and genetic diversity of *Allium victorialis* germplasms from the Changbai mountains. Mitochondrial DNA Part B. 2021;6(10):2915–23.
87. Liu JQ, Duan YW, Hao G, Ge XJ, Sun H. Evolutionary history and underlying adaptation of alpine plants on the Qinghai-Tibet plateau. J Syst Evol. 2014;52(3):241–9.
88. Marchese C. Biodiversity hotspots: A shortcut for a more complicated concept. Global Ecol Conserv. 2015;3:297–309.
89. Mosbrugger V, Favre A, Muellner-Riehl AN. Cenozoic evolution of geo-biodiversity in the Tibeto-Himalayan region. In: Hoorn C, Perrigio A, Antonelli A, editors. Mt Clim Biodiv. Chichester: Wiley-Blackwell; 2018. p. 429–48.
90. Rahbek C, Borregaard M, Antonelli A, Colwell R, Holt B, Nogués-Bravo D, Rasmussen C, Richardson K, Rosing M, Whittaker R, et al. Building mountain biodiversity: geological and evolutionary processes. Sci (New York NY). 2019;365(6458):1114–9.
91. Zheng W, Yan LJ, Burgess K, Yahuang L, Zou JY, Qin HT, Wang JH, Gao LM. Natural hybridization among three *Rhododendron* species (Ericaceae) revealed by morphological and genomic evidence. BMC Plant Biol. 2021;21(1):529.
92. Wu SD, Wang Y, Wang ZF, Shrestha N, Liu JQ. Species divergence with gene flow and hybrid speciation on the Qinghai–Tibet plateau. New Phytol. 2022;234(2):392–404.
93. Ding WN, Ree R, Spicer R, Xing YW. Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora. Sci (New York NY). 2020;369(6503):578–81.
94. Meyer B, Matschiner M, Salzburger W. Disentangling incomplete lineage sorting and introgression to refine species-tree estimates for lake Tanganyika cichlid fishes. Syst Biol. 2017;66(4):531–50.
95. Edelman N, Frandsen P, Miyagi M, Clavijo B, Davey J, Dikow R, Accinelli G, Belleghem SM, Patterson N, Neafsey D, et al. Genomic architecture and introgression shape a butterfly radiation. Sci (New York NY). 2019;366(6465):594–9.
96. Cai L, Xi Z, Lemmon EM, Lemmon AR, Mast A, Buddenhagen CE, Liu L, Davis CC. The perfect storm: gene tree Estimation error, incomplete lineage sorting, and ancient gene flow explain the most recalcitrant ancient angiosperm clade, Malpighiales. Syst Biol. 2021;70(3):491–507.
97. Huang DQ, Ma XG, Sun H. Phylogenomic analyses and chromosome ploidy identification reveal multiple cryptic species in *Allium Sikkimense* complex (Amaryllidaceae). Front Plant Sci. 2023;14:1268546.
98. Wang KL, Zhou XH, Liu DT, Li YZ, Yao Z, He WM, Liu YB. The uplift of the Hengduan mountains contributed to the speciation of three *Rhododendron* species. Glob Ecol Conserv. 2022;35:e02085.
99. Cheng XJ, Fritsch PW, Lin YJ, Li GH, Chen YQ, Zhang MY, Lu L. The role of pleistocene dispersal in shaping species richness of Sky Island Wintergreens from the Himalaya-Hengduan mountains. Mol Phylogenet Evol. 2024;197(3):108082.
100. Liu R, Wang H, Yang JB, Corlett RT, Randle CP, Li DZ, Yu WB. Cryptic species diversification of the *Pedicularis Siphonantha* complex (Orobanchaceae) in the mountains of Southwest China since the pliocene. Front Plant Sci. 2022;13:811206.
101. Ma XG, Ren YB, Sun H. Introgression and incomplete lineage sorting blurred phylogenetic relationships across the genomes of sclerophyllous Oaks from Southwest China. Cladistics. 2024;40(4):357–73.
102. Degnan JH, Rosenberg NA. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol Evol. 2009;24(6):332–40.
103. Arekar K, Parigi A, Karanth KP. Understanding the convoluted evolutionary history of the capped-golden langur lineage (Cercopithecidae: Colobinae)†. J Genet. 2021;100(2):79.
104. Olave M, Avila LJ, Sites JJ, Morando M. Detecting hybridization by likelihood calculation of gene tree extra lineages given explicit models. Methods Ecol Evol. 2017;9(1):121–33.
105. Taylor SA-O, Larson EL. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. Nat Ecol Evol. 2019;3(2):170–7.
106. Rose JP, Toledo CAP, Lemmon EM, Lemmon AR, Sytsma KJ. Out of sight, out of mind: widespread nuclear and plastid-nuclear discordance in the flowering plant genus *Polemonium* (Polemoniaceae) suggests widespread historical gene flowdespite limited nuclear signal. Syst Biol. 2021;70(1):162–80.
107. Sarver BAJ, Herrera ND, Sneddon D, Hunter SS, Settles ML, Kronenberg Z, Demboski JR, Good JM, Sullivan J. Diversification, introgression, and rampant cytonuclear discordance in Rocky mountains chipmunks (Sciuridae: *Tamias*). Syst Biol. 2021;70(5):908–21.
108. Talavera A, Nie ZL, Ma ZY, Johnson G, Ickert-Bond SM, Zimmer EA, Wen J. Phylogenomic analyses using a new 1013-gene Vitaceae bait-set support major groups of North American Vitis. Mol Phylogenet Evol. 2023;186(5):107866.
109. Mao K, Ruhsam M, Ma Y, Graham SW, Liu J, Thomas P, Milne RI, Hollingsworth PM. A transcriptome-based resolution for a key taxonomic controversy in Cupressaceae. Ann Botany. 2019;123(1):153–67.
110. Liu J, Nie ZL, Ren C, Su C, Wen J. Phylogenomics of *Aralia* sect. *Aralia* (Araliaceae): signals of hybridization and insights into its species delimitations and intercontinental biogeography. Mol Phylogenet Evol. 2023;181(S5):107727.
111. Gorospe JM, Záveská E, Chala D, Gizaw A, Tusiime FM, Gustafsson ALS, Piálek L, Kolář F, Brochmann C, Schmickl R. Ecological speciation with gene flow followed initial large-scale geographic speciation in the enigmatic Afroalpine giant senecios (*Dendrosenecio*). New Phytol. 2025;246(5):2307–23.
112. Hoffmann AA, Sgrò CM. Climate change and evolutionary adaptation. Nature. 2011;470(7335):479–85.
113. Cardoni S, Piredda R, Denk T, Grimm GW, Papageorgiou AC, Schulze ED, Scoppola A, Salehi Shanjani P, Suyama Y, Tomaru N, et al. 5S-IGS rDNA in wind-pollinated trees (*Fagus* L.) encapsulates 55 million years of reticulate evolution and hybrid origins of modern species. Plant Journal: Cell Mol Biology. 2022;109(4):909–26.
114. McLay TGB, Fowler RM, Fahey PS, Murphy DJ, Udovicic F, Cantrill DJ, Bayly MJ. 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(3):107869.

Publisher's note