Phylogenetic framework and evolution of Malvaceae s.l.

Authors
Wei-Long Yang, Bing Chen, Pei-Feng Liu, Shi-You Zuo, Dong-Zhu Jiang, ..., Hua-Feng Wang*, Hong-Yang Li*

Correspondence
hfwang@hainanu.edu.cn (Wang HF);
994484@hainanu.edu.cn (Li HY)

In Brief
Comparison of coalescent (ASTRAL) and concatenated maximum likelihood trees from nuclear data (left and right, respectively). Floral and fruit morphologies (A–V) correspond to major clades. Taxa: A, *Kleinhovia hospita*; B, *Ayenia grandifolia*; C–D, *Grewia biloba*; E–F, *Matisia cordata*; G–H, *Bombax ceiba*; I, *Talipariti tiliaceum*; J, *Hibiscus syriacus*; K–L, *Helicteres angustifolia*; M–N, *Durio zibethinus*; O, *Pentapetes phoenicea*; P, *Pterospermum heterophyllum*; Q–R, *Hainania trichosperma*; S, *Tilia tuan*; T, *Craigia yunnanensis*; U, *Brachychiton acerifolius*; V, *Sterculia lanceolata*. Photos: Weilong Yang, CVH, et al.

Graphical abstract

Highlights

- Malvaceae s.l. splits into two primary clades (Byttneriina and Malvadendrina) supported by 353 nuclear loci.

- Incomplete lineage sorting is the predominant driver (> 70% of loci) of nuclear gene tree discordance.

- Ancestral reconstructions indicate an African origin, with crown diversification at ~119.38 Ma.

- Significant topological discordance exists between nuclear and plastid phylogenies across eight subfamilies.

- Divergence time estimates suggest an Early Cretaceous origin (~134.31 Ma) and rapid radiation.

Citation: Yang WL, Chen B, Liu PF, Zuo SY, Jiang DZ, et al. 2026. Phylogenetic framework and evolution of Malvaceae s.l. *Tropical Plants* 5: e002 https://doi.org/10.48130/tp-0026-0001

Phylogenetic framework and evolution of Malvaceae s.l.

Wei-Long Yang[1,2], Bing Chen[1], Pei-Feng Liu[1], Shi-You Zuo[1,2], Dong-Zhu Jiang[1], Yu-Han Zhang[3], Hua-Feng Wang[1,4*] and Hong-Yang Li[5*]

[1] Sanya Nanfan Research Institute of Hainan University, Hainan Yazhou Bay Seed Laboratory, Sanya, Hainan 572025, China
[2] Collaborative Innovation Center of Nanfan and High-Efficiency Tropical Agriculture, School of Tropical Agriculture and Forestry, Hainan University, Haikou, Hainan 570228, China
[3] College of Agronomy and Biotechnology, China Agricultural University, Beijing 100083, China
[4] Key Laboratory of Tropical Biological Resources of Ministry of Education, College of Tropical Agricultural and Forestry, Hainan University, Haikou, Hainan 570228, China
[5] School of Breeding and Multiplication (Sanya Institute of Breeding and Multiplication), Hainan University, Sanya, Hainan 572025, China
* Correspondence: hfwang@hainanu.edu.cn (Wang HF); 994484@hainanu.edu.cn (Li HY)

Abstract

Understanding the evolutionary relationships and diversification of large, ecologically important plant families, such as Malvaceae s.l., is crucial for understanding angiosperm evolution and biogeographic patterns. Malvaceae s.l., known for its morphological diversity and complex evolutionary history, presents unique challenges in resolving phylogenetic relationships due to factors such as hybridization, introgression, polyploidy, and incomplete lineage sorting (ILS). This study addresses these complexities by reconstructing phylogenetic relationships, estimating divergence times, and inferring ancestral geographic distributions of Malvaceae s.l. using both plastid and nuclear genomic data. The analysis includes 134 species of Malvaceae s.l. and two outgroup species, strongly supports the division of Malvaceae s.l. into two primary clades, Byttneriina and Malvadendrina, while clarifying relationships among the subfamilies Dombeyoideae, Brownlowioideae, Sterculioideae, and Tilioideae. This study reconstructs the evolutionary history of Malvaceae s.l. based on plastid and nuclear genomic data, revealing deep phylogenetic discordance largely driven by incomplete lineage sorting, with additional signals of localized introgression within subfamilies. Divergence time estimates place the origin of Malvaceae s.l. at approximately 134.31 Ma (95% HPD = 123.16–138.33 Ma), representing its initial split from the outgroup lineage. The crown diversification of the family, corresponding to the divergence between its two major clades, Byttneriina and Malvadendrina, occurred around 119.38 Ma (95% HPD = 106.48–130.92 Ma). Ancestral range reconstructions support an African origin, followed by dispersal to tropical regions worldwide. Specifically, the ancestors of the Malvadendrina clade likely dispersed from Africa to South America, while Byttneriina shows strong ties to a North American origin.

Introduction

For large-scale phylogenetics, especially involving angiosperms, the in-depth exploration of species relationships and evolutionary history often encounters deep phylogenetic incongruence. With the widespread application of high-throughput sequencing technology and multi-gene datasets, substantial progress has been made in constructing large phylogenetic trees, revealing the complex mechanisms underlying species diversification[1−3]. However, in certain plant groups, particularly in families like Gesneriaceae[4] and Solanaceae[5], phylogenetic relationships still exhibit considerable uncertainty due to highly inconsistent gene histories and ambiguous evolutionary signals. These conflicts often arise from factors such as incomplete lineage sorting (ILS)[6], gene flow best represented as evolutionary networks[7,8], and polyploidy[4]. All these processes leave reticulate genetic footprints in different genomic datasets, making it particularly challenging to infer evolutionary relationships between species using molecular data.

The evolutionary relationships within Malvaceae s.l., with its two major clades, Byttneriina and Malvadendrina, and significant morphological variation, have been difficult to resolve[9−11]. Subfamilies within these clades vary greatly in morphology, and molecular data suggest a complex and often conflicting evolutionary history[11−13]. The Byttneriina clade is composed of two subfamilies: Byttnerioideae and Grewioideae. Byttnerioideae mainly consists of shrubs, characterized by petals that extend to the base, typically lacking an

epicalyx, and having a cup-shaped, curled margin (Fig. 1b). In contrast, Grewioideae includes trees and shrubs, featuring either separate or fused sepals and clawed petals (Fig. 1a). The Malvadendrina clade includes the remaining eight subfamilies, which display both morphological and molecular complexity and contentious relationships[10,11,14]. Helicteroideae mainly comprises trees and shrubs, characterized by fused sepals and clawed petals that are typically laterally constricted (Fig. 1c). The members of Sterculioideae comprise trees with petal-like sepals and prominently elongated stamens and pistils (Fig. 1d). Brownlowioideae is primarily composed of trees, with fused, bell-shaped sepals that irregularly split into two or three lobes (Fig. 1e). Dombeyoideae is characterized by a spiral leaf arrangement (Fig. 1f, g), persistent petals, and the presence of an epicalyx. Tilioideae exhibits a two-ranked leaf arrangement (Fig. 1h) and bears nectar glands on the abaxial side of its petals. The phylogenetic relationship of Dombeyoideae with the other subfamilies remains unresolved, with existing studies showing differing support for its sister-group relationship with Sterculioideae and Brownlowioideae[11,13], while some studies lean towards placing Dombeyoideae as a sister group to Tilioideae[15,16]. The subfamilies Bombacoideae and Malvoideae exhibit marked morphological differences; Bombacoideae primarily includes trees with palmate compound leaves, fused or absent sepals, and typically large seeds enclosed in capsules (Fig. 1i, j). In contrast, Malvoideae is predominantly composed of shrubs or herbaceous plants, characterized by five petals and typically forming

Fig. 1 Diversity of flowers and fruits of the Malvaceae Juss. (a) *Grewia biloba* var. parviflora (Bunge) Hand. -Mazz. Contributed by Maple, (b) *Waltheria indica* L., (c) *Helicteres angustifolia* L., (d) *Sterculia brevissima* H.H. Hsue ex Y. Tang, M.G. Gilbert & Dorr contributed by Malvaceae, (e) *Diplodiscus trichospermus* (Merr.) Y.Tang, M.G. Gilbert & Dorr, (f) *Melhania hamiltoniana* Wall., contributed by janstudio, (g) *Dombeya acutangula* Cav., (h) *Craigia yunnanensis* W.W.Sm. & W.E.Evans, (i) and (j) *Bombax ceiba* L., (k) *Urena lobata* L., (l) *Abutilon indicum* (L.) Sweet.

schizocarps or capsules (Fig. 1k, l). Molecular evidence strongly supports the sister-group relationship between Malvoideae and Bombacoideae[17] (Fig. 1, Supplementary Fig. S1).

Resolving these phylogenetic uncertainties is further complicated by the presence of conflicting signals across loci. Such discordance may reflect evolutionary processes including ILS, gene flow, and hybridization, which often obscure true species relationships[18]. Distinguishing whether observed discordance is due to ILS, introgression, or both remains challenging, particularly in lineages with complex coalescent histories. Dense taxon sampling and advanced analytical methods are therefore critical to disentangle these processes. Within Malvaceae s.l., the phylogenetic relationships among several subfamilies, including Tilioideae, Sterculioideae, Dombeyoideae, and Brownlowioideae, remain unresolved[13]. This provides an important opportunity to investigate how incomplete lineage sorting (ILS) and historical gene flow have influenced phylogenetic inference. In this study, the recently recognized subfamily Matisioideae is incorporated to re-examine the placement of all ten currently accepted subfamilies[19]. Additionally, divergence times and the reconstructed ancestral geographic ranges are estimated to explore the spatial and temporal patterns underlying the diversification of the family.

The specific objectives of this study are: (1) To analyze the phylogenetic relationships among the ten subfamilies of Malvaceae s.l. based on nuclear and plastid genomic data, using both concatenation and coalescence-based approaches. (2) To investigate the causes of phylogenetic discordance among the ten subfamilies of Malvaceae s.l. based on nuclear genomic data, by distinguishing the relative contributions of incomplete lineage sorting (ILS) and introgression using the quartet-based methods Phytop and QuIBL. (3) To reconstruct ancestral geographic distributions and examine biogeographic history and geographic expansion patterns of Malvaceae s.l. through analyses based on both nuclear and plastid phylogenies.

Materials and methods

Taxon sampling

This study focused on genus-level sampling within Malvaceae s.l. A total of 134 species were examined, representing 62 genera (ca. 25% of the 248 genera currently recognized in the family), along with two outgroup species. Specifically, the sampling includes 11 genera from Byttnerioideae (representing 50% of all recognized genera within this subfamily), six genera from Grewioideae (25%), five genera from Helicteroideae (41.67%), three genera from Brownlowioideae (27.27%), two genera from Tilioideae (66.67%), five genera from Dombeyoideae (31.25%), seven genera from Sterculioideae (53.85%), five genera from Bombacoideae (27.78%), one genus from Matisioideae (33.33%), and 17 genera from Malvoideae (13.49%). Additionally, two non-Malvaceae angiosperms were included as outgroups: *Anthoshorea assamica* P.S. Ashton & J. Heck (Dipterocarpaceae, Malvales) and *Mangifera indica* L. (Anacardiaceae, Sapindales). Samples included both newly collected individuals for *de novo* sequencing and species with publicly available genomic data retrieved from the NCBI database (Supplementary Table S1). Newly collected samples were obtained in compliance with local rules and regulations.

Genomic DNA was extracted from fresh leaf material using a modified cetyltrimethylammonium bromide (CTAB) protocol[20]. The DNA concentration was quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA), with samples exceeding a total amount of 0.8 μg selected for subsequent library construction. Whole-genome sequencing libraries were prepared using the VAHTS Universal DNA Library Prep Kit (Vazyme Biotech Co., Ltd., Nanjing, China), and sequencing was performed on the DNBSEQ-T7 platform by BGI Life Sciences Co., Ltd. (Hainan Province, China), generating approximately 10 Gb of raw data per sample. Quality control and filtering of the raw data were conducted using SOAPfilter v2.2[21]; reads containing more than 10% ambiguous

bases or low-quality bases (≤ 10) were removed to ensure high-quality data for downstream analyses.

Plastid genome generation and analysis

Plastid genomes were assembled from raw sequencing data using GetOrganelle v1.7.7.0[22] with k-mer sizes set to 21, 45, 65, 85, and 105 to optimize assembly across both conserved and variable regions. Assembly graphs were visualized and assessed with Bandage v0.9.02[23], and the parameters --max-kmer-coverage and --min-kmer-coverage were adjusted as needed to reduce coverage artifacts and low-depth noise. Assemblies failing to produce circular genomes were reprocessed by incrementally lowering --max-kmer-coverage values (e.g., 100, 50, 20) and re-evaluated for completeness and structure. Among the multiple FASTA outputs, the most complete and well-resolved circular genome—verified via Bandage visualization and Mauve v2.4.0[24] comparison against GetOrganelle's embplant_pt reference database—was selected for each sample. Genome annotation was performed using GeSeq[25], with gene features including protein-coding genes (CDS), rRNAs, and tRNAs transferred from reference plastomes based on Mauve alignments. Annotations were then exported in GenBank format and validated using the GB2Sequin tool (https://chlorobox.mpimp-golm.mpg.de/GenBank2Sequin.html). Detected annotation errors were manually corrected in Geneious Prime v22.1.1[26]. Final plastome statistics are summarized in Supplementary Tables S2 and S3.

For phylogenetic analysis, complete plastid genome sequences were aligned using MAFFT v7.505[27] implemented in PhyloSuite v1.2.3[28]. One copy of the inverted repeat (IR) region was removed prior to alignment to avoid redundancy in branch length estimation. The final alignment was manually partitioned in Geneious Prime according to structural regions (LSC, SSC, IR) and functional categories (coding vs non-coding). Maximum likelihood (ML) trees were inferred using IQ-TREE v2.2.6[29], with best-fit models selected using ModelFinder[30] under the Bayesian Information Criterion (BIC). Branch support was assessed with 1,000 ultrafast bootstrap replicates[31].

Nuclear dataset assembly and analysis

To reconstruct the nuclear phylogeny of Malvaceae s.l., a dataset is assembled using Easy353 v2.0.1[32] to extract 353 conserved nuclear loci of angiosperms from shallow whole-genome sequencing data[33,34]. A two-step procedure was employed: first, representative sequences were extracted from selected taxa of different subfamilies to serve as reference sequences; then, target gene assembly was guided via read mapping. This process ultimately yielded approximately 350 nuclear loci for downstream analyses.

Sequences for each locus were aligned using MAFFT v7.520, and individual gene trees were inferred in IQ-TREE v2.2.6 under their corresponding best-fit substitution models as determined by ModelFinder using the Bayesian Information Criterion. A species tree was reconstructed from the set of 353 nuclear gene trees using ASTRAL III v5.7.1[35]. Branch lengths in the ASTRAL species tree are expressed in coalescent units rather than substitutions per site or absolute time and were therefore not interpreted as estimates of divergence time or genetic distance. Given the potential for incomplete lineage sorting and introgression, the ASTRAL species tree was adopted as the principal framework for downstream evolutionary and biogeographic analyses. For comparison, a concatenated nuclear supermatrix was assembled in PhyloSuite v1.2.3, with partitions defined by individual loci. Maximum likelihood phylogenetic inference was conducted in IQ-TREE v2.2.6 under a partitioned scheme. with the optimal partitioning strategy and the best fitting nucleotide substitution model for each partition selected by ModelFinder according to the Bayesian Information Criterion, and branch support evaluated with 1,000 ultrafast bootstrap replicates. Due to its susceptibility to gene-tree discordance, the concatenated topology was used only as a supplementary reference.

Phytop analysis

To quantify the sources of nuclear phylogenetic discordance and to distinguish patterns attributable to incomplete lineage sorting (ILS) from those compatible with introgression or hybridization (IH), Phytop v0.3.2 is used[36]. Phytop is a quartet-based method that takes as input an ASTRAL species tree with its associated quartet support and, for each internal branch, partitions the supporting gene tree quartets into the three possible unrooted topologies (q1, q2, and q3) and uses their relative frequencies to compute branch-specific summary indices of ILS and IH. In this study, Phytop is applied to the ASTRAL species tree inferred from the 353 nuclear loci, extracted and retained the per-branch ILS and IH indices for subsequent interpretation of nuclear gene tree discordance patterns, and examined the default graphical output, including pie charts that depict the proportions of q1, q2, and q3 at each internal branch.

QuIBL analysis

To further investigate whether the observed phylogenetic discordance among species was due solely to incomplete lineage sorting (ILS) or also involved introgression, the QuIBL (Quantifying Introgression via Branch Lengths) method[37] is employed. QuIBL fits two models to the branch length distributions: one assuming ILS alone and another incorporating both ILS and introgression. Model comparisons are conducted using the Bayesian Information Criterion (BIC), with a ΔBIC > 10 favoring the ILS-only model and a ΔBIC < −10 supporting the ILS plus introgression model[38]. For analysis, gene trees constructed from non-overlapping 2-kb genomic windows spaced every 20 kb across the genome were utilized, a strategy designed to minimize the impact of intralocus recombination[37]. Only gene trees containing at least five parsimony-informative sites were retained for analysis. All gene trees were rooted using the designated outgroup, *Anthoshorea assamica*. QuIBL was executed with default parameters, and the resulting ΔBIC values were used to infer the predominant evolutionary processes contributing to phylogenetic discordance among species[39].

Divergence time estimation

To obtain a subset of loci suitable for divergence time estimation while keeping the analysis computationally tractable, SortaDate pipeline[40] is used to rank nuclear loci according to their degree of clock-likeness, overall tree length, and topological congruence with the ASTRAL species tree. The 25th highest-ranking loci were then selected, which combined near clock-like rate constancy with moderate evolutionary rates and low levels of gene-tree conflict. This selection strategy reduced both the computational burden and the impact of incomplete lineage sorting and other sources of model misspecification in relaxed-clock analyses.

Divergence times were estimated in BEAST v2.7.7[41] under a Birth-Death tree prior and an Optimized Relaxed Clock (ORC) model of branch-rate variation. A total of nine temporal calibrations (one secondary calibration and eight fossil constraints) were applied. A secondary calibration on the crown node of Malvales was implemented as a uniform prior between 110.48 and 138.33 Ma, following

Ramirez-Barahona et al.[42]. Eight fossil calibrations were assigned to the crown nodes of the major Malvaceae subfamilies. For all fossils, the lower bound of the uniform prior corresponded to the youngest limit of the stratigraphic interval reported for each fossil, whereas the upper bound was fixed at 138.33 Ma. The fossil constraints followed Hernandez-Gutierrez et al.[13] and comprised *Bombacoxylon langstoni* (Malvaceae, 72.1 Ma), *Bombax-type pollen* (Bombacoideae, 66 Ma), *Discoidites borneensis* (Brownlowioideae, 56 Ma), *Sphinxia ovalis* (Dombeyoideae, 47.8 Ma), *Grewioxylon indicum* (Grewioideae, 33.9 Ma), *Malvaciphyllum macondicus* (Malvoideae, 56 Ma), *Sterculiaephyllum australis* (Sterculioideae, 66 Ma), and *Craigia oregonensis* (Tilioideae, 47.8 Ma).

Nine independent MCMC chains were run for 100 million generations, sampling every 1,000 generations. Tracer v.1.7[43] was used to check for effective sample sizes (ESS > 200) with the first 25% discarded as burn-in. All the runs were combined using LogCombiner v2.7.7 after discarding the first 25% of trees of each as burn-in. TreeAnnotator v2.7.7 was used to generate the maximum clade credibility tree, displaying mean divergence time estimates with 95% highest posterior density (HPD) intervals. FigTree v.1.4.0 was used for tree visualization.

Ancestral area reconstructions

Occurrence data were downloaded from the GBIF database (DOI: 10.15468/dl.frwh2g). Raw data were cleaned using the rgbif package v3.5.2[44] in R 4.4.0 to remove erroneous records and ensure consistency in species identification and geographic coordinates. The cleaning process involved filtering out records with missing or obviously incorrect coordinates. After data cleaning, the R package BioGeoBEARS v1.1.3 is used to estimate ancestral geographic ranges under a maximum likelihood framework[45]. The time-calibrated divergence tree generated from BEAST v.2.7.7 was used as input for this analysis. BioGeoBEARS was chosen for its capability to compare various biogeographic models and to incorporate complex processes, such as dispersal, extinction, and founder-event speciation, into the analysis. Six different models—DEC, DEC + J, DIVALIKE, DIVALIKE + J, BAYAREALIKE, and BAYAREALIKE + J were applied—to reconstruct the ancestral distribution areas of Malvaceae s.l. To better reflect the biogeographic history of Malvaceae s.l., the global distribution is divided into six main regions: (A) Asia, (B) Africa, (C) North America, (D) South America, (E) Oceania, and (F) Europe. The regional division was based on the main endemic distribution areas and ecological–geographic characteristics of Malvaceae s.l.[46]. Because the taxon sampling includes only approximately one quarter of the genera of Malvaceae s.l., these ancestral range reconstructions should be regarded as hypotheses for the sampled lineages rather than as definitive family-wide scenarios. RASP v4.2 was used to reconstruct the ancestral distribution areas of Malvaceae s.l. using Statistical Dispersal–Vicariance Analysis (S-DIVA) and Bayesian Binary MCMC (BBM) analysis[47], providing additional insights into the historical biogeography of the group. The combination of BioGeoBEARS and RASP allowed us to cross-validate the results and ensure robustness by comparing different methodologies and models in reconstructing the ancestral areas.

Results

Phylogenetic analysis based on nuclear data

Both nuclear reconstructions, the ASTRAL species tree (Fig. 2) and the concatenated ML tree (Fig. 3), yielded a broadly congruent deep structure for Malvaceae s.l., with a primary split between Byttneriina and Malvadendrina and with most subfamilies supported as monophyletic. Within Malvadendrina, the backbone topology was largely congruent between the two analyses. Sterculioideae and Tilioideae formed a sister pair (ASTRAL: PP = 0.40; ML: BS = 100), and only minor differences were observed in the order of divergence among *Pterygota*, *Cola*, and *Heritiera* within Sterculioideae. Brownlowioideae was recovered as sister to the Sterculioideae + Tilioideae clade (PP = 0.50; BS = 100), and this inclusive lineage was, in turn, sister to Dombeyoideae (PP = 0.98; BS = 100). Bombacoideae and Matisioideae were inferred as sister groups (PP = 0.43; BS = 100), and this clade was sister to Malvoideae (PP = 1.00; BS = 100). The ASTRAL and ML trees also agreed in recovering Helicteroideae (sensu lato) as non-monophyletic, splitting the sampled taxa traditionally assigned to Helicteroideae into two well-supported, phylogenetically distinct lineages, one including *Durio* and the other comprising the remaining genera. These lineages have at times been treated as separate families or segregate subfamilies in previous classifications, and results support the use of a broad Malvaceae s.l. concept while explicitly acknowledging the polyphyly of Helicteroideae s.l. The main topological incongruence within Malvadendrina concerns the placement of *Durio*: in the ASTRAL species tree, it is resolved near the base of the clade comprising taxa traditionally placed in Helicteroideae, whereas in the concatenated ML tree, it is positioned as the sister lineage to Malvaceae s.s. By contrast, topological conflict is more pronounced within Byttneriina, where the relationships between Grewioideae and Byttnerioideae differ between the two analyses, indicating substantial gene tree discordance in this part of the tree.
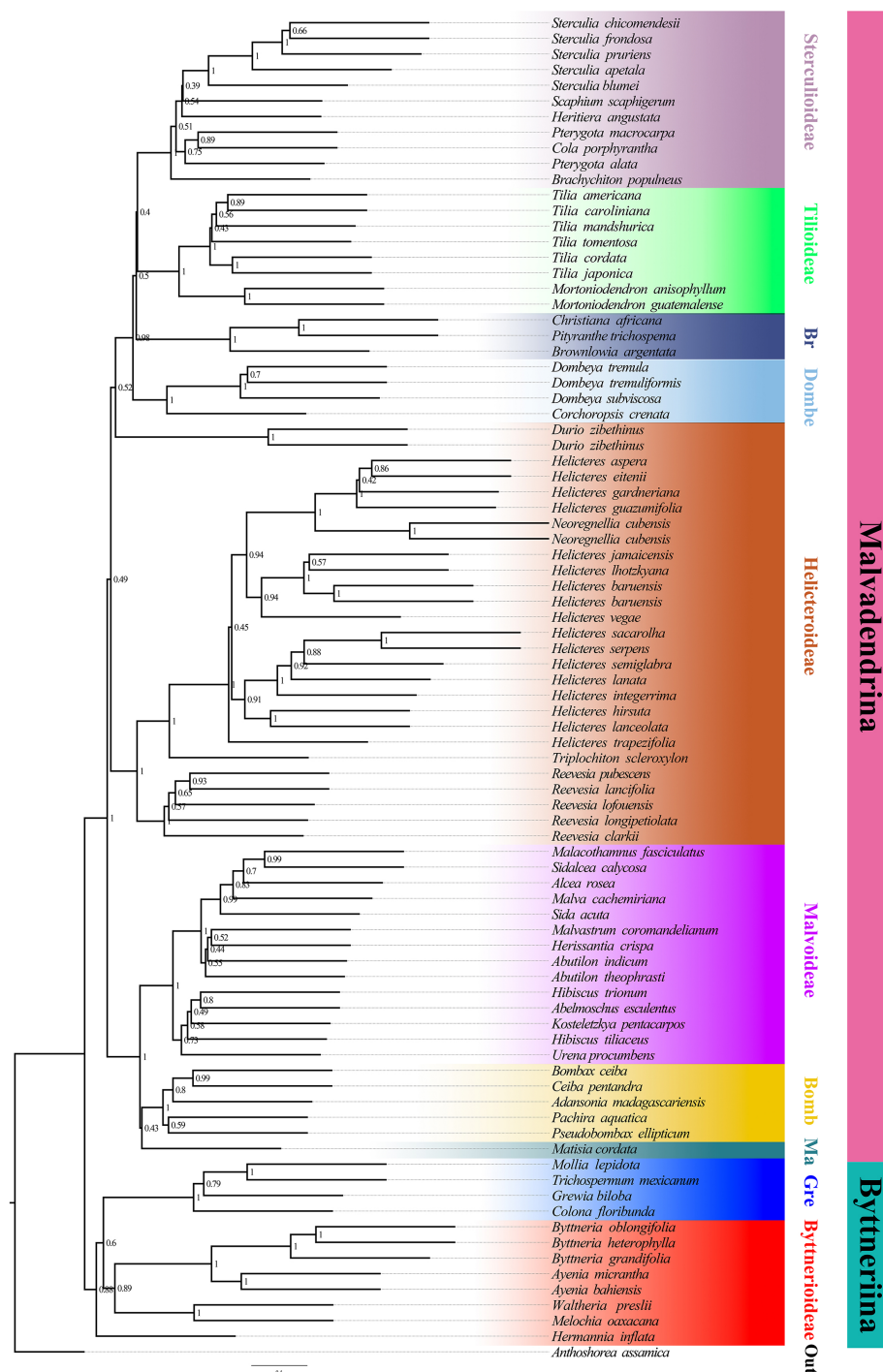
Phylogenetic analysis based on plastid data

The phylogenetic tree inferred from whole plastid genomes also recovered the monophyly of the ten subfamilies of Malvaceae s.l. (Fig. 4). However, this topology exhibited incongruence with the nuclear-based trees. For instance, Dombeyoideae was placed as sister to Tilioideae in the plastid tree, whereas in the nuclear trees it was recovered as sister to a clade comprising Tilioideae, Brownlowioideae, and Sterculioideae. Given the history of hybridization and polyploidy in Malvaceae s.l., it is expected that plastid data may not accurately reflect the underlying species relationships.

Phytop and QuIBL analysis

Phytop results indicated that, for most subfamily level branches within Malvadendrina, the species tree topology q1 accounted for the majority of supporting quartets at each node, typically with q1 clearly exceeding q2 and q3, and ILS indices ranging from approximately 30% to 60%, while IH indices were zero for most branches (Fig. 5). Within Byttneriina, several key nodes showed reduced q1 proportions and correspondingly higher contributions of q2 and q3, indicating stronger gene tree discordance than along the Malvadendrina backbone. Most of these nodes also had IH indices equal to zero. This pattern is consistent with a scenario dominated by ILS under rapid lineage diversification, and a strong node-level signals of introgression was not detected.

To further explore phylogenetic discordance across lineages, QuIBL analysis was employed. The resulting introgression matrix revealed that most high-probability introgression events occurred within subfamilies, whereas cross-subfamily comparisons consistently exhibited lower introgression proportions, typically in the range of 0.1–0.3 (Fig. 6a). This pattern implies that gene flow, if present, is likely restricted to closely related lineages. To evaluate
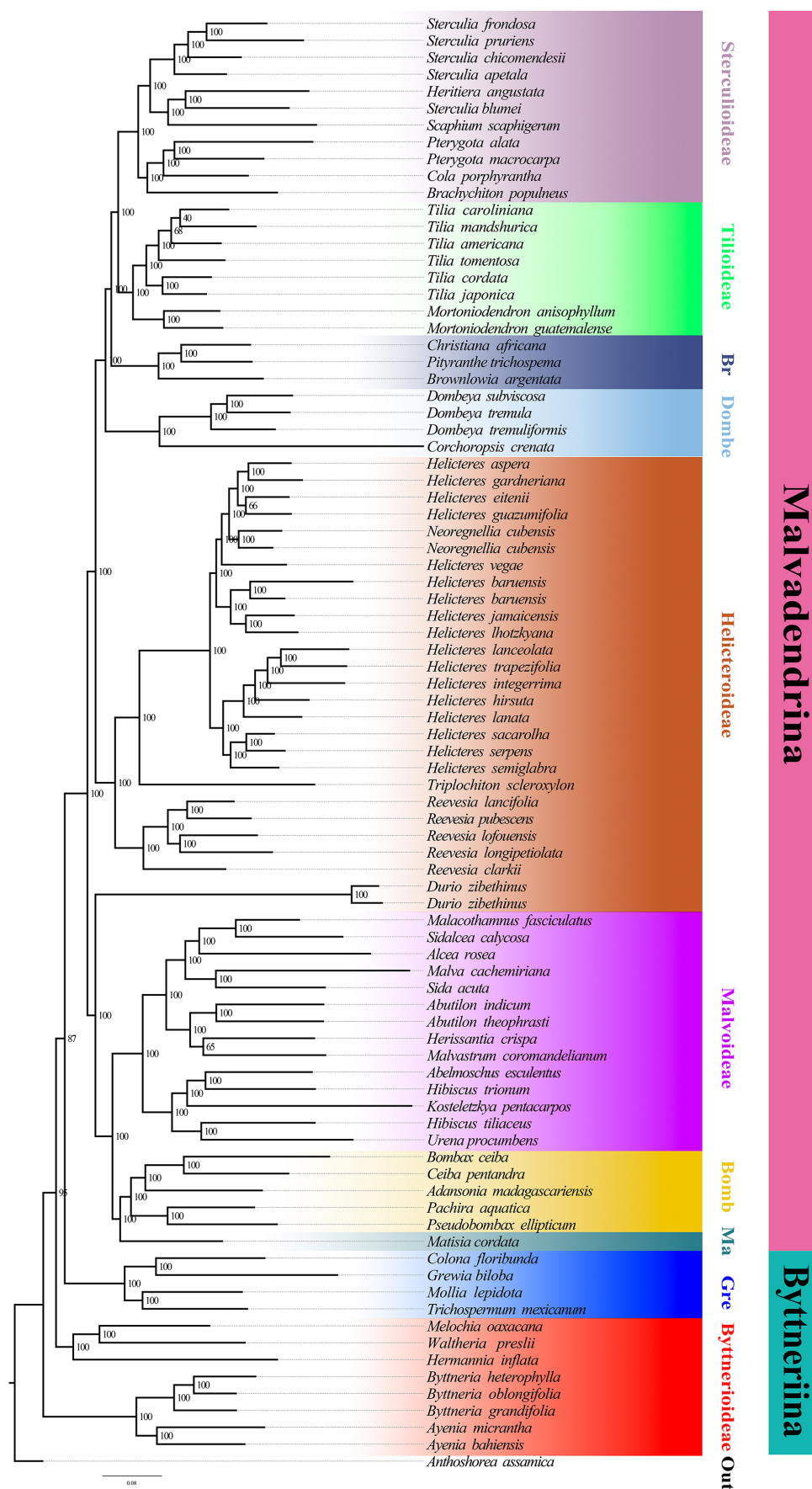
Fig. 2 Species tree inferred from the nuclear dataset using ASTRAL-III. Branch support is given by local posterior probabilities (PP). Major clades corresponding to currently recognized subfamilies are indicated by branch colours as visual references of relationships.

the model support across gene trees, the distribution of loci favoring the ILS + introgression model was compared between discordant and true topologies (Fig. 6c, d). The discordant topologies showed a broad and continuous band of high introgression support, particularly concentrated at low Non-ILS C values (x < 10), with y-axis values spanning from ~0.125 to nearly 1.0. In contrast, true topologies exhibited a more fragmented distribution, with two distinct regions: a small cluster of high-support points (y ≈ 0.75–1.0) and a broader set around y ≈ 0.3 –0.5. These patterns suggest that introgression contributes disproportionately to topological discordance, while loci supporting the species tree are more consistent

with ILS or localized introgression. Finally, the distribution of loci classified under the ILS vs ILS + introgression models (Fig. 6b) confirmed that ILS dominates the evolutionary signal across the dataset. However, a subset of loci showed strong preference for the introgression model, underscoring the potential for lineage-specific gene flow events, even in the absence of broader subfamily-level signals.
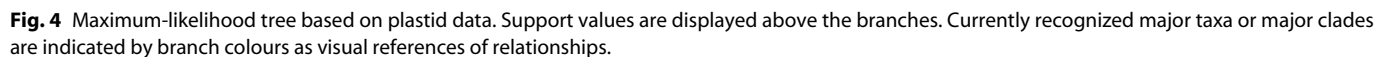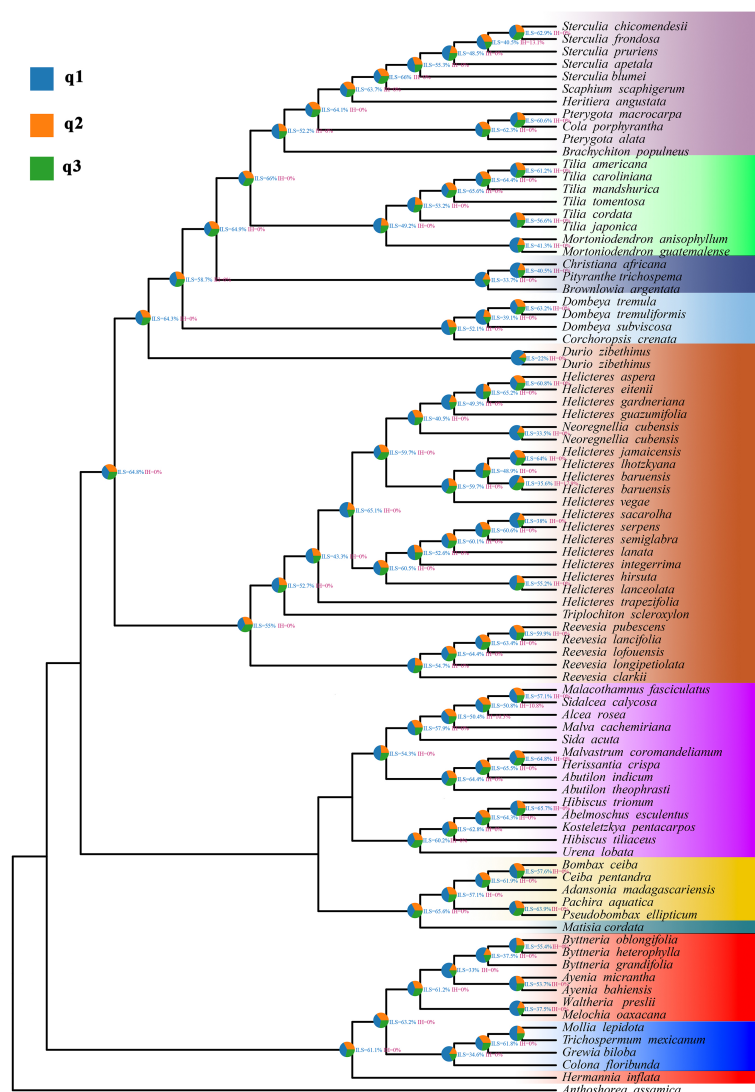
Divergence time estimation

Based on the nuclear gene dataset, the crown age of Malvaceae s.l. was estimated at 119.38 Ma (95% HPD = 106.48–130.92 Ma),

Evolutionary framework of Malvaceae s.l.

Fig. 3 Maximum-likelihood tree inferred from the concatenated nuclear dataset. Branch support is given by ultrafast bootstrap values (BS). Major clades or evolutionary lineages within the currently recognized subfamilies are indicated by branch colours as visual references of relationships.

Fig. 4 Maximum-likelihood tree based on plastid data. Support values are displayed above the branches. Currently recognized major taxa or major clades are indicated by branch colours as visual references of relationships.
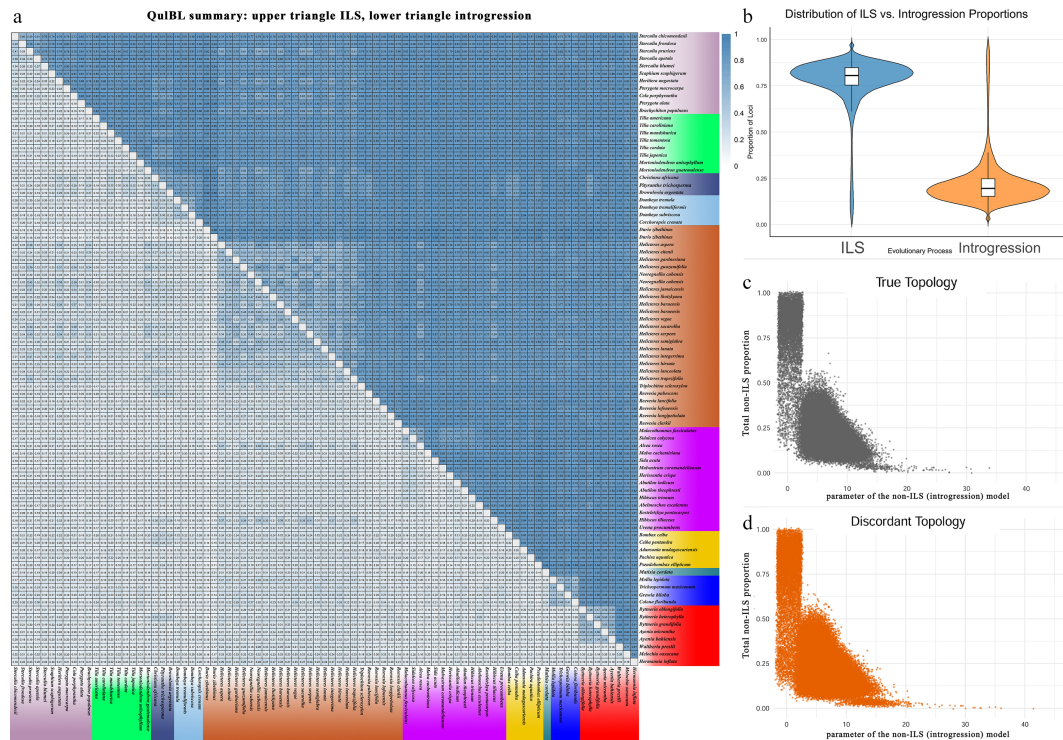
Fig. 5 Phytop analysis of quartet patterns on the ASTRAL species tree of Malvaceae s.l. Each internal node is annotated with a pie chart showing the relative frequencies of the three alternative quartet topologies (q1, q2, q3) around that branch, with blue indicating the topology concordant with the ASTRAL species tree and orange and green indicating the two discordant topologies. Numerical labels next to nodes give the ILS and IH indices inferred by Phytop.

corresponding to the divergence between Byttneriina and Malvadendrina (Fig. 7). Within Malvaceae s.l., the stem lineage leading to Byttneriina dates back to 128.30 Ma, whereas the crown age of Byttneriina, marking the first diversification among the sampled lineages, was inferred at 96.47 Ma. The crown age of Malvadendrina was estimated at 110.66 Ma (95% HPD = 96.86–124.34 Ma). At the subfamily level, Bombacoideae began diversifying around 80.67 Ma (95% HPD = 66.01–93.94 Ma), Dombeyoideae around 54.68 Ma (95% HPD = 47.80–67.90 Ma), Tilioideae at approximately 72.79 Ma (95% HPD = 53.31–89.12 Ma), and Brownlowioideae at 61.62 Ma (95% HPD = 56.00–73.98 Ma). The crown age of the main Helicteroideae clade excluding *Durio* was estimated at 84.11 Ma (95% HPD = 64.07–04.03 Ma). Although *Durio* is traditionally classified within Helicteroideae, in the nuclear divergence time tree, it branches off earlier along the Malvadendrina backbone, with a stem age of 110.66 Ma. The crown age of the sampled *Durio* lineage is much younger, at 13.52 Ma (95% HPD = 5.30–25.45 Ma), indicating a long stem branch associated with relatively recent diversification within the genus.

Ancestral area reconstruction

Ancestral area reconstructions were performed using RASP and BioGeoBEARS (Fig. 8). The DEC + J and DIVA-like + J models in BioGeoBEARS produced very similar results, with a ΔAIC difference of just 0.8 (Supplementary Table S4). Both models consistently supported that Malvaceae s.l. most likely originated in Africa (B). Under the DEC + J model, the root ancestral area of Malvaceae s.l. was reconstructed as a combination of Africa and South America (B + D), whereas the DIVA-like + J model inferred a single origin in Africa (B). Apart from this difference at the root, the Byttneriina clade was consistently reconstructed with a South American (D) origin in both models. Similarly, the Malvadendrina clade was predominantly reconstructed with an African (B) origin. Taken together, these results indicate that, although the choice of model slightly affects the precise reconstruction of the ancestral area of Malvaceae s.l., the broader biogeographical pattern is robust, with Africa and South America emerging as key regions in the diversification of the family.

Fig. 6 Tests for introgression. (a) Heatmap summarizing QuIBL results for pairwise species comparisons. For each species pair, the upper triangle shows the mean proportion of loci attributed to ILS, and the lower triangle shows the mean proportion attributed to introgression (mixprop2). Colours indicate proportions from zero to one, and empty cells mark pairs for which no informative triplets were available. (b) Distribution of the proportion of loci that exhibit a history of ILS or introgression across all discordant topologies, respectively. (c) Relationship between internal branch length (in coalescent units) and the proportion of non-ILS loci for triplets matching the true topology. (d) Same as (c), but for triplets with discordant topologies.

Discussion

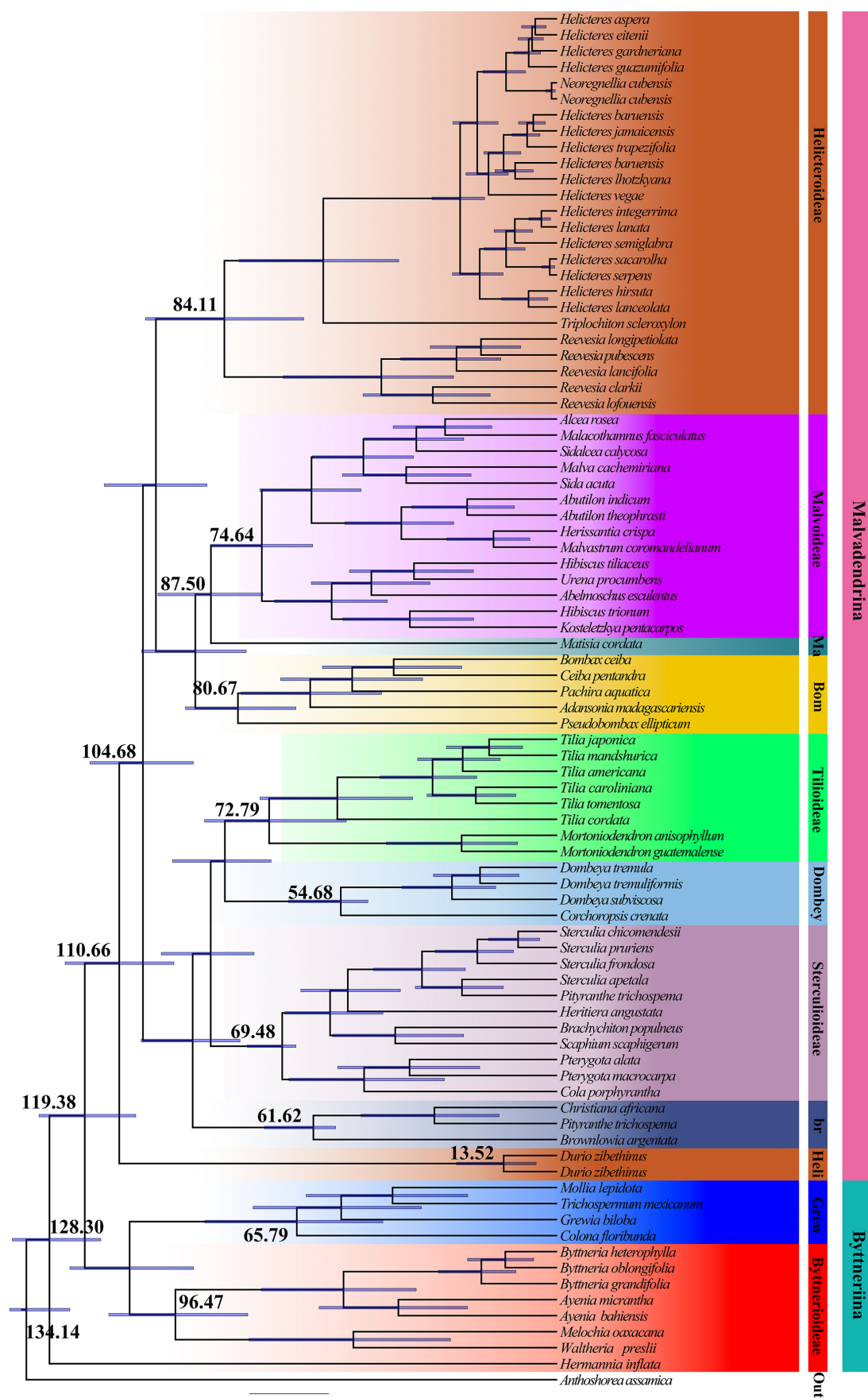## Discordance among nuclear gene trees and with the plastid genome

Phylogenetic incongruence arises from various complex factors, including genetic drift, selective pressures, gene flow, hybridization events, and variation in evolutionary rates. Particularly in Malvaceae s.l., hybridization has been identified as a key factor contributing to phylogenetic inconsistency, although other evolutionary mechanisms also play significant roles[13]. For instance, genetic drift in small populations can lead to random fluctuations in allele frequencies, resulting in divergences in the evolutionary trajectories of the nuclear and plastid genomes[48,49]. Moreover, differing selective pressures may cause nuclear and chloroplast genomes to respond independently to environmental conditions, thereby intensifying phylogenetic discordance between them[50]. In flowering plants, the nuclear genome evolves at a higher rate than the plastid genome, which tends to be more conserved. This disparity can lead to distinct evolutionary trajectories between the two genomes, thereby amplifying phylogenetic discordance[51,52]. Resolving these discrepancies is essential for accurately reconstructing the evolutionary history of complex plant families, such as Malvaceae s.l.

Many previous studies of Malvaceae s.l. relied on plastid genomic data[10,53]. However, as Hernandez-Gutierrez et al.[13] pointed out, nuclear genomic data offer unique advantages in addressing complex issues and provide a more comprehensive view of the evolutionary relationships within Malvaceae s.l. While plastid data are widely used, findings further highlight the necessity of prioritizing nuclear genomic data in phylogenetic analyses. The relatively rapid evolutionary rate of the nuclear genome further exacerbates the
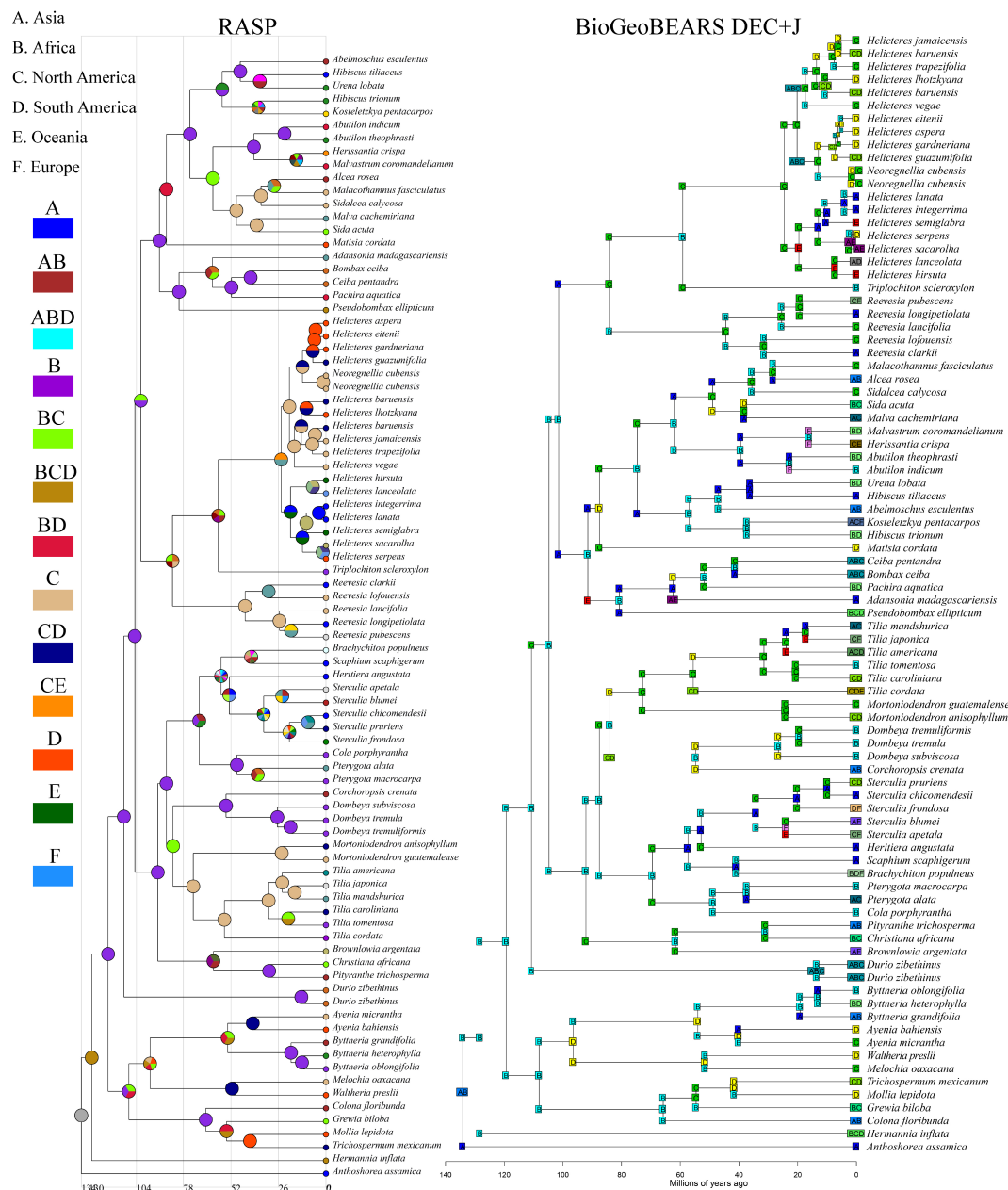
phylogenetic incongruence between nuclear and plastid data[52]. In Malvaceae s.l., nuclear and plastid phylogenetic trees exhibit marked topological discordance across multiple subfamilies[13,53]. Except for the sister group relationship between Grewioideae and Byttnerioideae, which is consistently recovered in both nuclear and plastid trees[13,19], the other eight subfamilies display distinct topologies in the two types of phylogenetic trees. This points to a broad and systematic discordance between nuclear and plastid signals. A major part of this discordance likely reflects the different modes of inheritance and evolutionary dynamics of the two genomes. Localized plastid incongruence is also observed for a small number of taxa, but these cases do not affect the higher-level relationships inferred in this study. The plastid genome is usually maternally inherited, has a much smaller effective population size than the nuclear genome, and generally evolves more conservatively, with lower substitution rates and fewer phylogenetically informative sites at deep nodes[54,55]. It was long regarded as essentially non-recombining, but recent studies have shown that plastids can, in fact, undergo recombination[56]. In particular, plastid capture is a frequent phenomenon in taxa with a history of hybridization, which can bias phylogenetic inferences by misrepresenting underlying species relationships[53,57]. In contrast, the nuclear genome is biparentally inherited and undergoes recombination, enabling it to integrate genetic information from both parental lineages and to reflect nuclear-specific evolutionary processes such as whole-genome duplication (WGD)[58]. While plastid data remain informative at certain phylogenetic levels, nuclear gene trees are generally considered more robust for inferring species-level relationships in lineages with complex evolutionary histories[11,13,59].

Even when based on the same multilocus nuclear dataset, different phylogenetic methods can in principle yield incongruent

Fig. 7 Divergence time analysis based on nuclear data using BEAST. Each major branch provides mean divergence times and 95% highest posterior density (HPD) intervals. A complete visualization of the 95% HPD intervals across all nodes is provided in Supplementary Fig. S1.

Fig. 8 The figure on the left is an estimate of the ancestral range of Malvaceae s.l. taxa in the RASP using the S-DIVA model. The figure on the right uses the BioGeoBEARS script to select the optimal output model based on six models to estimate the ancestral range of the Malvaceae s.l. groups, which are geographically distributed as (A) Asia, (B) Africa, (C) North America, (D) South America, (E) Oceania, and (F) Europe.

topologies, particularly in lineages that have experienced rapid diversification or gene exchange[13,18]. In this study, the species tree inferred under the multispecies coalescent in ASTRAL and the concatenated ML tree were largely congruent, recovering the same deep split between Byttneriina and Malvadendrina and identical placements for most subfamilies. Residual discrepancies were confined to a small number of branches, most notably the alternative positions of *Durio* within Malvadendrina and the internal relationships among Grewioideae and Byttnerioideae in Byttneriina. These remaining conflicts are best interpreted as the outcome of discordant signals among individual nuclear loci generated by pervasive incomplete lineage sorting and localized introgression, as suggested by the Phytop and QuIBL analyses. In this context, the multispecies coalescent framework implemented in ASTRAL, which explicitly accommodates gene tree heterogeneity due to ILS,

provides a more appropriate summary of the nuclear phylogenomic signal than concatenated ML approaches that assume a single underlying history for all loci[18,60].

Phylogenetic divergence and speciation complexity in Malvaceae s.l.

Resolving phylogenetic relationships among subfamilies within Malvaceae s.l. has long been hampered by pervasive incongruence among nuclear gene trees, particularly in groups such as Sterculioideae, Tilioideae, and Brownlowioideae. This phenomenon epitomizes a central challenge in phylogenomics, namely, how to distinguish discordance generated by incomplete lineage sorting from discordance caused by introgression resulting from historical gene flow and hybridization[57]. The results show a pattern in which a comparatively stable deep backbone contrasts with much more

complex relationships among several shallow lineages. The nuclear genomic species tree recovers the primary split between Byttneriina and Malvadendrina and supports the monophyly of most subfamilies, a result broadly consistent with recent classification frameworks[13,61]. Compared with earlier studies, the inclusion of large-scale nuclear genomic data combined with explicit conflict analyses enables a more precise localization of discordance, particularly among intergeneric relationships within Byttneriina and in the placement of recalcitrant lineages in Malvadendrina such as *Durio*.

The results of Phytop and QuIBL provide further quantitative evidence for the relative roles of incomplete lineage sorting and introgression. PhyTop indicates that, for most internal nodes within Malvadendrina, the species tree topology q1 clearly predominates among quartet frequencies, the ILS index is at intermediate levels, and the IH index is close to zero. QuIBL model comparisons likewise show that, across most species combinations, the ILS component generally concentrates around 70%–80%, whereas the introgression component rarely exceeds 30% and accounts for only a minority of loci (Fig. 5b). Together with the pairwise heatmap results, these analyses indicate that higher proportions of introgression are mainly restricted to closely related species within individual subfamilies (Fig. 5a). Within the assumptions of the models, the combined evidence thus points to a consistent conclusion: incomplete lineage sorting is the predominant source of nuclear gene tree discordance in Malvaceae s.l., whereas introgression exerts a much more limited, lineage specific, and spatially localized influence that does not alter the deep phylogenetic structure of the family. This pattern mirrors the findings of previous genomic studies, which likewise concluded that ILS, together with lineage-specific episodes of introgression, primarily drives local topological instability while exerting only modest effects on the deeper relationships of the family[56]. Recognizing the heterogeneity in the intensity and timing of ILS and introgression across lineages and evolutionary scales will be important for future attempts to integrate genomic, morphological, and ecological data when refining the classification of Malvaceae s.l., reconstructing trait evolution, and elucidating its biogeographic history.

Divergence time and ancestral distribution reconstruction

The analysis suggests that the stem age of Malvaceae s.l. dates to the Early Cretaceous, at approximately 134.14 Ma (95% HPD = 123.16−138.33 Ma), with the lineage likely originating in Africa. This timing coincides with significant geological events, such as the ongoing fragmentation of Gondwana. This geological reconfiguration created new environmental niches and isolated populations, facilitating the early diversification of Malvaceae s.l.[62]. A major phase of lineage diversification is inferred to have occurred between 119.38 and 54.68 Ma, likely promoted by continued continental drift and increasing regional biogeographic isolation[63]. During this time, the divergence of major clades such as Byttneriina and Malvadendrina highlights the role of geographic isolation in shaping evolutionary trajectories. The breakup of Gondwana likely facilitated this isolation by separating populations and leading to distinct evolutionary pathways. For example, the continued separation of Africa and South America during the Late Cretaceous provided opportunities for diversification driven by geographic and ecological factors[64].

The analyses also suggest that the early diversification of the sampled Malvaceae s.l. lineages most likely occurred in the tropical regions of northern Africa. The diversification of Malvaceae s.l. is consistent with the notion that early plant lineages diversified in tropical environments before dispersing to other regions via vicariance and long-distance dispersal[65]. Several evolutionary

mechanisms may have contributed to the geographic spread and diversification of Malvaceae s.l. The breakup of Gondwana during the Late Jurassic and Cretaceous probably created both migration routes and geographic barriers among tropical landmasses, which may in turn have facilitated early range expansion and spatial structuring of the family[66]. Hybridization and introgression can also facilitate ecological adaptation and contribute to rapid diversification in many plant lineages[67]. These evolutionary processes, combined with geographic isolation, likely played a critical role in the rapid diversification observed during the Late Cretaceous[68]. The complex biogeography of Malvaceae s.l. can also be partly attributed to long-distance dispersal, which allowed certain lineages to reach geographically distant regions. These divergence time estimates and biogeographic inferences provide an initial framework for understanding the early history of Malvaceae s.l., although further improvements in fossil sampling, taxon representation, and model specification will be essential for refining the temporal and spatial context of its evolution.

Conclusions

By integrating plastid and nuclear genomic data, this study provides a comprehensive framework for understanding the evolutionary history of Malvaceae s.l. Two major clades corresponding to Byttneriina and Malvadendrina were recovered, and the monophyly of most subfamilies was confirmed. Divergence time estimates and ancestral area reconstructions suggest an origin spanning Africa and South America. Furthermore, investigations into gene tree discordance identify incomplete lineage sorting (ILS) as the predominant driver of phylogenetic incongruence, whereas introgression appears to be restricted to localized events among closely related species. The findings from this study highlight the importance of integrating multiple genomic partitions and analytical approaches to resolve relationships within taxonomically challenging plant groups.

Author contributions

The authors confirm their contributions to the paper as follows: study conception and design: Wang HF, Li HY; experimental work and data analysis: Yang WL, Liu PF; draft manuscript preparation and revision: Yang WL, Liu PF, Zuo SY, Jiang DZ, Zhang YH, Chen B, Li HY, Wang HF. All authors reviewed the results and approved the final version of the manuscript.

Data availability

The datasets generated during and/or analyzed in the currentstudy are available from the corresponding author on reasonable request.

Acknowledgments

the Laboratories of Analytical Biology at the National Museum of Natural History, Smithsonian Institution, for technical support and assistance. We also thank the High-Performance Computing Platform of YZBSTCACC for providing computational resources that supported data analyses.

Conflict of 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.

References

[1] Soltis DE, Albert VA, Savolainen V, Hilu K, Qiu YL, et al. 2004. Genome-scale data, angiosperm relationships, and 'ending incongruence': a cautionary tale in phylogenetics. *Trends in Plant Science* 9:477–483

[2] Koenen EJM, Ojeda DI, Steeves R, Migliore J, Bakker FT, et al. 2020. Large-scale genomic sequence data resolve the deepest divergences in the legume phylogeny and support a near-simultaneous evolutionary origin of all six subfamilies. *New Phytologist* 225:1355–1369

[3] Zhang G, Ma H. 2024. Nuclear phylogenomics of angiosperms and insights into their relationships and evolution. *Journal of Integrative Plant Biology* 66:546–578

[4] Yang LH, Shi XZ, Wen F, Kang M. 2023. Phylogenomics reveals widespread hybridization and polyploidization in *Henckelia* (Gesneriaceae). *Annals of Botany* 131:953–966

[5] Deanna R, Barboza GE, Bohs L, Dodsworth S, Gagnon E, et al. 2025. A new phylogeny and phylogenetic classification for Solanaceae. *bioRxiv* Preprint

[6] Knowles LL, Huang H, Sukumaran J, Smith SA. 2018. A matter of phylogenetic scale: distinguishing incomplete lineage sorting from lateral gene transfer as the cause of gene tree discord in recent versus deep diversification histories. *American Journal of Botany* 105:376–384

[7] Solís-Lemus C, Yang M, Ané C. 2016. Inconsistency of species tree methods under gene flow. *Systematic Biology* 65:843–851

[8] Maurin O, Anest A, Bellot S, Biffin E, Brewer G, et al. 2021. A nuclear phylogenomic study of the angiosperm order Myrtales, exploring the potential and limitations of the universal Angiosperms353 probe set. *American Journal of Botany* 108:1087–1111

[9] Nyffeler R. 2005. Phylogenetic analysis of the Malvadendrina clade (Malvaceae s.l.) based on plastid DNA sequences. *Organisms Diversity & Evolution* 5:109–123

[10] Wang JH, Moore MJ, Wang H, Zhu ZX, Wang HF. 2021. Plastome evolution and phylogenetic relationships among Malvaceae subfamilies. *Gene* 765:145103

[11] Cvetković T, Areces-Berazain F, Hinsinger DD, Thomas DC, Wieringa JJ, et al. 2021. Phylogenomics resolves deep subfamilial relationships in Malvaceae *s.l.* *G3 Genes|Genomes|Genetics* 11:jkab136

[12] Escobar García P, Schönswetter P, Fuertes Aguilar J, Nieto Feliner G, Schneeweiss GM. 2009. Five molecular markers reveal extensive morphological homoplasy and reticulate evolution in the *Malva* alliance (Malvaceae). *Molecular Phylogenetics and Evolution* 50:226–239

[13] Hernández-Gutiérrez R, van den Berg C, Granados Mendoza C, Peñafiel Cevallos M, Freire ME, et al. 2022. Localized phylogenetic discordance among nuclear loci due to incomplete lineage sorting and

[14] introgression in the family of cotton and cacao (Malvaceae). *Frontiers in Plant Science* 13:850521

[14] Baum DA, Dewitt Smith SD, Yen A, Alverson WS, Nyffeler R, et al. 2004. Phylogenetic relationships of Malvatheca (Bombacoideae and Malvoideae; Malvaceae sensu lato) as inferred from plastid DNA sequences. *American Journal of Botany* 91:1863–1871

[15] Le Péchon T, Gigord LDB. 2014. On the relevance of molecular tools for taxonomic revision in Malvales, Malvaceae s.l., and Dombeyoideae. In *Molecular Plant Taxonomy*, ed. Besse P. Totowa, NJ: Humana Press. pp. 337–363 doi: 10.1007/978-1-62703-767-9_17

[16] Shamso E, Khattab A. 2016. Phenetic relationship between Malvaceae s.s. and its related families. *Taeckholmia* 36:115–135

[17] Cole TCH, Lei H, Yu WB. 2024. *MALVACEAE (MalvPP, Chinese)*. www.researchgate.net/publication/370818455_jinkuikexitongfayuhaibao-jinkuikedaibiaoshu_MALVACEAE_MalvPP_Chinese_2024

[18] Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* 24:332–340

[19] Colli-Silva M, Pérez-Escobar OA, Ferreira CDM, Costa MTR, Gerace S, et al. 2025. Taxonomy in the light of incongruence: an updated classification of Malvales and Malvaceae based on phylogenomic data. *Taxon* 74:361–385

[20] Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19:11–15

[21] Luo R, Liu B, Xie Y, Li Z, Huang W, et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* 1:18

[22] Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, et al. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* 21:241

[23] Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics* 31:3350–3352

[24] Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* 14:1394–1403

[25] Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, et al. 2017. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* 45:W6–W11

[26] Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649

[27] Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30:772–780

[28] Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, et al. 2020. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Molecular Ecology Resources* 20:348–355

[29] Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, et al. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37:1530–1534

[30] Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14:587–589

[31] Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* 35:518–522

[32] Johnson MG, Pokorny L, Dodsworth S, Botigué LR, Cowan RS, et al. 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology* 68:594–606

[33] Baker WJ, Bailey P, Barber V, Barker A, Bellot S, et al. 2022. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Systematic Biology* 71:301–319

[34] Zhang Z, Xie P, Guo Y, Zhou W, Liu E, et al. 2022. Easy353: a tool to get Angiosperms353 genes for phylogenomic research. *Molecular Biology and Evolution* 39:msac261

[35] Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19:153

[36] Shang HY, Jia KH, Li NW, Zhou MJ, Yang H, et al. 2025. Phytop: a tool for visualizing and recognizing signals of incomplete lineage sorting and hybridization using species trees output from ASTRAL. *Horticulture Research* 12:uhae330

[37] Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, et al. 2019. Genomic architecture and introgression shape a butterfly radiation. *Science* 366:594−599

[38] Tan X, Qi J, Liu Z, Fan P, Liu G, et al. 2023. Phylogenomics reveals high levels of incomplete lineage sorting at the ancestral nodes of the macaque radiation. *Molecular Biology and Evolution* 40:msad229

[39] McLay TGB, Fowler RM, Fahey PS, Murphy DJ, Udovicic F, et al. 2023. 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*. *Molecular Phylogenetics and Evolution* 187:107869

[40] Smith SA, Brown JW, Walker JF. 2018. So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. *PLoS One* 13:e0197433

[41] Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, et al. 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 15:e1006650

[42] Ramírez-Barahona S, Sauquet H, Magallón S. 2020. The delayed and geographically heterogeneous diversification of flowering plant families. *Nature Ecology & Evolution* 4:1232−1238

[43] Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Systematic Biology* 67:901−904

[44] Chamberlain S, Barve V, McGlinn D, Oldoni D, Desmet P, et al. 2021. *rgbif: Interface to the Global Biodiversity Information Facility API*. doi: 10.32614/CRAN.package.rgbif

[45] Matzke NJ. 2013. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography* 5:242−248

[46] Böhnert T, Luebert F, Merklinger FF, Harpke D, Stoll A, et al. 2022. Plant migration under long-lasting hyperaridity−phylogenomics unravels recent biogeographic history in one of the oldest deserts on Earth. *New Phytologist* 234:1863−1875

[47] Yu Y, Harris AJ, Blair C, He X. 2015. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. *Molecular Phylogenetics and Evolution* 87:46−49

[48] Yao Z, Wang X, Wang K, Yu W, Deng P, et al. 2021. Chloroplast and nuclear genetic diversity explain the limited distribution of endangered and endemic *Thuja sutchuenensis* in China. *Frontiers in Genetics* 12:658037

[49] Witharana EP, Iwasaki T, San MH, Jayawardana NU, Kotoda N, et al. 2025. Subfamily evolution analysis using nuclear and chloroplast data from the same reads. *Scientific Reports* 15:687

[50] Gu X, Li L, Li S, Shi W, Zhong X, et al. 2023. Adaptive evolution and co-evolution of chloroplast genomes in Pteridaceae species occupying different habitats: overlapping residues are always highly mutated. *BMC Plant Biology* 23:511

[51] Robbins EHJ, Kelly S. 2023. The evolutionary constraints on angiosperm chloroplast adaptation. *Biology and Evolution* 15:evad101

[52] Asar Y, Sauquet H, Ho SYW. 2024. Evolutionary rates of nuclear and organellar genomes are linked in land plants. *bioRxiv* 1−30

[53] Zhong Y, Bai B, Sun Y, Wen K, Qiao Y, et al. 2024. Comparative genomics and phylogenetic analysis of six Malvaceae species based on chloroplast genomes. *BMC Plant Biology* 24:1245

[54] Rokas A, Williams BL, King N, Carroll SB. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425:798−804

[55] Nakhleh L. 2013. Computational approaches to species phylogeny inference and gene tree reconciliation. *Trends in Ecology & Evolution* 28:719−728

[56] Liu H, Han B, Mou H, Xiao Y, Jiang Y, et al. 2025. Unraveling the extensive phylogenetic discordance and evolutionary history of spurless taxa within the *Aquilegia ecalcarata* complex. *New Phytologist* 246:1333−1349

[57] Mallet J, Besansky N, Hahn MW. 2016. How reticulated are species? *BioEssays* 38:140−149

[58] Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473:97−100

[59] Carvalho MR, Herrera FA, Jaramillo CA, Wing SL, Callejas R. 2011. Paleocene Malvaceae from northern South America and their biogeographical implications. *American Journal of Botany* 98:1337−1355

[60] Long C, Kubatko L. 2018. The effect of gene flow on coalescent-based species-tree inference. *Systematic Biology* 67:770−785

[61] Conover JL, Karimi N, Stenz N, Ané C, Grover CE, et al. 2019. A Malvaceae mystery: a mallow maelstrom of genome multiplications and maybe misleading methods? *Journal of Integrative Plant Biology* 61:12−31

[62] Le Péchon T, Dai Q, Zhang LB, Gao XF, Sauquet H. 2015. Diversification of Dombeyoideae (Malvaceae) in the Mascarenes: old taxa on young islands? *International Journal of Plant Sciences* 176:211−221

[63] McLoughlin S. 2001. The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Australian Journal of Botany* 49:271−300

[64] Smith JF, Stevens AC, Tepe EJ, Davidson C. 2008. Placing the origin of two species-rich genera in the late cretaceous with later species divergence in the tertiary: a phylogenetic, biogeographic and molecular dating analysis of *Piper* and *Peperomia* (Piperaceae). *Plant Systematics and Evolution* 275:9−30

[65] Givnish TJ, Renner SS. 2004. Tropical intercontinental disjunctions: Gondwana breakup, immigration from the boreotropics, and transoceanic dispersal. *International Journal of Plant Sciences* 165:S1−S6

[66] Hoorn C, van der Ham R, de la Parra F, Salamanca S, ter Steege H, et al. 2019. Going north and south: the biogeographic history of two Malvaceae in the wake of Neogene Andean uplift and connectivity between the Americas. *Review of Palaeobotany and Palynology* 264:90−109

[67] Stone BW, Wessinger CA. 2024. Ecological diversification in an adaptive radiation of plants: the role of de novo mutation and introgression. *Molecular Biology and Evolution* 41:msae007

[68] Swenson U, Hill RS, McLoughlin S. 2001. Biogeography of *Nothofagus* supports the sequence of Gondwana break-up. *Taxon* 50:1025−1041