

Research paper

Multi-disciplinary evidence illuminates the speciation history of a monophyletic yet dimorphic lily group

Yu Feng^a, Chaochao Yan^a, Wen-Qin Tu^b, Yu-Mei Yuan^{a, c}, Jing-Bo Wang^{a, c},
Xiao-Juan Chen^a, Chang-Qiu Liu^d, Yundong Gao^{a, *}

^a Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu 610213, China

^b Key Laboratory of Biodiversity Science and Ecological Engineering of the Ministry of Education, College of Life Sciences, Beijing Normal University, Beijing 100875, China

^c College of Life Sciences, University of Chinese Academy of Sciences, Beijing 101408, China

^d Guangxi Key Laboratory of Plant Conservation and Restoration Ecology in Karst Terrain, Guangxi Institute of Botany, Guangxi Zhuang Autonomous Region and Chinese Academy of Sciences, Guilin 541006, China



ARTICLE INFO

Article history:

Received 8 October 2024

Received in revised form

19 December 2024

Accepted 23 December 2024

Available online 28 December 2024

Keywords:

Speciation

Species boundary

Interspecific gene flow

Integrated species notion

Lilium

Isolation

ABSTRACT

Species boundaries are dynamic and constantly challenged by gene flow. Understanding the strategies different lineages adopt to maintain ecological and genetic distinctiveness requires employing an integrative species concept that incorporates data from a variety of sources. In this study, we incorporated genetic, ecological, and environmental evidence to assess the extent of speciation or evolutionary divergence within a monophyletic yet dimorphic group (i.e., clade *Leucolirion* consisting of six species) within the genus *Lilium*. This clade consists of two lineages that exhibit unexpectedly distinct perianth appearances: whitish trumpet (funnel form, encompassing four species) and orange recurved (reflex form, including two species), respectively, which are separated by completely different pollination syndromes. Transcriptome-based nuclear and plastome datasets indicate that these two lineages are isolated, with only weak ancient gene flow between them. Within each lineage, several taxa with incomplete isolation have diverged, as indicated by weak genetic structure, strong gene flow, and conflicts between nuclear and chloroplast phylogenies, especially in the trumpet lineage. Although these taxa are not entirely independent, our evidence indicates that they are diverging, with recent gene flow disappearing and multiple isolation strategies emerging, such as differences in flowering time and niche specialization. Taken together, our findings suggest that species divergence and maintenance in *Lilium* are driven by a combination of adaptive and non-adaptive processes, highlighting the complex interplay of historical climate changes, ecological adaptation, and gene flow in shaping biodiversity within this genus.

Copyright © 2024 Kunming Institute of Botany, Chinese Academy of Sciences. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Species are the fundamental units of biodiversity, playing a crucial role in conservation biology and ecological research (Wheeler and Meier, 2000; Coyne and Orr, 2004; Hong, 2016). However, the concept of species is complex and often controversial. Defining a species has been, and will continue to be, a major challenge for biological studies, affecting the formulation of

conservation strategies (Akçakaya et al., 2004). Despite numerous proposed species concepts, no single definition satisfies the entire spectrum of life forms, from unicellular organisms to complex multicellular beings like humans (De Queiroz, 2007; Wilkins, 2009). Therefore, different organisms may require distinct species concepts.

From an evolutionary perspective, species are dynamic entities, continuously evolving through time (Coyne and Orr, 2004; De Queiroz, 2007; Liu, 2016). While complete reproductive isolation has long been considered a key factor in speciation and maintaining species boundaries, particularly in the Biological Species Concept (Mayr, 1963), recent studies have revealed that gene flow is more pervasive than previously thought (Bolnick and Fitzpatrick, 2007; Nosil, 2008; Blanckaert et al., 2020; Wang et al., 2020). This is

* Corresponding author. No. 23 Qunxian South Street, Shuangliu District, Chengdu 610213, China.

E-mail address: gaoyd@cib.ac.cn (Y. Gao).

Peer review under the responsibility of Editorial Office of Plant Diversity.

especially true in plants, where hybridization and introgression occur not only between closely related taxa but also between more distantly related congeners or even across genera (Hu et al., 2021; Wu et al., 2022; Zhou et al., 2022; Mason-Gamer and White, 2024). Surprisingly, despite this gene flow, many species maintain distinct morphological and ecological identities. This paradox raises intriguing questions about the mechanisms of species persistence and boundary maintenance in the face of ongoing genetic exchange.

Contemporary biologists have proposed more integrative approaches to species delimitation. These incorporate multiple lines of evidence, including breeding systems, geography, morphology, DNA sequences, and ecology (Liu, 2016; Donoghue, 2022). Moreover, there is a growing recognition that speciation is a continuous process, with extant species representing various stages of divergence and independence (Liu, 2016; Gao et al., 2019; Zhang et al., 2023). This perspective acknowledges that species boundaries are dynamic and constantly challenged by gene flow, with different lineages adopting unique strategies to maintain their ecological and genetic distinctiveness.

The genus *Lilium*, spanning the temperate Northern Hemisphere, is renowned for its showy flowers and horticultural importance. Its success in cultivation stems from extensive interspecific hybridization, indicating weak reproductive barriers among species. *Lilium* species are typically self-incompatible and pollinator-dependent, making them prone to hybridization in both natural and cultivated settings (Marasek-Ciolakowska et al., 2018). Intriguingly, our observations indicate that while hybridization is common in *Lilium* species, species boundaries remain relatively stable in nature. In this genus, morphological diversity of flowers has led to various sectional classifications (Comber, 1949; Nishikawa et al., 1999; Hayashi and Kawano, 2000). For example, four perigone types correlate with specific pollinators in North America (Givnish et al., 2020), while in East Asia, the diversity is even greater. However, traditional classifications, such as assigning trumpet and recurved forms to sections *Leucolirion* and *Sinomartagon* respectively (Comber, 1949), have been challenged by recent phylogenetic studies (Kim et al., 2019; Duan et al., 2022; Li et al., 2022). Our previous research contradicts studies that suggested floral morphology reflects genetic relationships, instead indicating that floral morphology results from parallel evolution in response to similar habitats (Gao et al., 2015). This phenomenon is widely observed throughout the entire genus, making sub-generic classification much more difficult (Liu et al., 2019; Yuan and Gao, 2024; Zhou et al., 2024).

The monophyletic clade *Leucolirion* proposed by Gao et al. (2013) comprises six taxa exhibiting two distinct floral forms as mentioned above. This clade is endemic to central and southwestern China, with *Lilium sulphureum* Baker ex Hook. f. extending into neighboring Myanmar. Four members represented by *L. regale* E.H. Wilson—discovered and introduced by the famous plant hunter E.H. Wilson in the Min River valley—are characterized by the white giant trumpet flower with a strong fragrance. In contrast, the other two taxa (*L. henryi* Baker and *L. rosthornii* Diels), from central China, exhibit strongly recurved, inodorous orange to red petals adorned with black speckles. Despite their strikingly different appearances, these taxa share unique purplish bulbs characteristic of Chinese lilies. Our previous research (Liu et al., 2022) revealed that these divergent floral traits correspond to distinct pollinators: the trumpet-flowered species are pollinated by hawk moths at night, whereas butterflies visit the recurved flowers during the day.

After being categorized by floral traits into two groups, these groups were further separated into six taxa, which were accepted by *Flora of China* (Liang and Tamura, 2000). Although there were

pervasive diagnostic criteria to identify each of taxon based on morphology, these taxa still show some overlaps in traits (Liang and Tamura, 2000). Furthermore, their distributions are largely sympatric or partially overlapping, suggesting that gene flow among these taxa is unavoidable. As mentioned earlier, such genetic exchange would impede species formation and blur species boundaries (Mallet, 2008). However, this has not been observed in the field. Most of these taxa have maintained species boundaries, at least at the morphological level. Additional evidence (e.g., genetic and ecological data) may help clarify whether and how these plants maintain species boundaries.

In the present study, our overall aim is to better understand speciation and the maintenance of species boundaries in plants. For this purpose, we used evidence from genetic analyses, flowering time, and ecological niche modeling to delineate species boundaries and determine the divergence history within *Leucolirion*, a monophyletic yet dimorphic lily clade. Specifically, we clarified the species status of each taxon described in *Flora of China* (Liang and Tamura, 2000) and determined a complete set of genetic relationships and differentiation patterns based on population sampling of next-generation sequencing (NGS) data. We then examined how species boundaries are maintained between closely related sympatric species.

2. Materials and methods

2.1. Sampling and sequencing strategy

The extremely large genomes of *Lilium* species make *de novo* assembly of a reference genome and population resequencing a challenge. To address this, we generated a full-length transcriptome as a reference for population genomic analysis. Leaf and flower tissues were sampled from a batch of *L. regale* that had been transplanted to our garden. To maximize transcriptome coverage over the lifespan, tissue from three stages of blooming were collected. Samples were sent to Beijing Novogene Bioinformatics Technology Co., Ltd., where RNA extraction, library preparation, and sequencing using the PacBio Sequel platform were performed. Library preparation, sequencing, data processing, and gene function annotation followed the same procedures as described in Li et al. (2020).

For population sampling of the six *Lilium* species, fresh leaves and flowers were collected from natural populations and transplanted materials in our lab collected during 2018–2020. Two individuals, *L. duchartrei* Franch. (voucher: GYD1424) and *Cardiocrinum giganteum* (Wall.) Makino (voucher: GYD1400), were included as outgroup taxa. Materials for transcriptome sequencing were collected at the blooming stage, and both leaves and flowers were sampled and stored in liquid nitrogen in separated tubes. Samples were sent to Novogene for RNA extraction, library preparation and sequencing. Plastomes of each individual were obtained by genome skimming on the Illumina platform. A total of 142 individuals from 21 populations were sampled (Table S1 and Fig. 1).

2.2. Collection and analysis of flowering time data

Flowering time data for the six *Lilium* species were retrieved from the Chinese Virtual Herbarium (CVH, <https://www.cvh.ac.cn/>) and Plant Photo Bank of China (PPBC, <https://ppbc.iplant.cn/>). For each accession, the collection date the specimen or the shooting date of the photograph was imported into IBM SPSS Statistics for Windows, v.27.0 (IBM Corp., 2020). Data were then grouped by different species and by species sets that are close in morphology and flowering time. Normality tests were conducted on both the overall and the grouped data using the Kolmogorov–Smirnov and

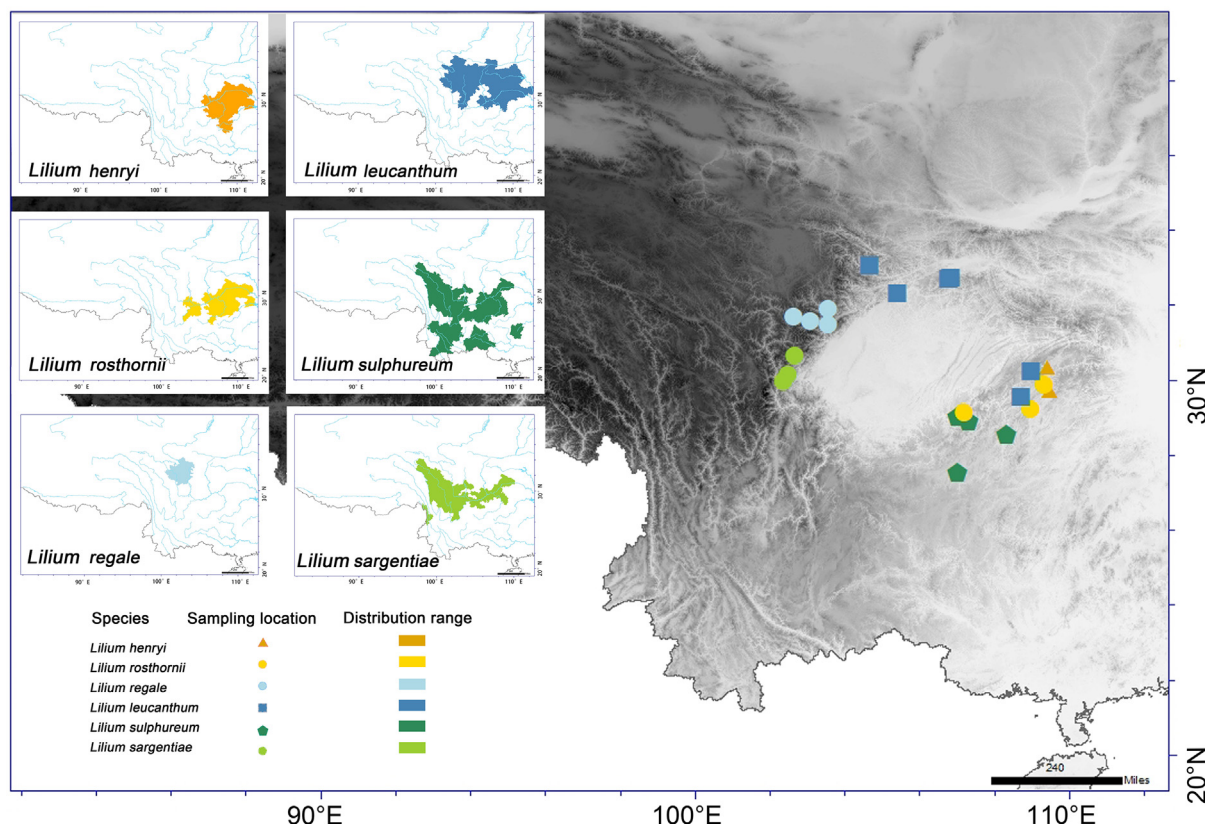


Fig. 1. Distribution ranges and population sampling locations for the six *Lilium* species.

Shapiro–Wilk tests. Results indicated that all tested datasets were non-normally distributed at a significance level of ($P < 0.05$) (Table S2). Subsequently, the variability of flowering time among subgroups in the monthly data was further analyzed using the Kruskal–Wallis one-way ANOVA (k samples) method in SPSS. Asymptotic significances (for two-sided tests) were utilized, with a significance level of 0.05. The significance values were adjusted using Bonferroni correction to account for multiple tests.

2.3. Ecological niche differentiation

To compare the niches of the six *Lilium* species, we quantified the niche overlap and width from two different perspectives: environmental space (E-space) and geographic space (G-space). Niche comparisons for G-space are based on the geographic range of species, which in many studies is often inferred from ecological niche modeling (ENM) (Lu et al., 2023), while niche comparisons for E-space typically quantify the niche of species using environmental information obtained directly from occurrence sites (Brown and Carnaval, 2019). Environmental variables used in this study include topographical factors and 19 contemporary bioclimatic factors with a resolution of 30 s (<http://www.worldclim.org>). After removing highly correlated environmental variables ($|r| > 0.75$), eight environmental factors were retained for further analyses: Isothermality (Bio3), Max Temperature of Warmest Month (Bio5), Min Temperature of Coldest Month (Bio6), Precipitation of Wettest Month (Bio13), Precipitation of Driest Month (Bio14), Precipitation Seasonality (Bio16), Slope, and Aspect Index.

To estimate the niche overlap between six *Lilium* species pairs in E-space, we employed Schoener's D metric (Schoener, 1968). This metric quantifies the degree of niche similarity between different species, with higher D values indicating greater niche overlap. For

each species, we generated a minimum convex polygon (MCP) and expanded the MCP's buffer zone based on an average annual diffusion distance of 10 km to define the geographical background. We then randomly chose 1000 points for each species in the background areas and extracted environmental data based on presence records and background points. Subsequently, the values of Schoener's D and niche similarity for the six *Lilium* species pairs were compared using the PCA-env method provided by the 'eco-spat' v.4.10 R package (Di Cola et al., 2017). In the comparison of ecological niche in G-space, we first employed the Maxent v.3.4.4 (Phillips and Dudík, 2008) to predict potential distribution for each species, which was then transformed to binary suitability maps by maximizing the sum of sensitivity and specificity (maxSSS) using 'dismo' R package v.1.3.9 (Hijmans et al., 2022). Finally, we used ENMTools v.1.4.4 (Warren et al., 2010) to assess the Schoener's D value based on the binary suitability maps of *Lilium* species.

2.4. Species-level phylogeny reconstruction and divergence time estimation

To reconstruct a species-level phylogeny, we merged all individuals from a species and assembled the transcriptomes for each species, treating different samplings within the same species as biological replicates. The Illumina NGS data was first filtered using fastp v.0.21.0 (Chen, 2023) with default parameters, and then assembled using Trinity v.2.9.1 (Haas et al., 2013) with default parameters. Assemblies were filtered using Trinity's built-in script (i.e., 'align_and_estimate_abundance.pl' and 'filter_low_expr_transcripts.pl') for initial quality control. Redundant sequences were removed using CD-HIT v.4.8.1 (Li and Godzik, 2006) with the following parameters: $-c$ 0.95, $-n$ 10, $-M$ 30000, and $-T$ 10. The integrity of the assemblies was evaluated using Benchmarking Universal Single-Copy Orthologs v.5.7

(BUSCO) (Manni et al., 2021), with the ortholog dataset set to Liliopsida ($n = 3236$). TransDecoder (<https://github.com/TransDecoder>) was used to translate transcriptome assemblies into protein sequences. OrthoFinder v.2.5.5 (Emms and Kelly, 2019) was used to cluster orthologous genes, with single-copy genes filtered and used for subsequent species tree construction. All coding sequences (CDS) were concatenated by species using custom python scripts. Four-fold sites were obtained with perl script 'obtain_4d.pl' (https://github.com/YanCCscu/comp_genomic).

A Maximum Likelihood (ML) tree was built using IQ-TREE v.2.0 (Minh et al., 2020) with ultrafast bootstrap set to 1000, and SH-like approximate likelihood ratio test set to 1000 replicates. The divergence time of the six species was estimated using BEAST v.2.17 (Bouckaert et al., 2019). BEAST analyses were performed under a strict clock approach with a Yule tree prior, random starting trees, and appropriate nucleotide substitution model (GTR+G+I). Calibration of the crown age of the six species was conducted using a normal distribution with a mean of 2.00 (Gao et al., 2015), standard deviation (s.d.) of 0.5, and 95% quantile ranging from 1.18 to 2.83 million years ago (Mya). Similarly, the crown age of the trumpet species was calibrated using a normal distribution with a mean of 1.90, s.d. of 0.4, and 95% quantile ranging from 1.24 to 2.56 Mya. Markov Chain Monte Carlo (MCMC) chains were run for one million generations, with a sampling frequency of 1000 per generation.

2.5. Plastome phylogeny

To obtain a chloroplast phylogeny, we conducted genome skimming on all individuals. DNA extraction, library preparation, and Illumina sequencing were detailed in our previous study (Yuan and Gao, 2024). Approximately 13 Gb of raw data were filtered and evaluated using fastp with default parameters. Complete plastomes were assembled by GetOrganelle v.1.7.6.1 (Jin et al., 2020). Orientation selection was refined through multiple sequence alignment with MAFFT v.7 (Katoh and Standley, 2013) and was manually verified in Geneious Prime v.2023.1.2 (Biomatters Ltd., Auckland, New Zealand). Haplotypes from each population were selected as representatives for tree construction using DnaSP v.6.12.03 (Rozas et al., 2017).

Sequence alignments were generated using MAFFT v.7. Filtering of ambiguously aligned sites was performed using Gblocks v.0.9b (Talavera and Castresana, 2007) with default parameters. Nucleotide substitution models were estimated using ModelTest-NG v.0.1.7 (Darrriba et al., 2020). Phylogenetic trees were inferred using Maximum Likelihood (ML) and Bayesian Inference (BI). ML trees were constructed using RAxML-NG v.1.2.0 (Kozlov et al., 2019) with GTR+I+G4 model and 1000 bootstrap replicates. BI analyses were performed using MrBayes v.3.2 (Ronquist et al., 2012) with the nucleotide substitution model GTR+G+I (Iset nst = 6; rates = invgamma). For each analysis, the posterior probability was estimated with two independent MCMC chains (10 million generations) with the preliminary 25% of sampled data discarded as burn-in.

2.6. Transcriptome genotyping, population genetic structure and nuclear phylogeny

Clean transcriptome sequencing data for each individual were mapped to the *de novo* assembled full-length transcriptome of *Lilium regale* utilizing BWA-MEM v.0.7.17 (Li, 2013). The alignment process incorporated soft clipping for supplementary alignments. Samtools v.1.2 (Li et al., 2009) was used to filter low-quality reads based on the following criteria: (1) mapping quality score less than 20, (2) multiple best hits, (3) unmapped reads (FLAG tag of 4), and (4) adjusted mapping quality below 50. The Genome Analysis

Toolkit (GATK) v.4.0.11.0 (Auwerwa and O'Connor, 2020) was utilized for read deduplication with MarkDuplicates, while variant calling was performed using HaplotypeCaller. The parameters for HaplotypeCaller included a minimum base quality score of 20, output-mode set to EMIT_ALL_ACTIVE_SITES, and ERC set to GVCF. The resulting GVCF files from all samples were merged via GenomicsDBImport and subjected to genotyping, filtering, and refinement processes using GenotypeGVCFs, SelectVariants, and VariantFiltration. Filtering parameters applied were 'QD < 2.0 || MQ < 40.0 || FS > 60.0 || SOR > 3.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0'.

Further SNP filtering was performed with bcftools (Li, 2011) to retain only bi-allelic sites and sites with a minor allele count (MAC) ≥ 2 . PLINK v.1.90b6.10 (Purcell et al., 2007) was used to filter out sites with high linkage ($r^2 > 0.1$). Population genetic structure was assessed using Bayesian clustering in ADMIXTURE v.1.23 (Alexander et al., 2009) based on unlinked SNPs. The number of genetic clusters (K) was set to vary from 2 to 8, with the optimal K determined by the lowest cross-validation error. To investigate population genetic relationships further, the SNP dataset was subjected to a principal component analysis (PCA) in GCTA v.1.24 (Yang et al., 2011) and a maximum likelihood (ML) tree analysis in IQ-TREE v.2.0 (Minh et al., 2020). For nucleotide alignment, filtered unlinked SNPs were processed using vc2phyloip v.2.0 (Ortiz, 2019). The alignment was trimmed with trimAl v.1.4 (Capella-Gutiérrez et al., 2009) to remove ambiguous sites. Phylogenetic relationships of all individuals were then constructed using IQ-TREE based on the unlinked SNPs.

2.7. Demographic history inference

To estimate the changes in population size over time for each species, we estimated folded site frequency spectrum (SFS) using easySFS.py (<https://github.com/isaacovercast/easySFS>), incorporating all samples from the six species. The SFS was then input into Stairway Plot 2 (Liu and Fu, 2020) to reconstruct demographic history for each species. A mutation rate (μ) of 2.5×10^{-8} per site per generation and a generation time of one year were used for the calculations. For simulating a comprehensive demographic scenario that includes divergence, population size changes, and gene flow among the six species, we utilized coalescent simulations with the composite-likelihood method implemented in fastsimcoal v.2.6 (Excoffier et al., 2013). Joint folded 2D-SFS were constructed using unlinked SNPs with easySFS.py. To minimize the impact of varying levels of missing data among groups, SNP data for each group were down-projected to an SFS with equal sampling size (26) across groups.

Given the different topologies in the phylogenetic trees reconstructed by SCNGs and unlinked SNPs, we initially tested four models to determine the most likely divergence scenario among the six species (Fig. S1). The existence and magnitude of gene flow between species, as well as changes in population size, were incorporated into the models, with additional manual fine-tuning of the parameters by comparing the differences between simulated and observed SFS (see supplementary dataset). Prior distributions of model parameters were set based on results from BEAST and Stairway Plot 2 analyses. For each scenario, we conducted 100,000 coalescent simulations per likelihood estimation ($-n$ 100,000) and performed 40 expectation-conditional maximization (ECM) cycles ($-L$ 40) as command line parameters for each run. The Akaike information criterion (AIC) was used to compare different models, where $AIC = 2k - 2\ln(\text{MaxEstLhood})$, with k representing the number of parameters estimated by each model, and MaxEstLhood being the maximum likelihood function value for each model. The best scenario was run 100 times to obtain optimal

parameter estimation. Subsequently, 100 independent DNA polymorphism datasets were simulated as joint SFSs conditional on the estimated demographic parameters. Maximum likelihood (ML) analysis was then applied to each joint SFS over 40 ECM cycles to obtain confidence intervals (CIs) for the final estimates.

2.8. Genomic diversity and detection of selection

Nucleotide diversity (π) for each species and population genetic differentiation (F_{ST}) between each pair of species were calculated using pixy v.1.2.7 (Korunes and Samuk, 2021), focusing on the first 10 kb of each transcript. To detect signatures of positive selection during the speciation of flowers with two distinct types (recurved and trumpet), we grouped the species as follows: *Lilium henryi* and *L. rosthornii* (recurved flowers) and *L. leucanthum*, *L. sargentiae*, *L. sulphureum* (trumpet flowers). Given that *L. regale* is restricted to a small region and is independent both geographically and genetically (see results), this species was excluded from the selection analysis to minimize the confounding effects of genetic drift.

For detecting regions under selection, we identified transcripts showing significantly elevated genetic differentiation (top 5% F_{ST}) between the flower groups and significant reductions in genetic diversity within either group (top and tail 5% π ratio). These transcripts were considered candidates for divergent selection. Further, functional enrichment analysis, including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses, was performed on the orthologous genes in *Arabidopsis thaliana* using KOBAS v.3.0 (Bu et al., 2021).

3. Results

3.1. Species level phylogeny reconstructed by transcriptome-based orthologous genes

The representative transcriptome assembled for each species ranged from 370 to 516 Mb in length, with scaffold N50 ranging from 823 to 1144 bp (Table S3). The completeness of the assemblies, as assessed by BUSCO, ranged from 89.9% (2909) to 94.9% (3071) (Fig. S2). Through Orthofinder2, a total of 7050 single-copy orthologous genes (SCNGs) shared by the six species were predicted. A maximum likelihood (ML) phylogenetic tree constructed based on these SCNGs revealed two major clades: one clade consisted of the two species with recurved flowers (*Lilium henryi* and *L. rosthornii*), while the other included the remaining four species with trumpet flowers (Fig. 2A and B). Within the latter clade, *L. regale* was identified as the basal taxon, followed by the divergence between *L. leucanthum* and the clade containing *L. sargentiae* and *L. sulphureum* (Fig. 2A). BEAST analysis estimated the crown age of the six species to be approximately 2.36 Mya. The divergence between *L. henryi* and *L. rosthornii* was estimated to have occurred around 0.63 Mya, while the diversification of the clade containing the four trumpet-flowered species was dated between 1.08 Mya and 0.67 Mya (Fig. 2A).

3.2. Flowering time and ecological differentiation of six *Lilium* species

To assess variation in flowering time among six *Lilium* species, we performed non-parametric tests using monthly data from independent samples. The results revealed significant differences in flowering time between the two recurved species, *L. henryi* and *L. rosthornii* ($P < 0.05$), although these differences were not significant after applying the Bonferroni correction ($P > 0.05$; Fig. 2C and

Table S4). Significant differences in flowering time were also found among the four species with trumpet flowers, with the exception of the pair *L. sargentiae* and *L. leucanthum* (Fig. 1C and Table S4). No significant differences were observed in flowering time between the recurved *L. henryi* and the trumpet-flowered *L. sulphureum* (Fig. 2C and Table S4). Moreover, the flowering time of *L. regale* did not resemble that of any other species or species sets (Table S4).

Ecological niche analysis indicated substantial niche overlap between the recurved-flower species *Lilium henryi* and *L. rosthornii* in both ecological (E-space) and geographic (G-space) contexts, with Schoener's D values of ≥ 0.446 ($P < 0.05$; Table 1). Similar patterns were found among the three trumpet-flower species (*L. sulphureum*, *L. sargentiae*, and *L. leucanthum*), where Schoener's D values were ≥ 0.454 , most of which were statistically significant. In contrast, *L. regale* exhibited few niche overlaps with the other trumpet-flowered species (Schoener's $D \leq 0.022$) and no overlap (Schoener's $D = 0$) with the recurved-flower species (Table 1). Members of recurved-flower species also showed relatively minimal overlap with those of the trumpet-flower species, particularly in E-space, with Schoener's D values of ≤ 0.162 (Table 1).

3.3. Full-length transcriptome assembly, population structure and phylogenetic relationships

A full-length transcriptome reference was obtained using the PacBio sequencing, which generated a total of 12,853,869 subreads, encompassing 21.36 Gb of data. After removing redundancy, the assembled transcriptome reached a length of 33,483,797 bp ($N_{50} = 2743$), containing 31,913 transcripts and 15,670 genes. Functional annotation of the transcriptome showed that 15,002 genes were annotated by at least one database (Table S5).

All population samples and outgroups were mapped against this reference transcriptome. Subsequent haplotype calling and filtering procedures resulted in 3,302,381 SNPs, with 624,637 unlinked SNPs retained for further population structure and phylogenetic analyses. Admixture analysis demonstrated a clear population structure from $K = 2$ to $K = 8$, with the lowest cross-validation error observed at $K = 3$ (Fig. S3). When $K = 2$, individuals of the two flower types formed two distinct clusters, and an additional cluster emerged for *Lilium regale* at $K = 3$ (Fig. 3A). At $K = 6$, the two species with recurved flowers were recognized by different genetic clusters, whereas the three trumpet-flower species (*L. sulphureum*, *L. sargentiae*, and *L. leucanthum*) exhibited signs of admixture and unstable ancestries (Fig. 3A). Principal Component Analysis (PCA) corroborated these findings, revealing three distinct genetic clusters and a close relationship among the three trumpet-flower species (Fig. 3B).

A phylogenetic tree constructed from unlinked SNPs showed a topology (Fig. 3C) similar to the species phylogeny (Fig. 2A). However, the relationships among individuals from the three species with trumpet flowers differed from the species-level phylogeny, with *Lilium leucanthum* nested within *L. sulphureum*, and both species forming a clade sister to *L. sargentiae* (Fig. 3C). The plastome haplotype tree presented a different topology, revealing conflicts between the two recurved-flower species and among the three trumpet-flower species (Fig. 3D). This suggests the possibility of hybridization and/or introgression within these groups.

3.4. Demographic history and gene flow

Stairway Plot 2 analysis revealed relatively stable population sizes for *Lilium leucanthum*, *L. regale* and *L. henryi*. In contrast, the other three species experienced varying degrees of population

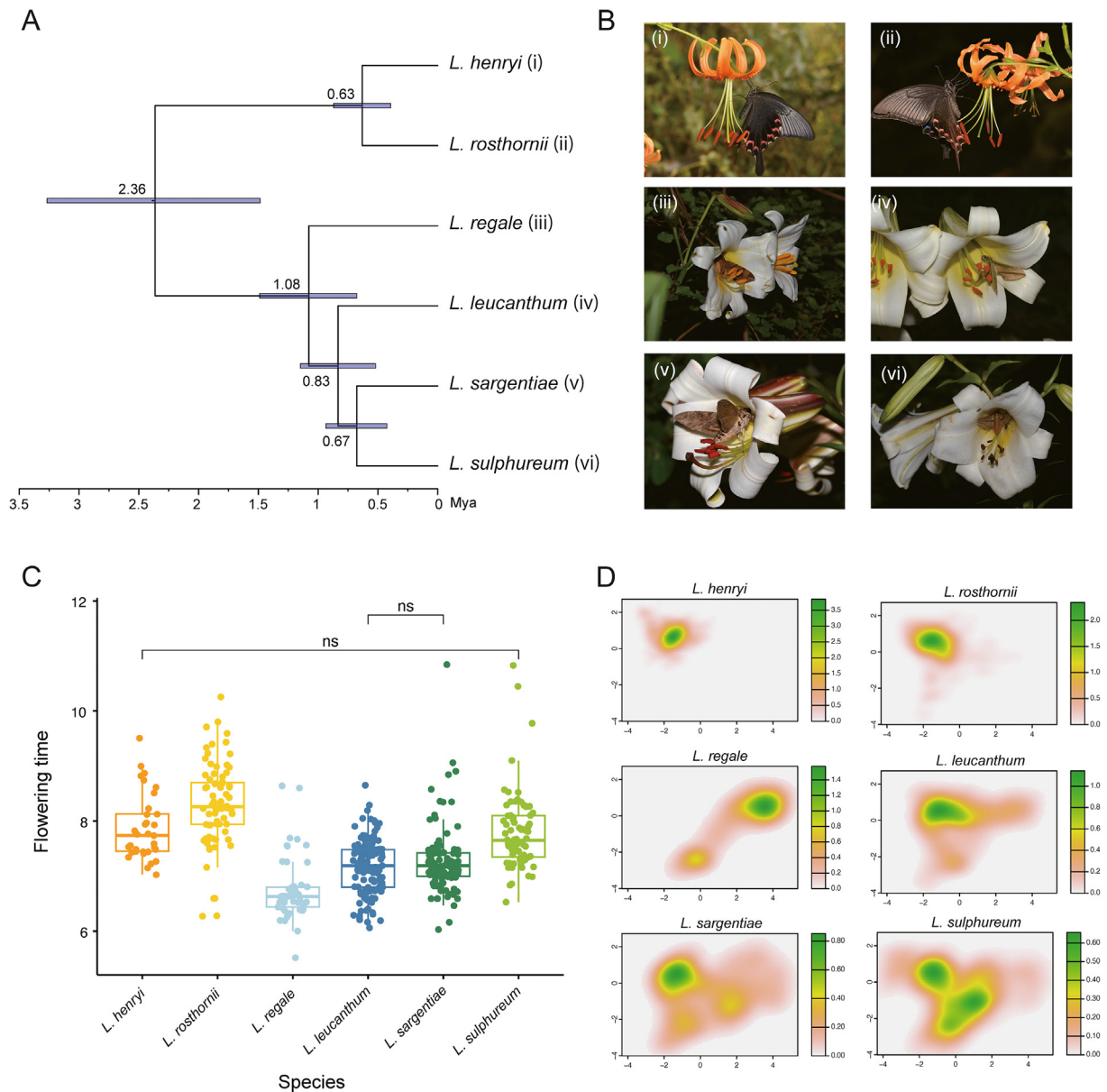


Fig. 2. Species phylogeny, flowering time, and niche differentiation of six *Lilium* species. (A) Dated phylogenetic tree reconstructed using BEAST based on single-copy orthologous genes shared by six *Lilium* species. Node heights indicate estimated divergence times, with bars indicating 95% highest posterior density (HPD) intervals. (B) Photos of flower and representative pollinators for (i) *L. henryi*, (ii) *L. rosthornii*, (iii) *L. regale*, (iv) *L. leucanthum*, (v) *L. sargentiae*, and (vi) *L. sulphureum*. (C) Boxplot showing distribution of flowering times among six *Lilium* species. Pairwise comparisons showed significant differences ($P < 0.05$) among all species, except for the two pairs marked as 'ns' (not significant). (D) Illustration of the density distribution on the first two environmental principal components for six *Lilium* species, analyzed by the 'ecospat' package.

Table 1
Schoener's D values for species pairs of six *Lilium* species in E-space (above the diagonal) and in G-space (below the diagonal).

Species	<i>L. henryi</i>	<i>L. rosthornii</i>	<i>L. regale</i>	<i>L. leucanthum</i>	<i>L. sargentiae</i>	<i>L. sulphureum</i>
<i>L. henryi</i>	1	0.667	0	0.382	0.373	0.317
<i>L. rosthornii</i>	0.446	1	0	0.2039	0.508	0.397
<i>L. regale</i>	0.022	0.022	1	0.173	0.093	0.131
<i>L. leucanthum</i>	0.139	0.5	0.126	1	0.533	0.454
<i>L. sargentiae</i>	0.078	0.134	0.006	0.5419	1	0.587
<i>L. sulphureum</i>	0.101	0.162	0.076	0.609	0.487	1

Note: Bolded values indicate that niche overlap for species pairs is significantly greater than the null hypothesis ($P < 0.05$).

contraction, notably significant in *L. sulphureum* and *L. sargentiae* (Fig. 4A). A composite likelihood test of demographic models (Table S6 and Fig. S1) indicated that *L. leucanthum* diverged from the most recent common ancestor (MRCA) of *L. sulphureum* and *L. sargentiae*. This finding aligns with the species phylogeny constructed using single-copy orthologous genes (SCNGs).

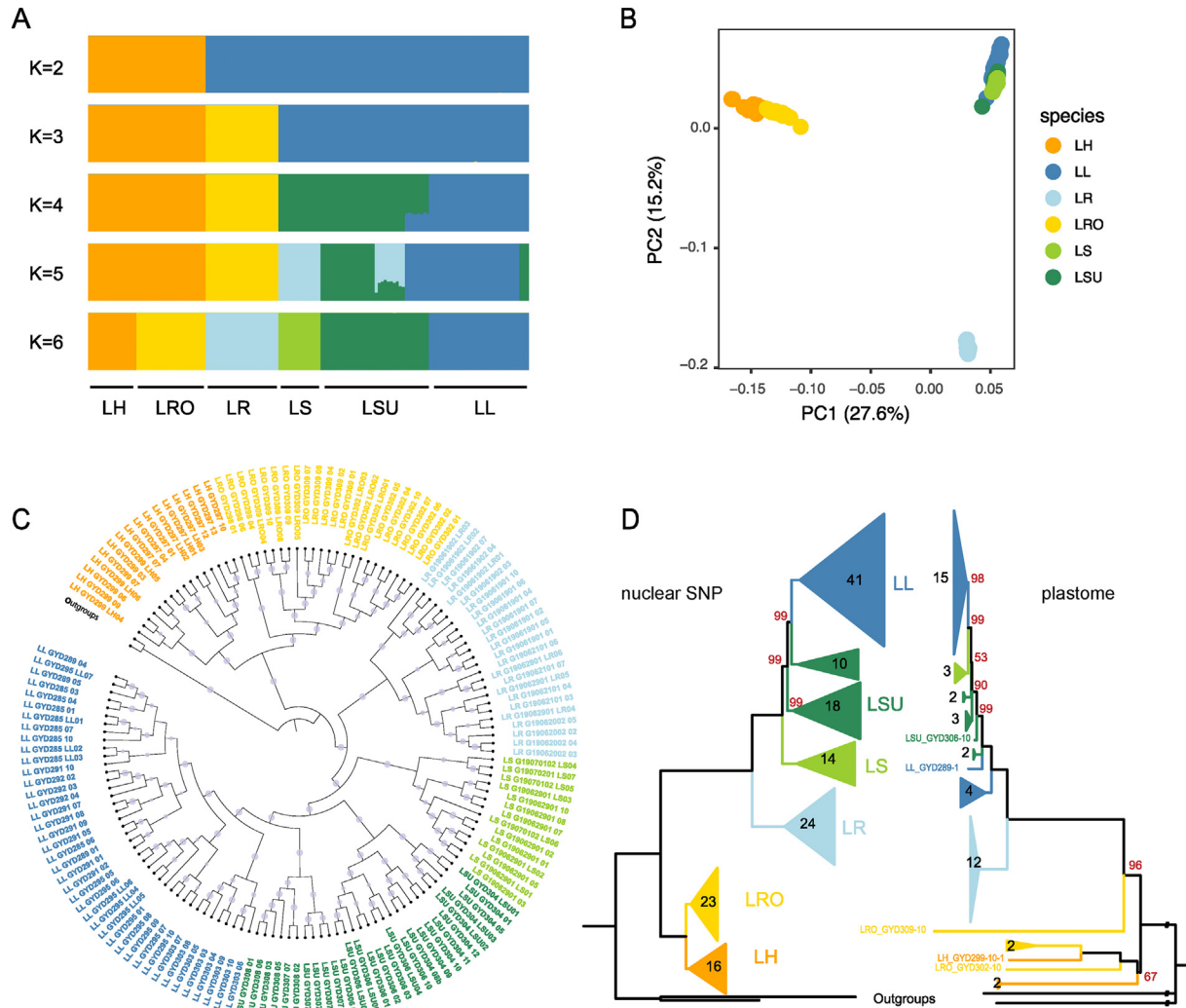


Fig. 3. Population genetic structure and phylogenetic relationships of six *Lilium* species. (A) Genetic structure inferred by ADMIXTURE, with K values from 2 to 6. (B) Principal Component Analysis (PCA) clusters, showing the plot of the first two principal components. (C) Maximum Likelihood tree reconstructed based on unlinked SNPs. Circle size on the branch denotes clade support. (D) Comparison of topologies inferred from nuclear SNP and plastome haplotype datasets. Numbers marked on/next to the clade indicate the number of individuals (nuclear tree) and the number of haplotypes (cp genome tree). Support values less than 100 are indicated by red numbers beside the nodes. Sample code is indicated on the cp genome tree if there is only one haplotype. LH, *L. henryi*; LRO, *L. rosthornii*; LR, *L. regale*; LS, *L. sargentiae*; LSU, *L. sulphureum*; LL, *L. leucanthum*.

Throughout the diversification of the six species, multiple instances of gene flow were detected, exhibiting various magnitudes and directions (Fig. 4B). Notably, continuous symmetric gene flow ($M3 = 6.21e-7$) was detected among *Lilium leucanthum*, *L. sulphureum*, and *L. sargentiae* since their divergence. Additionally, a significant pulse of symmetric gene flow occurred between *L. henryi* and *L. rosthornii* at approximately 15–20 thousand years ago (Kya) ($M1 = 3.87e-5$). Significant asymmetric gene flow was observed from the MRCA of species with trumpet flowers to the MRCA of the recurved flower clade at the onset of their differentiation (2.1–1.99 Mya, $M9 = 3.28e-5$; 1.99–1.23 Mya, $M8 = 6.06e-4$). Other instances of weak gene flow included that between *L. regale* and other trumpet-flowered species ($M2 = 1.09e-7$), as well as from trumpet-flowered species to recurved-flowered species since 1.23 Mya ($M7 = 1.3e-8$ from 1.23 Mya to 381 Kya; $M5 = 7.93e-8$ from 381 Kya to 20 Kya; $M4 = 5.6e-7$ from 20 Kya to 15 Kya). All gene flow events ceased after 15 Kya.

3.5. Genetic diversity and signatures of positive selection

Estimates of nucleotide diversity (π) suggested that the three species with trumpet flowers (*Lilium leucanthum*, *L. sulphureum*, and *L. sargentiae*) have relatively high genetic diversity compared to other species, while *L. regale* has lower π values. *L. rosthornii*, with recurved flowers, has the lowest genetic diversity (Fig. 5A). We observed significant genetic differentiation between the two flower shape types, with an average F_{ST} of 0.55 or higher (Fig. 5B). Genetic differentiation among the three species with trumpet flowers was minimal, likely due to continuous gene flow ($F_{ST} \leq 0.11$). However, *L. regale* was genetically independent from the other three trumpet-flowered species ($F_{ST} \geq 0.25$; Fig. 5B). The two species with recurved flowers also exhibited genetic divergence from each other ($F_{ST} = 0.23$; Fig. 5B).

A genomic scan for elevated genetic divergence between groups and lowered genetic diversity within either group identified 105

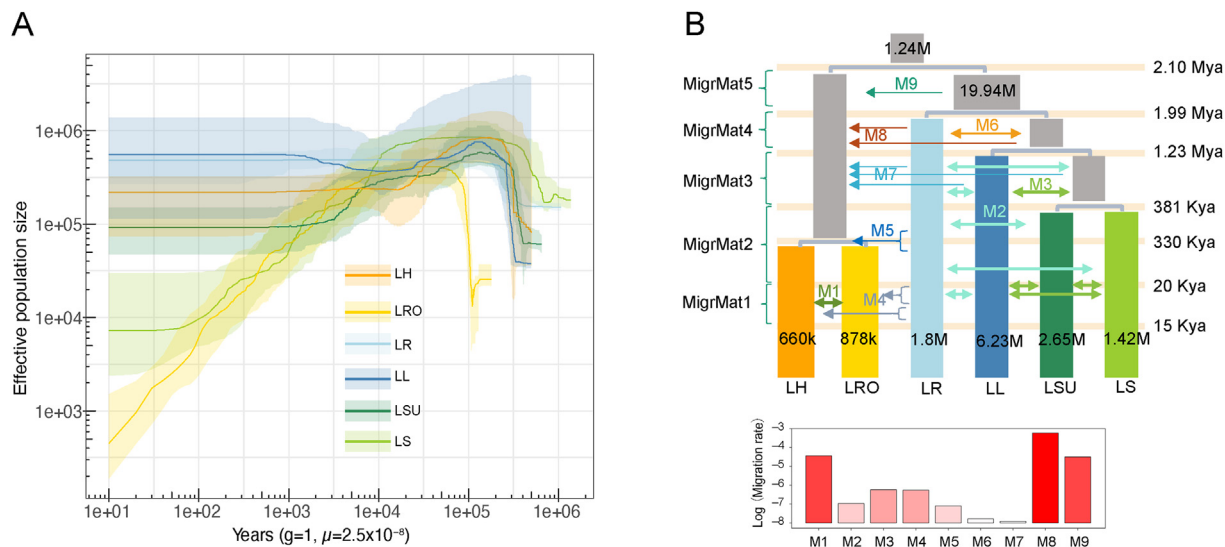


Fig. 4. Demographic history of six *Lilium* species. (A) Changes in population size through time inferred by Stairway Plot 2. (B) Illustration of the best-fit demographic scenario describing divergence and gene flow during the diversification of six *Lilium* species. The timeline for divergence and changes in migration matrices are indicated on the right. For detailed model configurations, refer to supplementary materials. See Fig. S4 for parameter names. Arrows indicate the existence and direction of gene flow, with colors corresponding to parameter names. Arrows with curly brackets indicate gene flow from all trumpet-flowered species. The estimated magnitude of gene flow is illustrated by bar plots at the bottom. Estimated effective population size is indicated for each population.

transcripts under potential positive selection (Fig. 5C). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses revealed that these genes are involved in responses to environmental stimuli (e.g., circadian rhythm and MAPK signaling pathway), various metabolic and biosynthetic processes (e.g., anthocyanin metabolism), and different cell wall and structural pathways (e.g., cellulose biosynthesis; Fig. 5D).

4. Discussion

The completely different flora forms (two forms in six species) in our study exemplify parallel evolution in the genus *Lilium*, which has been widely documented in previous studies (Gao et al., 2015; Yuan and Gao, 2024). This phenomenon creates challenges for systematic classification due to discordances between morphology-based taxonomy and molecular phylogeny. These distinct floral morphologies (whitish-trumpet and orange-recurved) were assigned to different sections based on their strong resemblance to these funnel and ‘turkey-cap’ members assembled as sect. *Leucolirion* and sect. *Sinomatagon*, respectively (Comber, 1949). However, molecular studies have confirmed their monophyletic status within *Lilium* (Gao et al., 2013; Du et al., 2014; Li et al., 2022). This monophyletic yet dimorphic lily group provides an excellent case study for investigating speciation and species boundary maintenance. Our comprehensive analysis of the six *Lilium* species using transcriptome-based orthologous genes, population genomics, flowering time data, and ecological data has provided valuable insights into their evolutionary history, speciation processes, and adaptive strategies. Our results indicate that diversification of these species has been shaped by a complex interplay between phylogenetic relationships, flower morphology, ecological differentiation, and gene flow.

4.1. History of species divergence within clade *Leucolirion*

Our transcriptome-based molecular dating estimates the crown age of the six species at early Pleistocene (ca. 2.36 Mya; Fig. 2A), when the recurved and trumpet flower groups diverged from each other. Subsequent diversification events within groups ranged from

the early to middle Pleistocene (between 1.08 Mya and 0.63 Mya), a period marked by pronounced climate fluctuations and glacial cycles that promoted rapid angiosperm diversification (Liu et al., 2023). Global climate oscillation during Pleistocene drove numerous speciation and lineage divergence events, causing rapid radiation in many genera, as reported in *Rhododendron* (Xia et al., 2022), *Rodgersia* (Ma et al., 2017) and *Primula* (Ren et al., 2015). However, range expansion-contraction during the glacial-interglacial cycle and/or incomplete range isolation often cause genetic exchange during divergence (Chen et al., 2019). Indeed, our demographic model, inferred by fastsimcoal2 (Fig. 4B), reveals significant intragroup gene flow, indicating historical hybridization or introgression. This genetic exchange likely contributed to reticulated evolution and, in some cases, blurred species boundaries. Gene flow can result in shared maternal genotypes among closely related species, potentially explaining the discrepancies we observed between nuclear and plastid phylogenies, particularly among the trumpet-flowered species (Fig. 3D; Du et al., 2009; Mo et al., 2022). Specifically, significant symmetric gene flow occurred between *Lilium henryi* and *L. rosthornii* approximately 15–20 Kya (Fig. 4B), likely causing their plastome haplotypes to nest with each other (Fig. 3D). Additionally, gene flow from trumpet members to recurved members resulted in chloroplast capture, leading some plastome haplotypes of recurved members to cluster with those of trumpet members (e.g., LRO_GYD309-10; Fig. 3D). Similarly, gene flow among the three trumpet-flowered species has shaped complex evolutionary relationships among their plastome haplotypes: some individuals of *L. leucanthum* were located at the base of the clade, reflecting either earlier divergence or potential gene flow from *L. regale* (Fig. 4B), while others clustered with *L. sulphureum* and *L. sargentiae*, resulting from more frequent gene flow among them.

4.2. Clear boundary between the recurved and trumpet flower groups

The separation of this clade into recurved and trumpet flower groups is strongly supported by multiple molecular datasets. The transcriptome-based phylogeny, constructed from 7050 single-

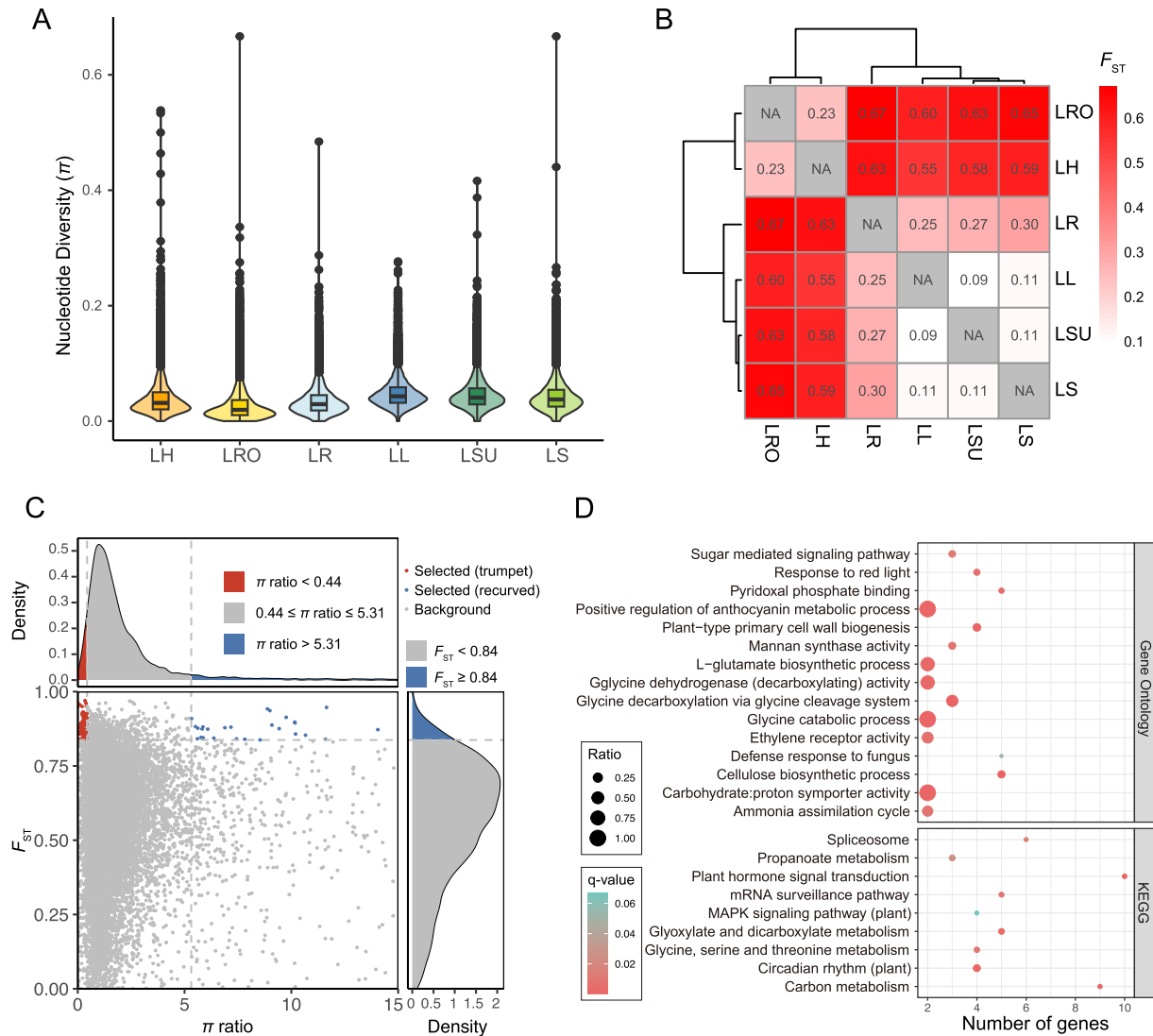


Fig. 5. Genetic diversity and signatures of positive selection between recurved and trumpet lily groups. (A) Boxplot showing the distribution of nucleotide diversity (π) for six *Lilium* species. (B) Heatmap plot indicating genetic differentiation (F_{ST}) between species pairs. (C) Distribution of F_{ST} and π ratio (trumpet/recurved group) values for each gene. The 5% left and right tails for π ratio distribution are indicated by left and right vertical dashed lines, respectively. The horizontal dashed line above indicates the distribution of the top 5% of F_{ST} . Identified genomic regions under selection are shown by red points for the trumpet group and blue points for the recurved group. (D) Representative GO and KEGG terms enriched by genes under selection in these two groups of *Lilium* species.

copy orthologous genes, clearly delineates two major clades corresponding to these flower morphologies (Fig. 2). This division is further corroborated by the nuclear SNP data, which shows distinct genetic clusters for recurved and trumpet-flowered species in both admixture and principal component analyses (Fig. 3), with significant genetic differentiation between groups ($F_{ST} \geq 0.55$). Pollinator-mediated reproductive isolation plays a crucial role in maintaining species boundaries (Grant, 1994; Brothers et al., 2014; Van der Niet et al., 2014). Our previous research on pollinator differentiation (Liu et al., 2022) demonstrated distinct pollination syndromes between recurved and trumpet-flowered *Lilium* species, with butterflies visiting the former during the day and hawk moths pollinating the latter at night. Therefore, the adaptation to different pollinators seems to play a quite important role in isolation of the two major lineages, despite some overlaps in flowering time and geographic distribution (Fig. 2 and Table 1).

Ecological niche analysis further supports this differentiation, showing substantial niche overlap within flower types but limited

overlap between them (Fig. 2D and Table 1). The genes under potential positive selection between the two flower types, particularly those involved in responses to environmental stimuli, metabolic processes, and structural pathways, provide a genetic basis for the observed ecological differentiation (Fig. 5C and D) despite the presence of gene flow (Fig. 4B). The most obvious difference between the two groups is flower color (whitish-trumpet and orange-recurved; Fig. 2B), which is reflected in the positive selection of genes related to anthocyanidin metabolic process (Fig. 5D). Notably, these species have evolved different adaptations to attract distinct pollinators—the trumpet-flowered species are pollinated by hawk moths at night, while butterflies visit the recurved flowers during the day—with multiple aspects of genetic evidence supporting this differentiation. For example, the significant difference in nectar volume between groups reported in Liu et al. (2022) is evidenced in the positive selection of genes related to pathways and biological processes included carbon metabolism, carbohydrate, and sugar mediated signaling pathways. Additionally, genes associated with

circadian rhythm (Fig. 5D) could contribute to the temporal display of attractive qualities, such as scent emission in a daily rhythmic fashion to attract pollinators (e.g., Fenske et al., 2018). These genetic adaptations highlight how selective pressures on specific pathways can drive ecological differentiation, especially pollinator-related divergence, allowing closely related species to occupy distinct temporal and environmental niches while maintaining reproductive isolation. Such adaptive changes likely facilitate the occupation of diverse niches and contribute to the maintenance of species boundaries despite potential gene flow (Nosil, 2008).

4.3. Asynchronous differentiation within groups

While substantial differentiation was found between the two groups of flower types, more complex patterns emerged within each group. Moderate genetic differentiation was observed between *Lilium regale* and other trumpet-flower species, as well as between the two recurved-flower species (Fig. 5B). *L. regale* showed nearly no gene flow with its trumpet-flowered relatives, and gene flow between the two recurved-flower species existed only for a short period (Fig. 4B). This moderate differentiation aligns with our findings on flowering time and ecological niche analysis (Fig. 2C and Table 1). For instance, *L. regale* blooms nearly a month earlier than the other trumpet members and exhibits little to no ecological and geographic overlap with other species. Despite sharing both distribution ranges and pollinators (hawk moths) with *L. leucanthum*, *L. regale* maintains substantial genetic separation from the latter (Fig. 3). In fact, the floral tubes of *L. regale* were significantly shorter than other trumpet-flowered species, attracting more hawk moths with short proboscises, such as *Deilephila elpenor* and *Ampelophaga rubiginosa* (Liu et al., 2022), likely contributing to its more independent evolution. Similarly, the recurved members *L. henryi* and *L. rosthornii* demonstrate a tendency for flowering time and niche differentiation. Specifically, *L. henryi* prefers sunny, drought-prone slopes, whereas *L. rosthornii* inhabits humid valleys along streams. These lines of evidence indicate a potential contribution to reproductive isolation within groups through temporal isolation (Coyne and Orr, 2004; Gaudinier and Blackman, 2020) and niche divergence (Schluter, 2000; Schluter and McPeck, 2000; Wogan and Richmond, 2015).

In contrast, the other three trumpet-flowered species (*Lilium leucanthum*, *L. sulphureum*, and *L. sargentiae*) show limited genetic differentiation among themselves (Fig. 5B), consistent with the continuous symmetric gene flow ($M3 = 6.21e-7$) detected since their divergence (Fig. 4B). This limited differentiation is further reflected in their substantial overlap in geographic distribution and ecological niches (Table 1). Additionally, unlike *L. regale*, with its relatively smaller flowers, these three species exhibit large trumpet-shaped flowers that are not only adapted to long-proboscid pollinators but also attract various hawkmoth species, as those with shorter proboscises can alight and crawl into the flowers in search of nectar, benefiting from an expanded pollination niche breadth (Liu et al., 2022). This enables potential cross-specific visits that may lead to gene flow between species and blur the species boundaries. However, trends toward divergence are evident. For example, the divergence between *L. leucanthum* and the *L. sargentiae* + *L. sulphureum* clade is supported by both molecular (Figs. 2A and 4B) and morphological evidence, with both *L. sargentiae* and *L. sulphureum* possessing axil bulblets while *L. leucanthum* lacks them (Liang and Tamura, 2000). Besides, *L. sulphureum* blooms significantly later than *L. leucanthum* and *L. sargentiae* (Fig. 2C). The demographic history inferences also reveal diverse population size trajectories: while *L. sargentiae* has experienced recent population contractions, *L. leucanthum* and *L. sulphureum* show population stability (Fig. 5B). These varied

histories likely reflect species-specific responses to historical climate and geographical changes, illustrating ecological resilience and the influence of historical biogeography on contemporary population dynamics (Avise, 2000).

5. Conclusions

In summary, the evolution of the Leucolirion clade is characterized by an initial split between recurved and trumpet-flowered lineages, driven by morphological adaptations to different pollination strategies. Subsequent diversification within each group involved flowering time shifts, ecological niche differentiation, and localized adaptations. The recurved-flower species (*Lilium henryi* and *L. rosthornii*) diverged primarily through temporal isolation, while trumpet-flowered species exhibit more complex diversification patterns. *L. regale* stands out with its distinct niche and genetic independence, whereas ongoing gene flow among *L. leucanthum*, *L. sulphureum*, and *L. sargentiae* suggests closer relationships and potentially blurred species boundaries. These findings collectively support the idea that species divergence and maintenance in *Lilium* are driven by a combination of adaptive and non-adaptive processes, including pollinator-mediated selection, ecological niche differentiation, historical demographic events, and natural selection. This research not only advances our understanding of *Lilium* evolution but also offers insights into the processes that underlie plant speciation and adaptation. Future studies incorporating broader species sampling and more detailed ecological data will further elucidate the evolutionary mechanisms in this diverse plant group. Notably, the recently published *L. davidii* genome (Xu et al., 2024) could enable more accurate phylogenetic reconstructions and provide deeper insights into the evolutionary history of *Lilium*. This genomic resource could also reveal structural variations that help elucidate the mechanisms driving adaptation and speciation within the genus, particularly in the context of ecological differentiation and reproductive isolation.

CRedit authorship contribution statement

Yu Feng: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Chaochao Yan:** Writing – original draft, Visualization, Methodology, Formal analysis, Data curation. **Wen-Qin Tu:** Writing – original draft, Methodology, Formal analysis, Data curation. **Yu-Mei Yuan:** Writing – original draft, Visualization, Formal analysis, Data curation. **Jing-Bo Wang:** Writing – original draft, Software, Formal analysis. **Xiao-Juan Chen:** Writing – original draft, Visualization, Formal analysis. **Chang-Qiu Liu:** Validation, Methodology, Investigation. **Yundong Gao:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Data availability

The sequencing data generated in this study has been deposited to NCBI SRA database under the BioProject accession PRJNA1141931. Supplementary datasets including configuration files and evaluation of demographic scenarios tested by fastsimcoal2 were deposit to Mendeley Data (<https://doi.org/10.17632/c7sgdwvrzh.1>). Other data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We gratefully acknowledge Prof. Xin-Fen Gao and Mr. Jun-Yi Zhang of Chengdu Institute of Biology, Chinese Academy of Sciences, Dr. Pan Li of Zhejiang University, and Mr. Melvyn Herbert for their support with sample collection. This work was supported by the National Natural Science Foundation of China (NSFC Grant Nos. 32171605, 31770406) and Natural Science Foundation of Sichuan Province (Grant No. 2023NSFSC0141) to Y.D. Gao.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.pld.2024.12.005>.

References

- Akçakaya, H.R., Burgman, M.A., Kindvall, O., et al., 2004. *Species Conservation and Management: Case Studies*. Oxford University Press, Oxford, New York.
- Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. <https://doi.org/10.1101/gr.094052.109>.
- Auwera, G. van der, O'Connor, B.D., 2020. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. O'Reilly Media, Incorporated, Sebastopol, California.
- Avise, J.C., 2000. *Phylogeography: the History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts.
- Blanckaert, A., Bank, C., Hermisson, J., 2020. The limits to parapatric speciation 3: evolution of strong reproductive isolation in presence of gene flow despite limited ecological differentiation. *Philos. Trans. R. Soc. B-Biol. Sci.* 375, 20190532. <https://doi.org/10.1098/rstb.2019.0532>.
- Bolnick, D.I., Fitzpatrick, B.M., 2007. Sympatric speciation: models and empirical evidence. *Annu. Rev. Ecol. Evol. Syst.* 38, 459–487. <https://doi.org/10.1146/annurev.ecolsys.38.091206.095804>.
- Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., et al., 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 15, e1006650. <https://doi.org/10.1371/journal.pcbi.1006650>.
- Brothers, A.N., Atwell, J.W., Caruso, E.C., 2014. The role of pollinator-mediated selection in the divergence of floral traits between two closely related plant species. *Int. J. Plant Sci.* 175, 287–295. <https://doi.org/10.1086/673883>.
- Brown, J.L., Carnaval, A.C., 2019. A tale of two niches: methods, concepts, and evolution. *Front. Biogeogr.* 11, e44158. <https://doi.org/10.21425/F5F8G44158>.
- Bu, D., Luo, H., Huo, P., et al., 2021. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res.* 49, W317–W325. <https://doi.org/10.1093/nar/gkab447>.
- Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>.
- Chen, J., Huang, Y., Brachi, B., et al., 2019. Genome-wide analysis of cushion willow provides insights into alpine plant divergence in a biodiversity hotspot. *Nat. Commun.* 10, 5230. <https://doi.org/10.1038/s41467-019-13128-y>.
- Chen, S., 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* 2, e107. <https://doi.org/10.1002/imt2.107>.
- Comber, H.F., 1949. A new classification of the genus *Lilium*. *R.H.S. Lily Year Book* 13, 86–105.
- Coyne, J.A., Orr, H.A., 2004. *Speciation*. Oxford University Press, Oxford, New York.
- Darriba, D., Posada, D., Kozlov, A.M., et al., 2020. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* 37, 291–294. <https://doi.org/10.1093/molbev/msz189>.
- De Queiroz, K., 2007. Species concepts and species delimitation. *Syst. Biol.* 56, 879–886. <https://doi.org/10.1080/10635150701701083>.
- Di Cola, V., Broennimann, O., Petitpierre, B., et al., 2017. ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecography* 40, 774–787. <https://doi.org/10.1111/ecog.02671>.
- Donoghue, M.J., 2022. What in the world is a species? *Arnoldia* 79, 48–53.
- Du, F.K., Petit, R.J., Liu, J.Q., 2009. More introgression with less gene flow: chloroplast vs. mitochondrial DNA in the *Picea asperata* complex in China, and comparison with other Conifers. *Mol. Ecol.* 18, 1396–1407. <https://doi.org/10.1111/j.1365-294X.2009.04107.x>.
- Du, Y., He, H., Wang, Z., et al., 2014. Molecular phylogeny and genetic variation in the genus *Lilium* native to China based on the internal transcribed spacer sequences of nuclear ribosomal DNA. *J. Plant Res.* 127, 249–263. <https://doi.org/10.1007/s10265-013-0600-4>.
- Duan, Q., Liu, F., Gui, D., et al., 2022. Phylogenetic analysis of wild species and the maternal origin of cultivars in the genus *Lilium* using 114 plastid genomes. *Front. Plant Sci.* 13, 865606. <https://doi.org/10.3389/fpls.2022.865606>.
- Emms, D.M., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 238. <https://doi.org/10.1186/s13059-019-1832-y>.
- Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., et al., 2013. Robust demographic inference from genomic and SNP data. *PLoS Genetics* 9, e1003905. <https://doi.org/10.1371/journal.pgen.1003905>.
- Fenske, M.P., Nguyen, L.P., Horn, E.K., et al., 2018. Circadian clocks of both plants and pollinators influence flower seeking behavior of the pollinator hawk-moth *Manduca sexta*. *Sci. Rep.* 8, 2842. <https://doi.org/10.1038/s41598-018-21251-x>.
- Gao, Y.-D., Gao, X.-F., Harris, A., 2019. Species boundaries and parapatric speciation in the complex of alpine shrubs, *Rosa sericea* (Rosaceae), based on population genetics and ecological tolerances. *Front. Plant Sci.* 10, 321. <https://doi.org/10.3389/fpls.2019.00321>.
- Gao, Y.-D., Harris, A., Zhou, S.-D., et al., 2013. Evolutionary events in *Lilium* (including *Nomocharis*, Liliaceae) are temporally correlated with orogenies of the Q–T Plateau and the Hengduan Mountains. *Mol. Phylogenet. Evol.* 68, 443–460. <https://doi.org/10.1016/j.ympev.2013.04.026>.
- Gao, Y.-D., Harris, A.J., He, X.-J., 2015. Morphological and ecological divergence of *Lilium* and *Nomocharis* within the Hengduan Mountains and Qinghai-Tibetan Plateau may result from habitat specialization and hybridization. *BMC Evol. Biol.* 15, 147. <https://doi.org/10.1186/s12862-015-0405-2>.
- Gaudinier, A., Blackman, B.K., 2020. Evolutionary processes from the perspective of flowering time diversity. *New Phytol.* 225, 1883–1898. <https://doi.org/10.1111/nph.16205>.
- Givnish, T., Skinner, M., Rešetnik, I., et al., 2020. Evolution, geographical spread and floral diversification of the genus *Lilium* with special reference to the lilies of North America. *Lily Yearb. North Am. Lily Soc. Inc* 74, 26–44.
- Grant, V., 1994. Modes and origins of mechanical and ethological isolation in angiosperms. *Proc. Natl. Acad. Sci. U.S.A.* 91, 3–10. <https://doi.org/10.1073/pnas.91.1.3>.
- Haas, B.J., Papanicolaou, A., Yassour, M., et al., 2013. *De novo* transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat. Protoc.* 8, 1494–1512. <https://doi.org/10.1038/nprot.2013.084>.
- Hayashi, K., Kawano, S., 2000. Molecular systematics of *Lilium* and allied genera (Liliaceae): phylogenetic relationships among *Lilium* and related genera based on the *rbcL* and *matK* gene sequence data. *Plant Species Biol.* 15, 73–93. <https://doi.org/10.1046/j.1442-1984.2000.00025.x>.
- Hijmans, R.J., Phillips, S., Leathwick, J., et al., 2022. dismo. Available at: version 1.3-9. <https://cran.r-project.org/web/packages/dismo/> (Accessed 21 May 2022).
- Hong, D.-Y., 2016. Biodiversity pursuits need a scientific and operative species concept. *Biodivers. Sci.* 24, 979. <https://doi.org/10.17520/biods.2016203>.
- Hu, L., Yang, R., Wang, Y.-H., et al., 2021. The natural hybridization between species *Ligularia nelumbifolia* and *Cremathodium stenoglossum* (Senecioneae, Asteraceae) suggests underdeveloped reproductive isolation and ambiguous intergeneric boundary. *AoB Plants* 13, plab012. <https://doi.org/10.1093/aobpla/plab012>.
- IBM Corp., 2020. *IBM SPSS Statistics for Windows (Computer software), Version 27.0*.
- Jin, J.-J., Yu, W.-B., Yang, J.-B., et al., 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* 21, 241. <https://doi.org/10.1186/s13059-020-02154-5>.
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. <https://doi.org/10.1093/molbev/mst010>.
- Kim, H.T., Lim, K.-B., Kim, J.S., 2019. New insights on *Lilium* phylogeny based on a comparative phylogenomic study using complete plastome sequences. *Plants* 8, 547. <https://doi.org/10.3390/plants8120547>.
- Korunes, K.L., Samuk, K., 2021. pixy: unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol. Ecol. Resour.* 21, 1359–1368. <https://doi.org/10.1111/1755-0998.13326>.
- Kozlov, A.M., Darriba, D., Flouri, T., et al., 2019. RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35, 4453–4455. <https://doi.org/10.1093/bioinformatics/btz305>.
- Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>.
- Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. <https://doi.org/10.48550/arXiv.1303.3997>.
- Li, H., Handsaker, B., Wysoker, A., et al., 2009. Genome Project Data Processing Subgroup, 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Li, J., Cai, J., Qin, H.-H., et al., 2022. Phylogeny, age, and evolution of tribe Liliaceae (Liliaceae) based on whole plastid genomes. *Front. Plant Sci.* 12, 699226. <https://doi.org/10.3389/fpls.2021.699226>.
- Li, Q., Xiang, C., Xu, L., et al., 2020. SMRT sequencing of a full-length transcriptome reveals transcript variants involved in C18 unsaturated fatty acid biosynthesis and metabolism pathways at chilling temperature in *Pennisetum giganteum*. *BMC Genomics* 21, 52. <https://doi.org/10.1186/s12864-019-6441-3>.
- Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>.
- Liang, S.-Y., Tamura, M., 2000. *Lilium* L. In: Wu, Z.Y., Raven, P.H. (Eds.), *Flora of China*, vol. 24. Science Press, Beijing; & Missouri Botanical Garden Press, St. Louis, pp. 135–159.
- Liu, C.-Q., Gao, Y.-D., Niu, Y., et al., 2019. Floral adaptations of two lilies: implications for the evolution and pollination ecology of huge trumpet-shaped flowers. *Am. J. Bot.* 106, 622–632. <https://doi.org/10.1002/ajb2.1275>.

- Liu, C.-Q., Niu, Y., Lu, Q.-B., et al., 2022. *Papilio* butterfly vs. hawkmoth pollination explains floral syndrome dichotomy in a clade of *Lilium*. *Bot. J. Linn. Soc.* 199, 678–693. <https://doi.org/10.1093/botlinnean/boab074>.
- Liu, J., 2016. The “integrative species concept” and “species on the speciation way.”. *Biodivers. Sci.* 24, 1004–1008. <https://doi.org/10.17520/biods.2016222>.
- Liu, X., Fu, Y.-X., 2020. Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biol.* 21, 280. <https://doi.org/10.1186/s13059-020-02196-9>.
- Liu, Y., Lai, Y.-J., Ye, J.-F., et al., 2023. The Sino-Himalayan flora evolved from lowland biomes dominated by tropical floristic elements. *BMC Biology* 21, 239. <https://doi.org/10.1186/s12915-023-01746-4>.
- Lu, W.-X., Wang, Z.-Z., Hu, X.-Y., et al., 2023. Echoes of the past: niche evolution, range dynamics, and their coupling shape the distribution of species in the *Chrysanthemum zawadskii* species complex. *Front. Ecol. Evol.* 11, 1250491. <https://doi.org/10.3389/fevo.2023.1250491>.
- Ma, X., Sun, W., Zhu, W., et al., 2017. Resolving the phylogenetic relationships and evolutionary history of the East Asian endemic genus *Rodgersia* (Saxifragaceae) using multilocus data. *Perspect. Plant Ecol. Evol. Syst.* 25, 20–28. <https://doi.org/10.1016/j.ppees.2016.12.005>.
- Mallet, J., 2008. Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Phil. Trans. Biol. Sci.* 363, 2971–2986. <https://doi.org/10.1098/rstb.2008.0081>.
- Manni, M., Berkeley, M.R., Seppey, M., et al., 2021. BUSCO: assessing genomic data quality and beyond. *Curr. Protoc.* 1, e323. <https://doi.org/10.1002/cp21.323>.
- Marasek-Ciolakowska, A., Nishikawa, T., Shea, D.J., et al., 2018. Breeding of lilies and tulips—Interspecific hybridization and genetic background—. *Breed. Sci.* 68, 35–52. <https://doi.org/10.1270/jsbbs.17097>.
- Mason-Gamer, R.J., White, D.M., 2024. The phylogeny of the Triticeae: resolution and phylogenetic conflict based on genome wide nuclear loci. *Am. J. Bot.* 111, e16404. <https://doi.org/10.1002/ajb2.16404>.
- Mayr, E., 1963. *Animal Species and Evolution*. Harvard University Press, Cambridge, Massachusetts. <https://doi.org/10.4159/harvard.9780674865327>.
- Minh, B., Schmidt, H., Chernomor, O., et al., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
- Mo, Z.-Q., Fu, C.-N., Zhu, M.-S., et al., 2022. Resolution, conflict and rate shifts: insights from a densely sampled plastome phylogeny for *Rhododendron* (Ericaceae). *Ann. Bot.* 130, 687–701. <https://doi.org/10.1093/aob/mcac114>.
- Nishikawa, T., Okazaki, K., Uchino, T., et al., 1999. A molecular phylogeny of *Lilium* in the internal transcribed spacer region of nuclear ribosomal DNA. *J. Mol. Evol.* 49, 238–249. <https://doi.org/10.1007/pl00006546>.
- Nosil, P., 2008. Speciation with gene flow could be common. *Mol. Ecol.* 17, 2103–2106. <https://doi.org/10.1111/j.1365-294X.2008.03715.x>.
- Ortiz, E.M., 2019. vcf2phylip v2.0: Convert a VCF Matrix into Several Matrix Formats for Phylogenetic Analysis. <https://doi.org/10.5281/zenodo.2540861>.
- Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31, 161–175. <https://doi.org/10.1111/j.0906-7590.2008.5203.x>.
- Purcell, S., Neale, B., Todd-Brown, K., et al., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. <https://doi.org/10.1086/519795>.
- Ren, G., Conti, E., Salamin, N., 2015. Phylogeny and biogeography of *Primula* sect. *Armerina*: implications for plant evolution under climate change and the uplift of the Qinghai-Tibet Plateau. *BMC Evol. Biol.* 15, 161. <https://doi.org/10.1186/s12862-015-0445-7>.
- Ronquist, F., Teslenko, M., van der Mark, P., et al., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. <https://doi.org/10.1093/sysbio/sys029>.
- Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., et al., 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. <https://doi.org/10.1093/molbev/msx248>.
- Schluter, D., 2000. *The Ecology of Adaptive Radiation*. Oxford University Press, Oxford, New York.
- Schluter, D., McPeck, A.E.M.A., 2000. Ecological character displacement in adaptive radiation. *Am. Nat.* 156, S4–S16. <https://doi.org/10.1086/303412>.
- Schoener, T.W., 1968. The Anolis Lizards of Bimini: resource partitioning in a complex fauna. *Ecology* 49, 704–726. <https://doi.org/10.2307/1935534>.
- Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56, 564–577. <https://doi.org/10.1080/10635150701472164>.
- Van der Niet, T., Peakall, R., Johnson, S.D., 2014. Pollinator-driven ecological speciation in plants: new evidence and future perspectives. *Ann. Bot.* 113, 199–211. <https://doi.org/10.1093/aob/mct290>.
- Wang, J., Dong, S., Yang, L., et al., 2020. Allopolyploid speciation accompanied by gene flow in a tree fern. *Mol. Biol. Evol.* 37, 2487–2502. <https://doi.org/10.1093/molbev/msaa097>.
- Warren, D.L., Glor, R.E., Turelli, M., 2010. ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* 33, 607–611. <https://doi.org/10.1111/j.1600-0587.2009.06142.x>.
- Wheeler, Q.D., Meier, R., 2000. *Species Concepts and Phylogenetic Theory*. Columbia University Press, New York.
- Wilkins, J.S., 2009. *Species: A History of the Idea*. University of California Press, Berkeley, California.
- Wogan, G.O.U., Richmond, J.Q., 2015. Niche divergence builds the case for ecological speciation in skinks of the *Plestiodon skiltonianus* species complex. *Ecol. Evol.* 5, 4683–4695. <https://doi.org/10.1002/ece3.1610>.
- Wu, D., Hu, Y., Akashi, S., et al., 2022. Lateral transfers lead to the birth of momilactone biosynthetic gene clusters in grass. *Plant J.* 111, 1354–1367. <https://doi.org/10.1111/tpj.15893>.
- Xia, X.-M., Yang, M.-Q., Li, C.-L., et al., 2022. Spatiotemporal evolution of the global species diversity of *Rhododendron*. *Mol. Biol. Evol.* 39, msab314. <https://doi.org/10.1093/molbev/msab314>.
- Xu, S., Chen, R., Zhang, X., et al., 2024. The evolutionary tale of lilies: giant genomes derived from transposon insertions and polyploidization. *Innovation* 5, 100726. <https://doi.org/10.1016/j.xinn.2024.100726>.
- Yang, J., Lee, S.H., Goddard, M.E., et al., 2011. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011>.
- Yuan, Y., Gao, Y., 2024. *Lilium liangiae*, a new species in the genus *Lilium* (Liliaceae) that reveals parallel evolution within morphology. *Front. Plant Sci.* 15, 1371237. <https://doi.org/10.3389/fpls.2024.1371237>.
- Zhang, W., Wang, H., Zhang, T., et al., 2023. Geographic–genomic and geographic–phenotypic differentiation of the *Aquilegia viridiflora* complex. *Hortic. Res.* 10, uhad041. <https://doi.org/10.1093/hr/uhad041>.
- Zhou, B.-F., Yuan, S., Crowl, A.A., et al., 2022. Phylogenomic analyses highlight innovation and introgression in the continental radiations of Fagaceae across the Northern Hemisphere. *Nat. Commun.* 13, 1320. <https://doi.org/10.1038/s41467-022-28917-1>.
- Zhou, N., Miao, K., Liu, C., et al., 2024. Historical biogeography and evolutionary diversification of *Lilium* (Liliaceae): new insights from plastome phylogenomics. *Plant Divers.* 46, 219–228. <https://doi.org/10.1016/j.pld.2023.07.009>.