

RESEARCH

Open Access



Incomplete lineage sorting and introgression among genera and species of Liliaceae tribe Tulipeae: insights from phylogenomics

Zhihui Zhang^{1†}, Meizhen Wang^{2†}, Zhaoping Yang³, Hans Peter Comes⁴, Xin Zhong⁵, Ryan A. Folk⁶, Yigang Song⁵, Dana A. York⁷, Kenneth M. Cameron^{8*} and Pan Li^{1*}

Abstract

Background Phylogenetic research in *Tulipa* (Liliaceae), a genus of significant economic and horticultural value, has relied on limited nuclear (mostly nuclear ribosomal internal transcribed spacer, nrITS) and plastid DNA sequences, resulting in low-resolution phylogenetic trees and uncertain intrageneric classifications. The genus, noted for its large genome, presents discordant relationships among *Amana*, *Erythronium*, and *Tulipa*, likely due to incomplete lineage sorting (ILS) and/or reticulate evolution. Thus, phylogenomic approaches are needed to clarify these relationships and the conflicting signals within the tribe Tulipeae.

Results We newly sequenced 50 transcriptomes of 46 species of tribe Tulipeae (including multiple accessions of all four genera) and one outgroup species of the sister tribe Lilieae (*Notholirion campanulatum*), and downloaded 15 previously published transcriptomes of tribe Tulipeae to supplement the sampling. One plastid dataset (74 plastid protein-coding genes, PCGs) and one nuclear dataset (2594 nuclear orthologous genes, OGs) were constructed, with the latter used for species tree inference based on maximum likelihood (ML) and multi-species coalescent (MSC) methods. To investigate causes of gene tree discordance, “site con/discordance factors” (sCF and sDF1/sDF2) were calculated first, after which phylogenetic nodes displaying high or imbalanced sDF1/2 were selected for phylogenetic network analyses and polytomy tests to determine whether ILS or reticulate evolution best explain incongruence. Key relationships not resolved by this technique, especially those among *Amana*, *Erythronium*, and *Tulipa*, were further investigated by applying D-statistics and QuIBL.

Conclusions We failed to reconstruct a reliable and unambiguous evolutionary history among *Amana*, *Erythronium*, and *Tulipa* due to especially pervasive ILS and reticulate evolution, likely caused either by obscured minority phylogenetic signal or differing signals among genomic compartments. However, within *Tulipa* we confirmed the monophyly of most subgenera, with the exception of two species in the small subgenus *Orithya*, of which *Tulipa heterophylla*

[†]Zhihui Zhang and Meizhen Wang contributed equally to this work.

*Correspondence:
Kenneth M. Cameron
kmcameron@wisc.edu
Pan Li
panli_zju@126.com

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



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

was recovered as sister to the remainder of the genus, whereas *T. sinkiangensis* clustered within subgenus *Tulipa*. In contrast, most traditional sections of *Tulipa* were found to be non-monophyletic.

Keywords Phylogeny, Incomplete lineage sorting (ILS), Introgression, Reticulate evolution, Tulipeae, Transcriptome sequencing

Background

Tulipa L. (Liliaceae, Lilioideae) is a northern temperate to subtropical Old World genus of some 76 species [1] world-renowned for its considerable economic value and horticultural interest (Fig. 1). All tulips are perennial, bulbiferous herbs (geophytes) native to Europe/northwestern Africa to China, but they reach their highest diversity in Central Asia [2]. Although the Liliaceae family has undergone many taxonomic changes, the status of *Tulipa* and its closest relatives as members of a tribe Tulipeae (or

rarely Liliaceae *s.l.*) has remained stable in both traditional and APG classification systems [3, 4].

Apart from *Tulipa*, the tribe Tulipeae also includes three other genera: *Amana* Honda, *Erythronium* L., and *Gagea* Salisb. (including *Lloydia* Salisb. ex Rchb.) [5–7]. There is substantial evidence that *Gagea* is sister to the other members of Tulipeae. However, there are ongoing debates regarding the phylogenetic relationships among the latter three genera. Analyses based on the nuclear ribosomal internal transcribed spacer (nrITS)



Fig. 1 Representative members of Liliaceae tribe Tulipeae. **A** *Gagea granulosa*. **B** *G. serotina* (= *Lloydia serotina*). **C** *Amana edulis*. **D** *Erythronium sibiricum*. **E** *Tulipa heterophylla*. **F**, *T. heteropetala*. **G** *T. uniflora*. **H** *T. clusiana* (subgenus *Clusianae*). **I** *T. montana* (subgenus *Clusianae*). **J** *T. dasystemon* (subgenus *Eriostemones*). **K** *T. saxatilis* (subgenus *Eriostemones*). **L** *T. turkestanica* (subgenus *Eriostemones*). **M** *T. sinkiangensis*. **N** *T. kaufmanniana* (subgenus *Tulipa*). **O** *T. thianschanica* (subgenus *Tulipa*). **P** *T. altaica* (subgenus *Tulipa*). Photograph H by Xiang-Yu Zhou, all others by Xin Zhong

Table 1 Comparison of the phylogenetic relationship among *Erythronium*, *Amana*, and *Tulipa*. The gene regions and methods used for analysis are indicated. Sister groups are abbreviated as their initials (e.g., TA|E means sister relationship between *Tulipa* and *Amana*)

Article	TA E	TE A	EA T
Hayashi and Kawano, 2000 [8] ^a	<i>rbcL</i> (MP)		
Allen and Soltis et al., 2003 [9]	<i>matK</i> (MP)	<i>matK</i> (MP)	<i>matK</i> (ML)
Rønsted and Law et al., 2005 [10]			<i>matK</i> (MP)
Zarrei and Wilkin et al., 2009 [11]	nrITS (MP) nrITS + five plastid regions ^b (MP)		five plastid regions ^b (MP)
Clennett and Chase et al., 2012 [12]			nrITS, <i>rps16</i> + <i>matK</i> intron, morphological data ^d (MP, Bayesian)
Kim and Hong et al., 2013 [13]			nrITS + five plastid regions ^c (MP, Bayesian)
Li and Lu et al., 2017 [5]			whole plastomes (ML, Bayesian)
Kim and Kim, 2018 [6]		<i>rbcL</i> + <i>matK</i> + <i>ndhF</i> + <i>atpB</i> (ML, Bayesian)	
Lu and Yang et al., 2021 [7]			whole plastomes (ML)

^a *matK* was also used in this study to reconstruct phylogenetic relationship, but the branch of *Erythronium*, *Amana*, and *Tulipa* collapsed into a polytomy

^b Five plastid regions: *rpl16* intron, *trnL* intron, *trnL-F* spacer, *matK*, and the *psbA-trnH* spacer

^c Five plastid regions: *trnL* intron and *trnL-trnF* spacer, *rpl16* intron, *rps12-rpl20* intergenic spacer, and *matK*

^d Whether nrITS, *rps16* + *matK* intron, and morphological data alone or combined together, the topology of EA is supported. Note: In this study, *Amana* embedded in *Erythronium* sometimes

and several plastid loci (*matK*, *rbcL*, *rps16*, etc., and even whole plastomes) employing different models (Maximum Parsimony [MP], ML, Bayesian) have yielded inconsistent results with support for all possible resolutions of relationships among the three genera (Table 1) [5–13]. Although there is increasing molecular systematic evidence, particularly from whole plastomes, to support a sister relationship between *Amana* and *Erythronium*, the true evolutionary history may be otherwise, since there is still no published study based on low copy nuclear genes. Inconsistent results from different studies indicate that complex evolutionary phenomena like plastid capture or incomplete lineage sorting may lead to phylogenetic incongruence between nuclear and plastid genomes, but without a broad sampling of nuclear genomic loci we lack a path to test this.

Moreover, many taxonomic issues remain unresolved within *Tulipa*. Contemporary studies mostly rely on the classification system of Zonneveld [14, 15], which divides *Tulipa* into four subgenera (*Orithyia*, *Clusianae*, *Eriostemones*, and *Tulipa*) according to nuclear DNA content and morphology. Based on sequences of nrITS and five plastid regions (*trnL* intron, *trnL-F* intergenic spacer, *rpl16* intron, *rps12* intron, partial *matK*), Christenhus et al. [1] confirmed the monophyly of three subgenera in this classification system, but the monophyly of *Orithyia* could not be tested since only one species of that group (*T. uniflora*) had been sampled. However, limited by very low resolution and support for relationships below the level of subgenus, some clades in their study collapsed into polytomies, meaning that no study to date has had

sufficient sampling of multi-genomic loci and taxa to adequately assess species relationships or evaluate the natural status of recognized taxonomic sections within the subgenera.

An ideal approach to further resolve the phylogeny of *Tulipa* and its relationship to related genera would be whole-genome-sequencing (WGS), but this is time-consuming and costly. In addition, the very large genome size (e.g., *Tulipa*, DNA 2C-value = 32–69 pg) [15] makes this methodologically prohibitive at the present time for most of the species. RNA sequencing (RNA-Seq), on the other hand, is an appropriate alternative by which numerous genes and their different evolutionary histories can be analyzed to uncover phylogenetic relationships and analyze gene tree conflict [16–18]. Robust gene sampling of both the nuclear and plastid genomes is possible because plastid protein-coding genes (PCGs) are almost entirely transcribed [19].

Of course, an increase in the amount of available molecular phylogenetic data, on its own, is unlikely to resolve conflicting phylogenetic signals among different gene regions and genomes (nuclear, plastid, mitochondrial) without consideration of alternative methods of phylogenetic analysis, because standard species tree inference methods do not explicitly differentiate causes of gene tree conflict [20]. Instead, it is necessary to try to reveal underlying evolutionary causes of gene tree discordance, such as incomplete lineage sorting (ILS; in the retention of ancestral allelic polymorphisms among descendant lineages) and/or reticulation (the transfer of gene material between species [21]).

In this study, we newly sequenced 50 transcriptomes of 46 species of Tulipeae (including multiple accessions of several species), representing all four genera (and their major clades [1, 5, 12]). We also included *Notholirion campanulatum* Cotton & Stearn as an outgroup taxon for two reasons. First, it is a member of the tribe Lilieae, a group that stands as the sister lineage to the tribe Tulipeae. Second, its native range—the Qinghai-Tibet Plateau (QTP) and adjacent regions—serves to minimize potential gene flow [22]. These characteristics make it an ideal outgroup for rooting the phylogenetic trees. Based on this dataset plus 15 previously published transcriptomes of tribe Tulipeae, one plastid dataset and one nuclear dataset (with its two sub-datasets) were constructed for estimating phylogenetic trees by different methods (ML and MSC). Reticulate evolution analyses and polytomy tests were used for exploring inconsistent phylogenetic signals among different trees and thus inferring a most likely phylogeny of tribe Tulipeae, focusing on the genus *Tulipa*. We also paid special attention to the phylogenetic relationships among *Tulipa*, *Erythronium*, and *Amana* using D-statistics and QuIBL to assess the alternative contributions of ILS and introgression to the substantial gene tree discordance among these lineages.

Results

Phylogenetic tree reconstruction

In our study, well-supported clades were defined as those with ultrafast bootstrap support values (UF value, for ML) ≥ 95 or local posterior probabilities (LPP, for ASTRAL) ≥ 0.95 . All of the phylogenetic trees described hereafter can be found in Additional file 1: File S1.

First, considering the plastid genome tree reconstructed from 74 PCGs, about two-thirds of the nodes are highly supported (Fig. 2).

The monophyly of all four ingroup genera with the topology of (*Gagea*, (*Tulipa*, (*Erythronium*, *Amana*))) is well supported. Within genus *Tulipa*, the subgenera *Clusianae*, *Eriostemon*, and *Tulipa* are supported to be monophyletic, whereas subgenus *Orithyia* is not: *T. heterophylla* and *T. heteropetala* are sister to all other taxa of *Tulipa* successively, and *T. sinkiangensis* is sister to subgenus *Tulipa*.

However, these plastid-based results show limited phylogenetic resolution at the species level. Some presumably closely related samples were not sister or even in the same clade, such as *T. patens* (this study)/*T. patens* (NCBI), *T. humilis*/ *T. humilis*2, *T. thianschanica* (this study)/*T. thianschanica* (NCBI), *T. iliensis* (this study)/*T. iliensis* (NCBI), and *T. xgesneriana* (this study)/*T. xgesneriana* (NCBI). This outcome is influenced not only by the inherent characteristics of plastid PCGs but also by factors such as misidentifications or species

complexes, which will be further discussed in the “Discussion” section.

Turning to the low-copy nuclear genome topology, the reconstructed ML trees and ASTRAL trees from 2594 nuclear OGs (and its subsets) are generally well supported, albeit with a few nodes that lack support (Figs. 3 and 4). Notably, the nodes with lowest support values in both the ML and coalescent trees are related to *T. sinkiangensis*.

At the genus level, the monophyly of *Amana*, *Eriostemon*, *Gagea*, and *Tulipa* is highly supported in the nuclear trees. However, the topology of the ingroup was resolved differently from the plastid genome tree, recovering (*Gagea*, (*Erythronium*, (*Tulipa*, *Amana*))), which is only weakly supported in the coalescent tree based on 2594 OGs. Conversely, in the dataset of 1594 OGs with a relatively low copy number, the topology recovered was (*Gagea*, (*Tulipa*, (*Erythronium*, *Amana*))) with high support (Fig. 4). Interestingly, the phylogenetic relationships reconstructed by ASTRAL with the remaining higher-copy 1000 OGs are exactly the same as those reconstructed with all OGs, with even greater support. There are no topological differences between ML trees based on 1594 and 2594 OGs (Fig. 3).

At the subgenus level within *Tulipa* all nuclear trees are consistent with the plastid tree, and the subgenera of genus *Tulipa* are well supported except for subg. *Orithyia*. Subgenus *Eriostemon* (excluding *T. patens*) gives rise to two main clades, one of which contains all *T. biflora* samples. We refer to this clade as Clade E1, and the other as Clade E2. Subgenus *Tulipa* (excluding *T. sinkiangensis*) similarly gives rise to two primary clades. One clade is represented by *T. iliensis*, which we abbreviate as Clade T1. The other is referred to as Clade T2. For ease of discussion, within Clade T2, we define three subclades: Clade alt, representing *T. altaica* and its sister taxa; Clade acu, representing *T. acuminata* and its sister taxa; and Clade eic, representing *T. eichleri* and its sister taxa.

In contrast, most recognized sections of *Tulipa* (Additional file 2: Table S1) are not supported in this study, with the exception of Neo-tulipae, sect. *Saxatiles* (embedded in sect. *Sylvestres*) and sects. *Multiflorae* and *Spiranthera* (only one sample).

Regarding species placements, *T. turkestanica* was found to be embedded in the clade of *T. biflora*, and the phylogenetic position of *T. patens*, *T. sprengeri*, *T. sinkiangensis*, and *T. armena* are different among analyses (Figs. 3 and 4). One important point is that *T. thianschanica* var. *sailimuensis* is no closer to *T. thianschanica* than *T. iliensis* and *T. tetraphylla*.

Finally, we used a simulation-based approach to characterize ILS expectations based on the nuclear topology.

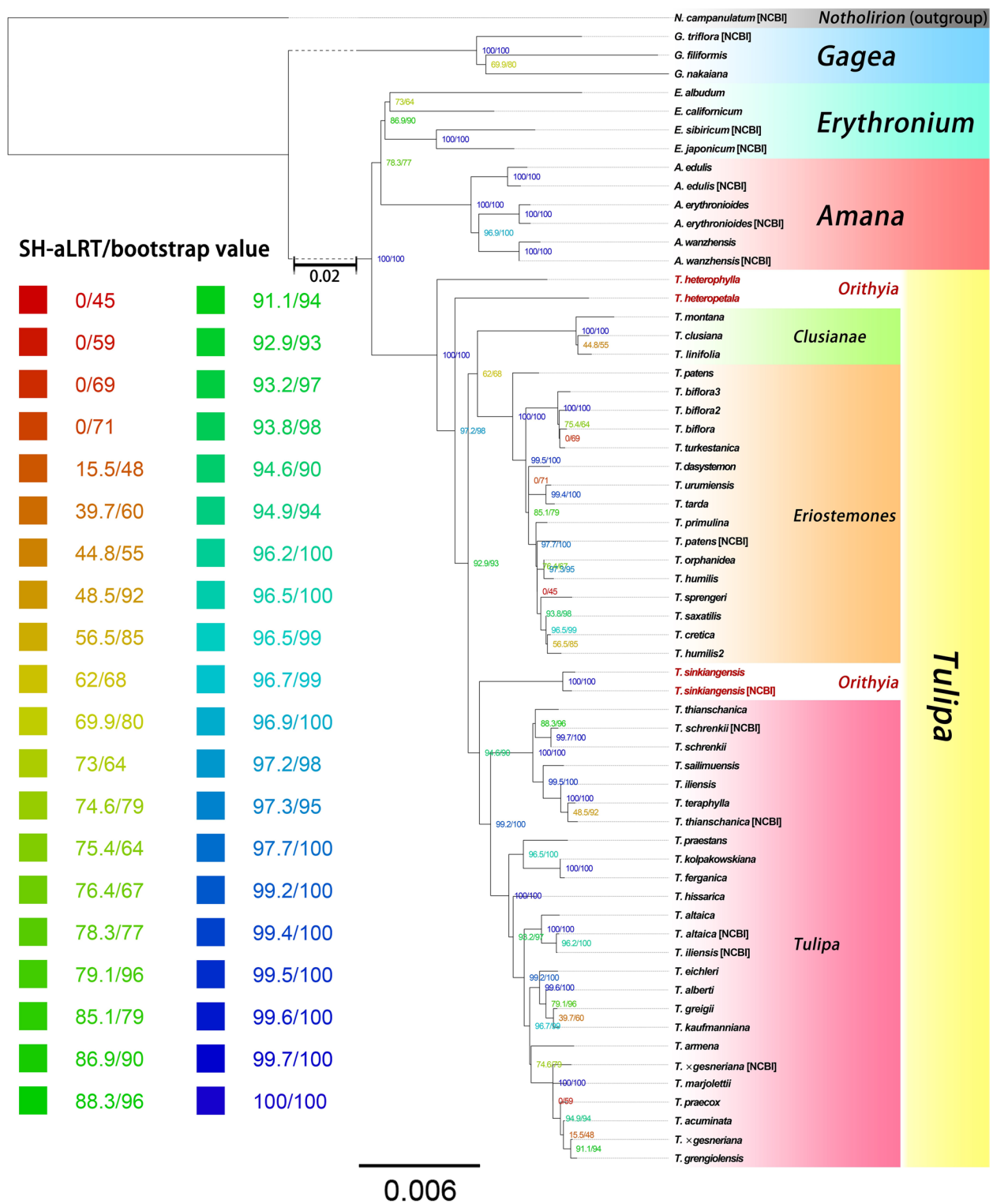


Fig. 2 ML tree based on 74 PCGs. Support values are marked in the form of SH-aLRT/bootstrapped value. The support values from low to high corresponds to the gradient color from red to blue

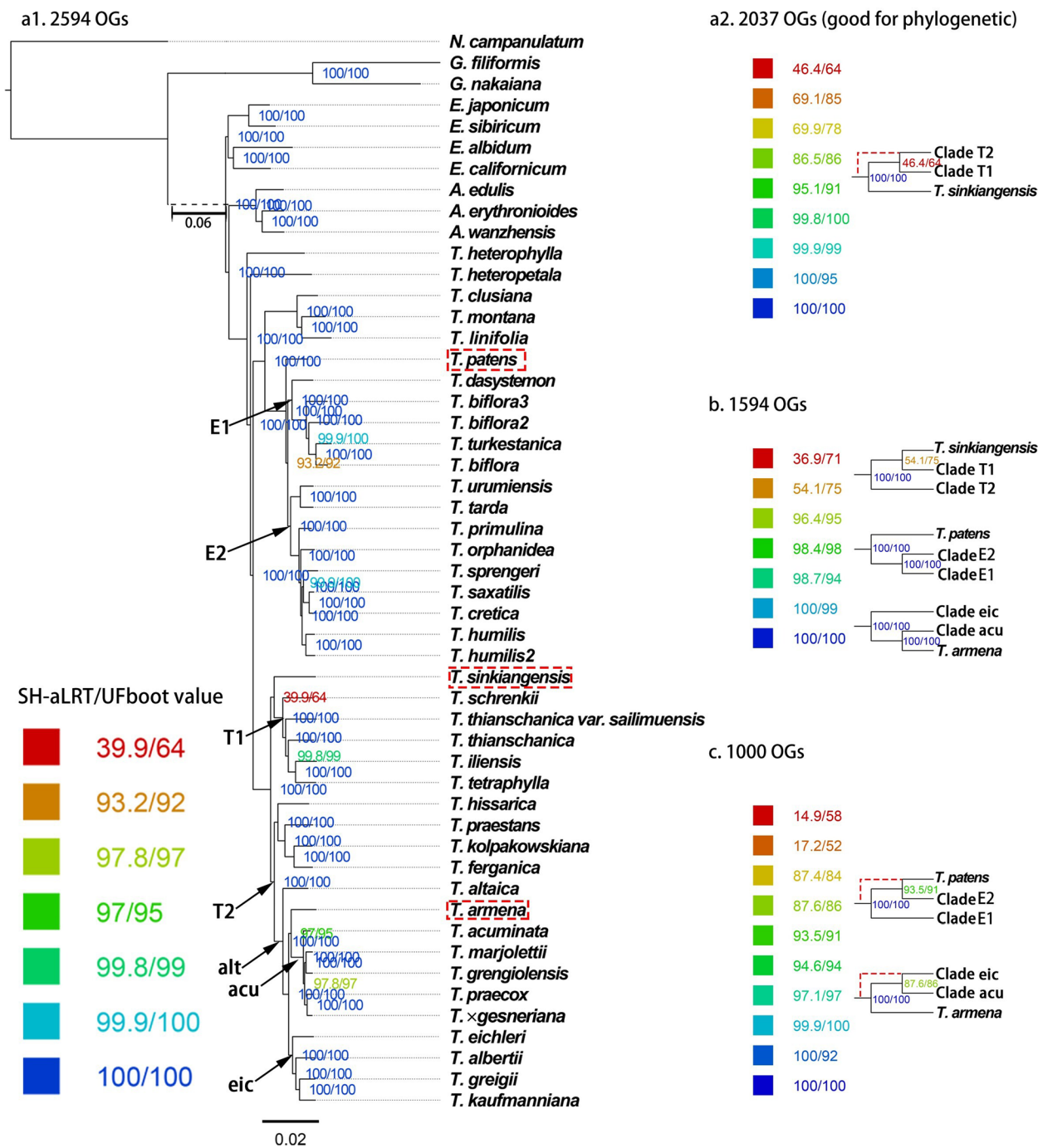
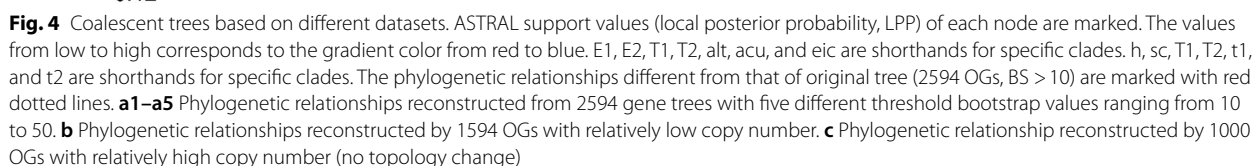
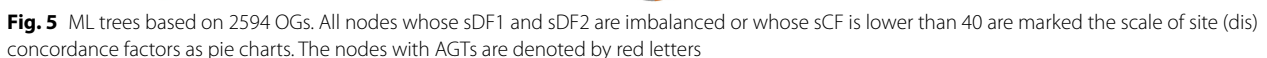


Fig. 3 ML tree based on different datasets. Support values are marked in the form of SH-aLRT/bootstrap value. Support values from low to high corresponds to the gradient color from red to blue. E1, E2, T1, T2, alt, acu, and eic are shorthands for specific clades. The phylogenetic relationships different from that of original tree (2594 OGs) are marked with red dotted lines. **a1** Phylogenetic relationships reconstructed from concatenated 2594 OGs, among them, 2037 OGs were considered suitable for phylogenetic analysis by IQ-TREE2, thus **a2** was also generated. **b** Phylogenetic relationship reconstructed by 1594 OGs with relatively low copy number (no topology change). **c** Phylogenetic relationship reconstructed by 1000 OGs with relatively high copy number



It is worth noting that we also observed conflict among nuclear analyses; there are several nodes in the ML tree that have high support but conflict with those in our coalescent tree. These nodes may be those most affected

by ILS, and the coalescent tree is therefore likely a better reflection of the real evolutionary history of these taxa [23], because concatenation is positively misleading under high ILS with introgression (although both concatenation and coalescent methods are inconsistent in the presence of introgression [21, 24, 25] as our result indicated (see below, *Conflicting nuclear phylogenetic signals*). Therefore, we have taken the 1594 OGs-based coalescent tree as the main basis for our discussion, and will further elaborate on the logic of this choice further below.



For those nodes that may contain taxa involved in ILS or introgression, the ratios of sCF/sDF1/sDF2 are shown in Figs. 5 and 6 (pie charts; shown only to two significant digits; further data can be found in Additional file 1: File S2). Unlike the consistent support recovered across the

Almost no node in tribe Tulipeae (especially *Tulipa*) has a sCF higher than 50 (%). In ML tree (Fig. 5), anomalous gene trees (AGTs, i.e., incongruent gene trees [IGTs] with a higher proportion than congruent

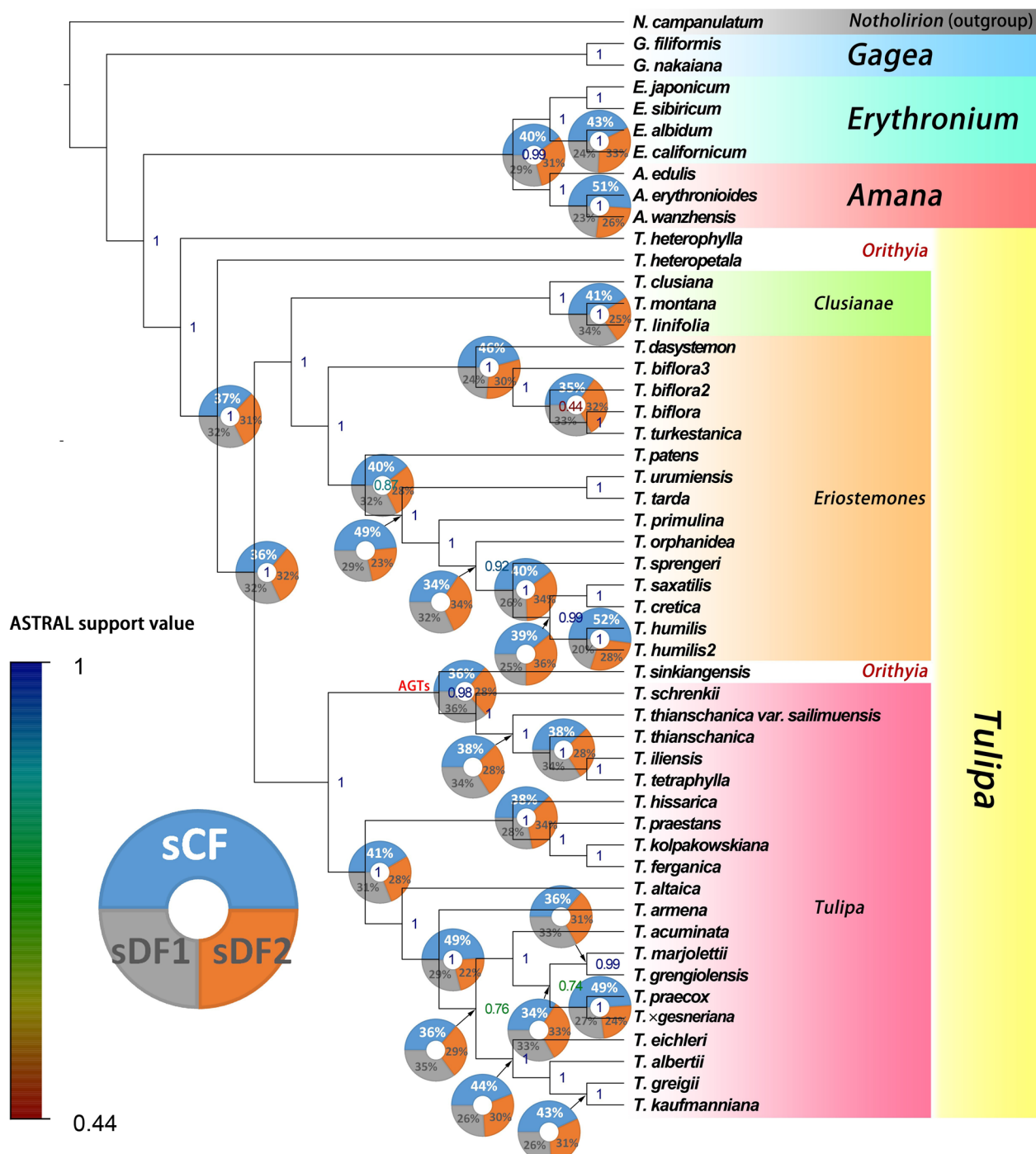


Fig. 6 Coalescent trees based on 1594 OGs (BS > 10). All nodes whose sDF1 and sDF2 are imbalanced or whose sCF is lower than 40 are marked the scale of site (dis)concordance factors as pie charts; One node with AGTs is denoted by red letters

gene tree [CGT] [26], as detailed in the “Distinguishing between ILS and introgression” section in “[Methods](#)”) were detected at five nodes which corresponds to five problematic taxa with unstable positions among the

analyses: genus *Amana*, *T. patens*, *T. sprengeri*, *T. sinkiangensis*, and *T. armena*.

Considering that only one node in the ASTRAL tree has a detected AGTs (the node of *T. sinkiangensis*, Fig. 6,

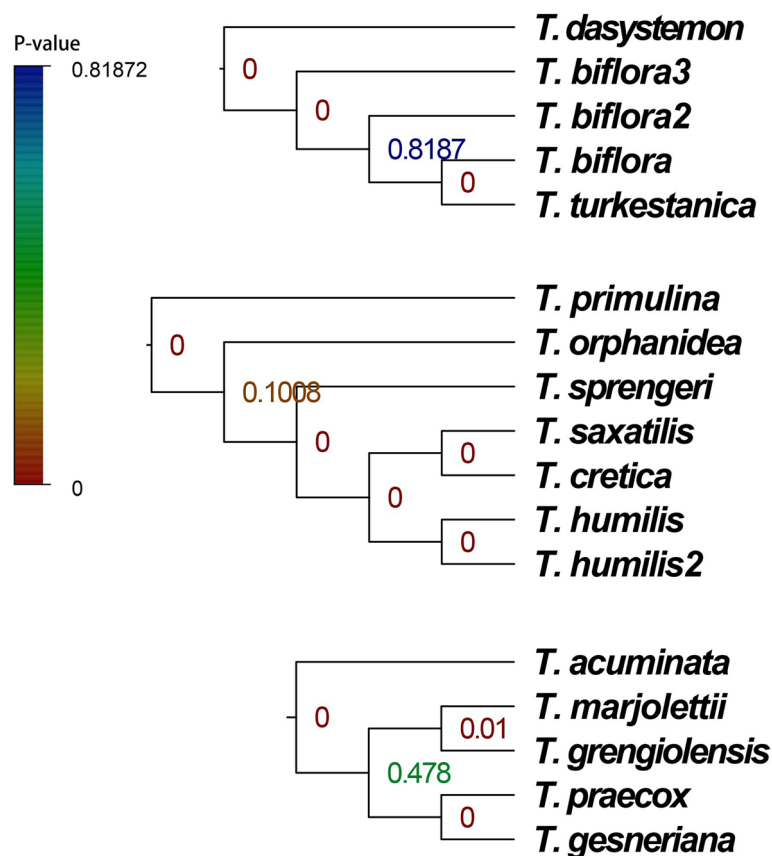


Fig. 7 Four nodes with P -value > 0.01 in polytomy tests

vs. five nodes in ML tree), we feel that the phylogenetic relationships in our ASTRAL tree is the more reliable estimate and was the basis of our later analyses. Although most nodes have imbalanced sDF1/2 in the ASTRAL tree, we focused on selecting taxa near those nodes with $|sDF1 - sDF2| \geq 5$ (%) as problematic taxa also, which may be affected by introgression but with stable position: (1) *E. albidum*, *E. californicum*, and Eurasian branch of *Erythronium*; (2) *T. linifolia*, *T. montana* and *T. clusiana*; (3) *T. dasystemon*; (4) *T. orphanidea*; (5) *T. thianschanica* var. *sailimuensis*; (6) *T. hissarica*.

Polytomy tests and reticulate evolution

Considering that some signals of phylogenetic conflict may arise from polytomy, we first conducted a polytomy test on the 1594 OGs-based ASTRAL tree (Additional file 1: File S3, partly shown in Fig. 7). There are three nodes with p -value > 0.05 , which suggests polytomies for *T. biflora*/*T. turkestanica*, *T. gesneriana*/*T. praecox*/Neo-tulipae, and *T. orphanidea*/*T. sprengeri*/sect. *Saxatiles*. If taking a lower cutoff value p -value > 0.01 , then *T. marjolettii* and *T. grengiolensis*, two species of Neo-tulipae, could also be considered to form a polytomy.

Figure 8 shows the best-fit network analyses to detect reticulate evolution for major problematic taxa.

First, to explore the deeper relationships among *Amana*, *Erythronium*, and *Tulipa*, representatives of all major taxa of tr. Tulipeae were selected for reticulate evolution analysis (Figs. 8a, b).

This analysis predicted two different topologies. The first topology, denoted as ((*Tulipa*, *Amana*), *Erythronium*) (abbreviated as TA|E, Fig. 8a), appeared more robust and yielded higher scores. This network prediction suggested that *Tulipa* acquired 12.9% of its genome (as represented by transcriptomic loci) from the common ancestor of *Amana*, *Erythronium*, and *Tulipa*. Additionally, a significant gene flow of 40% was detected between *Amana* and the American branches of *Erythronium*. These findings therefore partially explain why *Amana* and *Erythronium* are occasionally sister. Furthermore, within the genus *Erythronium*, it was found that the ancestor of *E. californicum* (western North American) contributed approximately 15 and 35% of genetic components to *E. japonicum* (Eurasia) and *E. albidum* (Eastern North American), respectively.

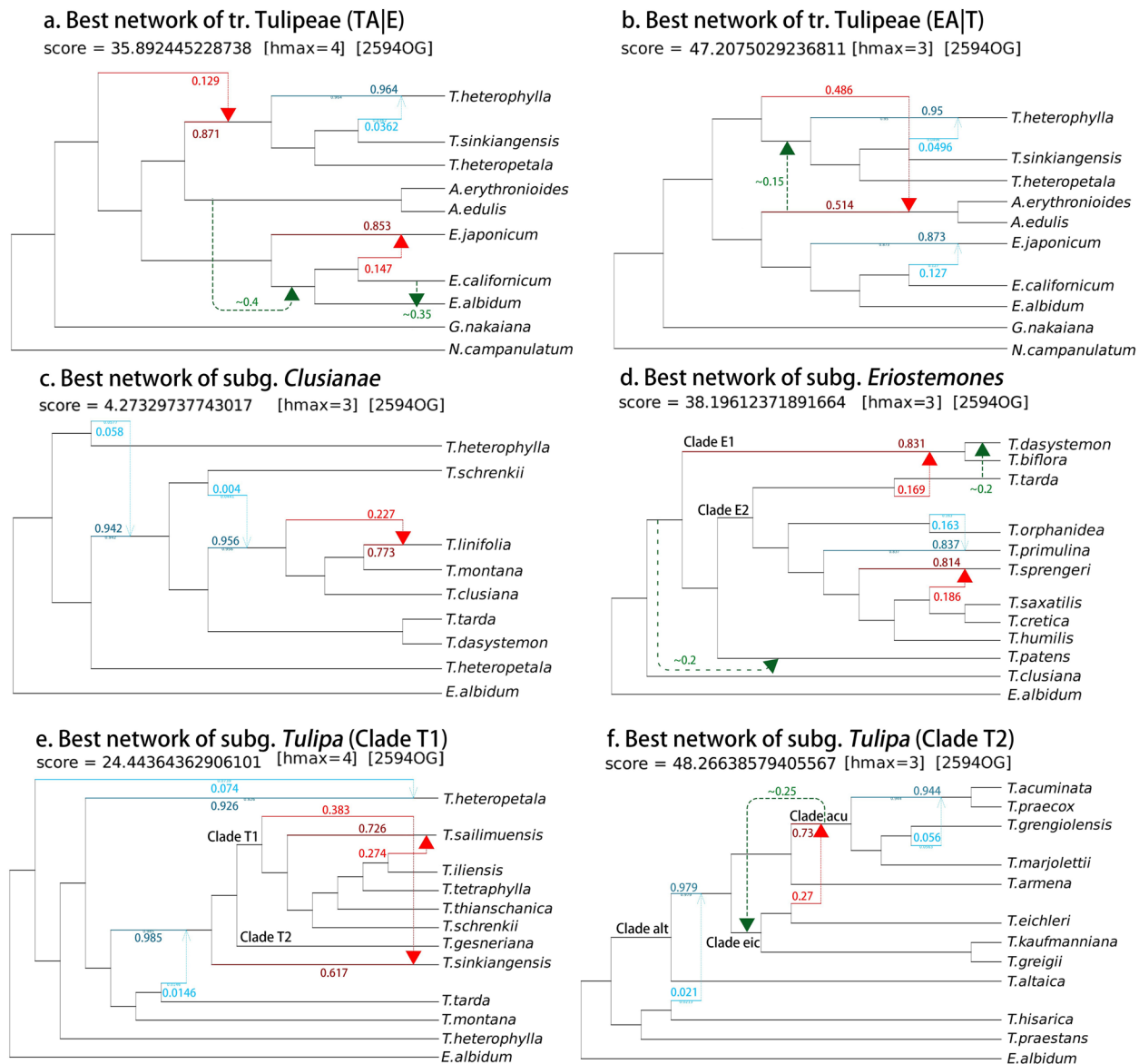


Fig. 8 Best networks predicted by PHYLONETWORKS. The gene flows involving problematic taxa are the focuses, which are shown in red. The unimportant gene flows are represented in blue and not highlighted. The gene flows appeared many times in suboptimal networks are also displayed in form of green arc dotted line. Those values next to each hybrid edge indicate inheritance probabilities

Figure 8b illustrates the network topology identified as ((*Erythronium*, *Amana*), *Tulipa*) (abbreviated as EA|T). Under this network, nearly 50% gene flow from the ancestor of *Tulipa* to *Amana* suggests that *Amana* is a hybrid product between *Erythronium* and *Tulipa*. Conversely, alternative predictions propose that the ancestor of *Amana* provided 15% of genetic components to *Tulipa*. However, upon conducting subsequent rounds of analysis, wherein higher levels of gene flow were considered, all networks with the EA|T structure transitioned to the TA|E configuration.

In other groups, we mainly discuss the phylogenetic relationship of each subgenus within the genus *Tulipa*, taking into account the following issues: (1) *N. campanulatum* is distantly related to the ingroup; (2) introgressions between *Gagea* and *Erythronium*/*Amana*/*Tulipa* were detected frequently; and (3) the relationship between *Amana* and *Tulipa* is still unclear, with evident gene flow but different details regarding this gene flow depending on the favored network. We replaced the outgroup with *E. albidum*, a more suitable outgroup choice for lower phylogenetic levels that also does not overlap

geographically with *Tulipa* and which yields no evidence of recent gene flow between the two genera aside from the ancient events identified in the backbone.

In subgenus *Clusianae*, as previously predicted, two possible introgressions among *T. clusiana*, *T. montana*, and *T. linifolia* were detected (Fig. 8c). The best network shows that *T. linifolia* acquired more than 20% of its genome from the ancestor of subgenus *Clusianae*.

In subgenus *Eriostemon* (Fig. 8d), the optimal network reveals that clade E1 acquired 16.9% of its genome from *T. tarda*. This elucidates why *T. patens* is intermittently positioned as the basal taxa of subg. *Eriostemon* and sometimes only as the sister of Clade E2. Additionally, gene flow from the ancestor of subg. *Eriostemon* to *T. patens* in another network further explain this.

Another network also detected gene flow from *T. tarda* to *T. dasystemon*, providing an explanation for *T. dasystemon*'s inclusion as a member of Clade E2. The optimal network places *T. sprengeri* as the sister of *T. humilis*, *T. saxatilis*, and *T. cretica*, aligning with the plastid phylogenetic tree. The reason for this cytonuclear discord may therefore be the 18.6% genome exchange between *T. sprengeri* and *T. saxatilis* as recovered in this network.

In subgenus *Tulipa*, *T. sinkiangensis* is supported as the sister to the entire subgenus. However, *T. sinkiangensis* obtained 38.3% of its genome from the ancestral lineage of Clade T1. This may be why *T. sinkiangensis* is considered the sister of Clade T1 in the nuclear gene phylogeny (Fig. 8e). *T. thianschanica* var. *sailimuensis* has obtained 27.8% of its genetic components from *T. iliensis* (Fig. 8e).

T. eichleri contributed 27% of its genetic components to Clade acu, while another network predicted gene flow between Clade acu and Clade eic. This explains the fluctuation in the position of *T. armena* between the base of Clade acu and the base of Clade acu and Clade eic. (Fig. 8f).

Discussion

Strategies for discriminating between introgression and ILS

Introgression is a common phenomenon in plant evolution [27] and distinguishing it from ILS is a crucial step in phylogenetic analysis. The processes of introgression and ILS both create gene tree discordance, but they predict differing gene tree distributions that can be distinguished using a variety of probabilistic approaches [28]. Common phylogenetic tools to distinguish ILS and introgression, such as D-statistics (a site-based method) [29] and quantifying introgression via branch lengths (QuIBL, a tree-based method) [30], can quickly detect the existence of introgression by testing whether the null hypothesis (that only ILS occurred) can be rejected. However, these methods cannot directly quantify the proportion

of introgression present and often impose hard bounds on the maximum number of taxa and topologies that can be tested in one analysis, limiting their application in species-rich clades [28]. On the other hand, although some phylogenetic tools for reticulate evolution, such as PHYLONETWORKS [31], can infer networks of arbitrary shape with quantified introgression, these methods require an unrealistically large amount of computational resources and time even with moderate numbers of taxa [32]. Hence, these methods are limited either by a priori sampling limitations (for D-statistics and QuIBL) or by indirect sampling limitations via computational resource scaling (for PHYLONETWORKS), creating the need for “divide-and-conquer” approaches for the common case where it is unclear which taxa specifically are involved in hybridization.

One feasible “dive-and-conquer” approach to roughly identify nodes affected by ILS/introgression, and thus to adjust the strategy and scope of introgression analysis accordingly, is to identify globally on a phylogeny those problematic nodes that would form the basis for smaller-scale tests with more elaborate probabilistic approaches. As implemented here, first we calculate for each node estimates of the site concordance factor (sCF) and two site discordance factors (sDF1 and sDF2). As sDF1/2 actually represents the proportion of two IGTs (detailed in the section “Distinguishing between ILS and introgression” in “Methods”), in a quartet (based on parsimony-informative sites), the taxa affected by ILS and introgression can be roughly determined from the high/imbalanced values of sDF1/2 respectively.

We then reduce the scope of hybridization inference to those nodes with suggestive patterns of discord on the basis of site discordance factors, in this case those nodes with imbalanced sDF1/2, yielding on a quantitative basis for identifying candidates for a more computationally feasible sub-analysis. Specifically, there are two possible scenarios to consider: for *Amana* (A), *Erythronium* (E), and *Tulipa* (T), we can determine that the bipartition EA|T corresponds to CGT based on a plastid derived tree topology, whereas the alternative bipartition TA|E corresponds to the IGT of higher proportion (versus TE|A). This indicates two possible introgressions: one is between A and T, and the other is directional gene flow from an unsampled common ancestor/sister of “E, A and T” to E, two alternative scenarios that can be distinguished by divergence time, internal branch length, and other non-topological information.

Phylogenetic framework of tribe Tulipeae

The tribe Tulipeae consists of four monophyletic genera: *Amana*, *Erythronium*, *Gagea* (including *Lloydia*), and *Tulipa*. *Gagea* has always been regarded as sister

to the others, which is supported by our phylogenetic study and others [5–7]. The relationships among the three remaining genera, however, remains ambiguous. In addition to detecting high levels of ILS within this tribe, we also documented numerous cases of introgression among *Amana*, *Erythronium*, and *Tulipa*, which makes their phylogenetic relationship particularly challenging to resolve. Previous studies showed substantial disagreement, recovering all three possible resolutions of these genera, but there are very few studies that strongly support a sister relationship between *Erythronium* and *Tulipa* [6, 9]; neither does our result.

Most previous studies support a topology of EA|T with *Erythronium* sister to *Amana* [5, 7, 10, 11, 13] (Table 1). This is consistent with the plastid ML tree and the ASTRAL tree based on our lower copy nuclear dataset (1594 OGs). The sDF values also show that the parsimony-informative sites slightly favor *Amana* as sister to *Erythronium* (Figs. 5 and 6).

In contrast, most of the networks predicted by PHYLONETWORKS, including the best ones, indicate that *Amana* is sister to *Tulipa* (not *Erythronium*). Those networks with a structure of TA|E scored much lower (better) than the networks with structure of EA|T, even if we deliberately used the guide tree of EA|T structure. We must take this “dissent” seriously. When combined with high levels of ILS, even low levels of extant introgression (e.g., among non-sister clades) can cause AGTs and thus misleading species tree estimations, while even massive introgression may not necessarily lead to AGTs [33, 34]. With the simultaneous presence of ILS and introgression, neither concatenation nor coalescent methods are robust, while coalescent-based networks are consistent [24]. In essence, the MPL model has the ability to take both ILS and introgression into account [35], therefore potentially resolving discord quite differently from methods assuming a tree model, which gives the result substantial weight.

Because of the uncertain relationship, we used three topologies: ((*Erythronium*, *Amana*), *Tulipa*), ((*Tulipa*, *Erythronium*), *Amana*), and ((*Tulipa*, *Amana*), *Erythronium*) to carry out D-statistic analyses (shorthand as EA|T, TE|A and TA|E, respectively) and chose *N. campanulatum* (N, outgroup), all spp. in *Erythronium* and *Amana*, and 6 spp. in *Tulipa* as samples for 72 independent sets of D-statistic runs. All 72 test groups using topology TE|A detected significant introgression between *Erythronium* and *Amana* (Additional file 2: Table S2), which indicated that TE|A is an unlikely evolutionary history because introgression only occurred between *Erythronium* and *Amana*, which would lead to the dominance of TEA (polytomy) or EA|T but would not explain a close relationship between *Amana* and *Tulipa*. Similar

to TE|A, the results based on TA|E showed that there was highly significant introgression between *Erythronium* and *Amana* in all the 72 test groups (P -value = 0, Additional file 2: Table S2). This result is unusual because the D values do not show a particularly obvious species correlation, if the structure of TA|E is true, then each branch of *Amana* and *Erythronium* (geographically isolated) all have experienced continuous introgressions, which is very doubtful. EA|T's results showed that 38 out of 72 test groups detected introgression between *Amana* and *Tulipa*; 5 test groups showed introgression between *Erythronium* and *Tulipa*; the remaining 29 test groups did not detect introgression (Additional file 2: Table S2). All E3 (18) and almost all E2 (17) test groups showed introgression between *Amana* and *Tulipa*, while most E1 (15) and E4 (13) test groups did not support the existence of any introgression. It is intriguing that all instances of introgression between *Erythronium* and *Tulipa* are exclusively linked to E4. Given the general belief that the divergence time of *Erythronium* is earlier than that of *Tulipa* and *Amana* [6, 12, 36], this result is reasonable, because the introgression signal is related to each branch of *Erythronium*, but is almost independent of *Amana* and *Tulipa*. These findings suggest that either *Erythronium* and *Amana* have undergone extensive and continuous gene flow in their history (TA|E), and/or they are actually sister taxa (EA|T).

However, the analyses based on internal branch lengths provided different results (Additional file 2: Table S3). Based on the same groupings described above for the D-statistic, QuIBL were used for performing internal branch length analysis of the 72 test groups. The mean frequencies of gene trees supporting specific topologies are as follows: 28.07% (TE|A), 36.02% (TA|E), and 35.37% (EA|T). This indicates that both TA|E and EA|T are potential candidates for being consistent with the true species tree. However, given that introgression between *Amana* and *Tulipa* affects approximately 66.67% of triplets and 20.56% of loci, which is greater than the impact of introgression between *Amana* and *Erythronium* (47.22% and 19.76%), TA|E is the best supported topology.

Due to the close proximity of the proportions of gene trees that support the three triplets, it was difficult for us to interpret the results of QuIBL. As a result, we used ASTRAL for calculating the corresponding coalescent trees of 72 triplets with N as an outgroup, and the phylogenetic relationships were inferred finally (Table 2 and Additional file 1: File S4). In terms of quantity, TA|E is more frequent than EA|T, but interestingly, the Eurasian branch of *Erythronium* (E1 & E2) strongly supports *Tulipa* to be sister to *Amana*, while the Western North America branch (E4) strongly supports *Erythronium* and

Table 2 Topology statistic for coalescent trees of *Erythronium*, *Amana*, and *Tulipa* (2594 OGs)

outgroup = <i>N. campanulatum</i>											
E1 = <i>E. japonicum</i> E2 = <i>E. sibiricum</i> E3 = <i>E. albidum</i> E4 = <i>E. californicum</i>											
A1 = <i>A. edulis</i> A2 = <i>A. erythronioides</i> A3 = <i>A. wanzhensis</i>											
T1 = <i>T. heterophylla</i> T2 = <i>T. heteropetala</i> T3 = <i>T. tarda</i> T4 = <i>T. montana</i> T5 = <i>T. schrenkii</i> T6 = <i>T. sinkiangensis</i>											
Triplets	Sister	Support	Triplets	Sister	Support	Triplets	Sister	Support	Triplets	Sister	Support
111	TA	0.93	211	EA	0.69	311	EA	0.96	411	EA	1
112	TA	1	212	TA	1	312	EA	0.74	412	EA	1
113	TA	1	213	TA	1	313	TA	0.61	413	EA	1
114	TA	0.97	214	TA	1	314	EA	0.64	414	EA	1
115	TA	0.98	215	TA	1	315	EA	0.99	415	EA	1
116	TA	1	216	TA	1	316	TA	0.77	416	EA	1
121	EA	1	221	EA	1	321	EA	0.93	421	EA	1
122	TA	1	222	TA	1	322	TA	0.9	422	EA	1
123	TA	1	223	TA	1	323	TA	0.64	423	EA	1
124	TA	0.96	224	TA	1	324	TA	0.98	424	EA	1
125	TA	0.66	225	TA	1	325	EA	0.52	425	EA	1
126	TA	0.9	226	TA	1	326	TA	0.52	426	EA	1
131	EA	0.68	231	TA	1	331	EA	0.98	431	EA	1
132	TA	1	232	TA	1	332	EA	0.77	432	EA	1
133	TA	0.96	233	TA	1	333	EA	0.72	433	EA	1
134	TA	1	234	TA	1	334	TA	0.79	434	EA	0.99
135	TA	0.99	235	TA	1	335	EA	0.6	435	EA	1
136	TA	1	236	TA	1	336	TA	0.55	436	EA	1

Amana as sister groups, and the Eastern North America branch (E3) supports both of them. As the basal taxon of gen. *Tulipa*, *T. heterophylla* (T1) seems to have a special status, because it supports the Eurasian branch of *Erythronium* to be sister to *Amana* in 2/3 cases.

In order to prevent the instability of phylogenetic tree reconstructed by only four species, we used *N* as an outgroup, together with both of the *Gagea* spp., A1, A3, T1-T6, *T. biflora* and *T. hissarica*, a total of 13 species to tested different combinations of E1-E4 (Table 3 and Additional file 1: File5). The result was decisive: all the trees including E1 or E2 showed sister relationship of *Amana* and *Tulipa*; However, E3, E4 and their combinations supported the sister relationship between *Erythronium* and *Amana*. Moreover, if the same taxa were reconstructed by ML method, *Amana* and *Tulipa* would be clustered into one branch with high support, which clearly showed that ILS has misled the ML inference.

Overall, the Eurasian group of *Erythronium* species possesses some unique characteristics compared to its North American counterpart. Specifically, it exhibits a disjunct distribution pattern, yet slightly overlaps with the distribution of *Amana*. In this scenario, *Erythronium*'s plastid and nuclear genes appear to have undergone different evolutionary trajectories, with the former

Table 3 Analysis of phylogenetic tree topology based on different *Erythronium* species combinations

1 = <i>E. japonicum</i> 2 = <i>E. sibiricum</i> 3 = <i>E. albidum</i> 4 = <i>E. californicum</i>			
Combinations of <i>Erythronium</i>	Method	Sister	Support
1	Coalescent	TA	1
13	Coalescent	TA	0.97
14	Coalescent	TA	0.98
134	Coalescent	TA	0.99
2	Coalescent	TA	1
23	Coalescent	TA	1
24	Coalescent	TA	0.98
234	Coalescent	TA	0.98
3	Coalescent	EA	1
4	Coalescent	EA	1
34	Coalescent	EA	0.99
3	Concatenation (ML)	TA	76
4	Concatenation (ML)	TA	96
34	Concatenation (ML)	TA	100

supporting a sister relationship with *Amana*, whereas the latter favor it as a sister to both *Tulipa* and *Amana*. Thus, better understanding its genome may be one of the keys

to resolving the phylogenetic problems among the three genera. Additionally, *E. albidum* (eastern North America) showed the characteristics of branches that are half like those of taxa from Eurasian and half like branches of Western North America taxa, which may be related to its homologous polyploidy [12]. In the reticulate evolution analysis not presented in the main text, introgression of about 50% of genetic components from the ancestor of *Erythronium* to *E. albidum* was also detected, indicating that *E. albidum* may retain a relatively ancient genome that sheds light on the phylogenetic history of this genus.

Reconstructing the evolutionary history of tribe Tulipeae

In our study, when using low-copy nuclear datasets (1594 OGs and the datasets used in pre-experiments), ML trees always conflict with coalescent trees, which can be simply explained by ILS. However, when a further 1000 OGs with slightly higher copies are added to the analyses, the support value of this node in coalescent tree decreases rapidly, and finally collapses to a sister relationship between *Amana* and *Tulipa*, albeit with poor support. We have reason to believe that although the additional 1000 OGs added later are fewer in number, they may possess stronger phylogenetic signals than the initial 1594 OGs. Moreover, due to their slightly higher copy numbers, they may evolve more quickly. We have noticed that when we add more OGs with higher copies or remove the branches of gene trees with low support values, then the support value at this node tends to increase. It can be predicted, therefore, that if whole genomes could be used in the future for reconstructing their phylogenetic relationships, then the sister relationship between *Amana* and *Tulipa* will likely emerge as true. This unusual situation is similar to a scenario discussed by Solís-Lemus et al. [24]. Under cases of high-level introgression (> 30%), ASTRAL will be more and more firmly misled by the wrong structure of introgression with the addition of more well-reconstructed gene trees. Mallet et al. [21] put forward the same view based on a case of the *Anopheles gambiae* complex: abundant introgression may lead to a “tyranny of the majority,” such that with an increasing number of genes involved, more and more gene trees that are inconsistent with the species tree will impose a quantitative advantage on the phylogenetic result. Therefore, the real history may be similar to the anomaly zone envisaged by Solís-Lemus et al. [24]. In brief, they hypothesize that when introgression occurs between two rapid differentiation events, alleles affected by the introgression from another lineage may immediately experience the second differentiation event without sorting, which makes the two descendant populations (i.e., sister taxa, like *Erythronium* and *Amana*) inherit the alleles vertically and by introgression, respectively, and

thus lead to an inconsistency between gene trees and the species tree.

With all this said, we want to emphasize that we are not arguing that TA|E must be the wrong topology for these genera. Considering the results of the Robinson-Foulds-based metric (see “Phylogenetic tree reconstruction” in the “Results” section), it is necessary to invoke introgression to account for the inconsistency between the plastid and nuclear gene topologies. Therefore, there is another possibility that *Amana*, indeed, may have experienced multiple hybridization and introgression events with *Erythronium*, eventually leading to the capture of its plastid. One reason we say this is that the coalescent tree results of 72 quartets showed that *E. californicum* (from western North America) may have an especially close nuclear relationship with *Amana* from Asia (Table 2). In fact, instances of introgression between *Amana* and North American *Erythronium* were frequently detected (Fig. 8a). In D-statistic analyses, not only *E. californicum* but also *E. japonicum* (East Asia) showed a close relationship with *Amana* (East Asia). In addition, the average divergence time for *Tulipa*, *Erythronium*, and *Amana* is about 22.82 Mya, 24.38 Mya (15.88 Mya for Eurasian branch), and 10.96 Mya respectively [6, 36]. Together these lines of evidence support one of the reasonable migration hypotheses [12]—that the ancestor of the three genera originated in the Far East near the western part of North America, with the ancestor of *Erythronium* diverging first and spreading westward to form the Eurasian branch, and eastward to form two North American branches. The ancestor of *Tulipa* also diverged around the same time, with one lineage moving westward to the Tianshan-Pamir region. Another lineage of *Tulipa* (or possibly *Erythronium*) remained in the Far East and diverged into early lineages of *Amana* during the middle to late Miocene.

Another fact to consider is that *Amana* shares several morphological similarities with *Tulipa*. Both typically have bulbous (rather than cylindric) bulbs, leaves that are usually free of pigmentation or reticulate venation, and tepals that do not recurve or reflex. These traits distinguish them from *Erythronium*. However, despite these morphological similarities with *Tulipa*, *Amana* has a geographic range more akin to that of *Erythronium* [5, 12]. Differing genomic partitions alternately correlating with morphology or geography is characteristic of many ancient plastid capture events [37, 38].

We further speculate that the historical distribution range of *Amana* may have been much broader and substantially overlapping with that of *Tulipa* and/or *Erythronium*. This is because under the influence of glaciation, lowland taxa likely underwent southward migration, range contraction, and fragmentation [39]. The migration

history of *Amana* aligns with this scenario [40]. Additionally, plastid capture has been detected in other monocotyledonous herbs (e.g., *Hemerocallis*, Asphodelaceae) [38], and so a bolder hypothesis is that the ancestor of *Amana* once inhabited a shared distribution between East Asia and North America, captured its plastid by hybridization.

Phylogenetic relationships within *Tulipa*

Following the classification framework of genus *Tulipa* proposed by Veldkamp et al. [14, 15] and Christenhusz et al. [1], our research tested the monophyly of the four currently recognized subgenera. Three of these, namely subg. *Tulipa* (including *T. sinkiangensis*), subg. *Clusianae* and subg. *Erythronium*, are strongly supported as monophyletic. The phylogenetic relationship among these three subgenera is clear, that is: (*Tulipa*, (*Eriostemon*, *Clusianae*)).

Subgenus *Orithyia*, however, is not monophyletic. This is not surprising, since there has been a lack of research on the systematics of this subgenus. Zonneveld et al. [14, 15] defined subg. *Orithyia* with a combination of three characters: a style nearly as long as the ovary, bulb tunics usually naked inside, and DNA 2C value of only 38–39 pg. They recognized three species, *T. uniflora*, *T. heterophylla*, and *T. heteropetala*, as members of subg. *Orithyia*, but did not rule out the possibility of *T. sinkiangensis* and *T. thianschanica* also being part of this subgenus. However, we now know that *T. thianschanica* actually belongs to the subg. *Tulipa*. Subsequently, Christenhusz et al. used only *T. uniflora* to represent the subgenus in their phylogenetic analysis and warned that the remaining species need to be sampled in order to evaluate the monophyly of the group [1]. Our study shows quite strikingly that *T. heterophylla* and *T. heteropetala* are successive sister species to the remainder of the genus *Tulipa*, whereas *T. sinkiangensis* is sister to subgenus *Tulipa* and might be best classified within it. Since *T. heterophylla* is similar to other *Tulipa* spp. in morphology, geographical distribution, and phylogeny, there is no sufficient reason to separate it as a monotypic genus as some have proposed. Therefore, following the study by Wilson [36], we suggest the old taxon *Eduardoregelia* Popov (which includes a single species, *Eduardoregelia heterophylla* (Regel) Popov, a synonym of *T. heterophylla*) should be redefined as a new subgenus of *Tulipa*. As for *T. heteropetala* and *T. uniflora*, we only sampled the former, but they share a highly similar morphology, and considering that a previous study has shown their clustering together [6, 36], it is reasonable to continue recognizing them within a now smaller circumscribed subg. *Orithyia*.

In the small subg. *Clusianae*, we reconstructed a topology different from Christenhusz et al. [1]. Here we find that *T. clusiana* is sister to *T. montana* and

T. linifolia. An introgression event, where *T. liniflora* acquired more than 20% of its genetic components from the ancestor of subg. *Clusianae* may be reasonable for this conflict (Fig. 8c). Additionally, our plastid phylogenetic tree indicates that *T. montana* is the basal taxon of subg. *Clusianae* (Fig. 2), conflicting with results based on nuclear genes. Wilson's study also demonstrates the same scenario [36], suggesting potential discordance in the evolutionary history between the plastid and nuclear genomes of *T. montana*.

In subg. *Eriostemon*, we found that *T. turkestanica* is embedded within *T. biflora* (2, 3), and they are supported to form a polytomy (Fig. 8). Research also indicates that they form a complex characterized by white flowers [36]. Furthermore, three species, *T. patens*, *T. dasystemon*, and *T. sprengeri* have unstable phylogenetic positions, with some of the reasons explained by the results of reticulate evolution analysis in the *Polytomy tests and reticulate evolution* in “Results” section (Fig. 8d). Based on the plastid phylogenetic tree (Fig. 1), the *T. patans* sample from NCBI separates from our *T. patans* sample, but its position aligns with Wilson et al.'s study [36]. This suggests that the species referred to as “*Tulipa patens*” may not constitute a natural group, although more data are needed to confirm this. *T. dasystemon* exhibits cytonuclear conflict, a scenario consistent with findings in Wilson's study. Therefore, it is also possible that the nuclear and chloroplast genomes of *T. dasystemon* have different evolutionary history.

As mentioned earlier, we consider *T. sinkiangensis* to be a member of subg. *Tulipa*, even though *T. sinkiangensis* has a style (an original plesiomorphy). Although many phylogenetic trees in our study supported this anomalous species to be the basal taxon of Clade T1 (Figs. 3 and 4), our plastid phylogenetic tree and the reticulate evolution analysis support it as the sister taxon of the entire subg. *Tulipa* (Figs. 2 and 8e). Some other results that stand out within the topology of subg. *Tulipa* are that *T. thianschanica* is sister to *T. iliensis* and *T. tetraphylla*, rather than *T. thianschanica* var. *sailimuensis* (Figs. 3 and 4). For this reason, we feel that *T. thianschanica* should not be treated as a synonym of *T. iliensis* and that *T. thianschanica* var. *sailimuensis* should be treated as a distinct species rather than a variety of *T. thianschanica* [1]. However, it is important to be cautious, the sample of *T. thianschanica* from NCBI shows a closer relationship with *T. tetraphylla* (Fig. 1), indicating a potential for *T. thianschanica* to be part of species complexes. Additionally, we can safely conclude that the *T. iliensis* sample from NCBI has been misidentified; it should actually be classified as *T. altaica*. Such identification errors seem to be quite common in subg. *Tulipa* [36, 41].

For Clade T2 (Fig. 8f), reticulate evolution analysis suggests that *T. armena* is more likely the sister taxon of Clade acu. We also found that *T. praecox* and *T. acuminata* are not closely related, so at least one of them cannot be treated as a synonym of *T. agenensis*. We wish to point out that because we were unable to sample *T. undulatifolia*, it is still unclear whether *T. eichleri* is a synonym of that species.

Finally, we also evaluated the monophyly of the various sections classified within the larger subgenera, but most of these, except for sect. Neo-tulipae, *Saxatiles* (embedded in sect. *Sylvestres*), *Multiflorae* and *Spiranthera* (only one sample), were found to be not monophyletic. To highlight a few examples, *T. tarda* and *T. urumienensis*, which belong to sect. *Biflores*, are embedded in sect. *Sylvestres*; the species of sect. *Kolpakowskianae* are scattered across three clades; species of sect. *Vinistriatae* and sect. *Tulipanum* are mixed together; and sect. *Tulipa* is divided between two distant clades. Previous studies have delved deeper into the sections of tulips and offered specific recommendations [42]. Further examination of morphological, genetic, and other characters in light of these phylogenetic results is warranted in order to better classify the horticulturally popular tulips.

Conclusions

In this study, we newly sequenced 50 transcriptomes of taxa from tribe Tulipeae (targeting all four genera and most of their major subclades) and one outgroup species from tribe Lilieae (*Notholirion campanulatum*). Using an improved method to splice the transcript fragments, 2594 nuclear OGs and 74 plastid PCGs were employed to reconstruct phylogenetic relationships within the tribe, which confirmed the monophyly of the genera and most of the subgenera in *Tulipa* (except *Orithyia*), but indicated that most sections within the *Tulipa* subgenera are not monophyletic. The phylogenetic positions of specific problematic taxa has been discussed in detail.

Furthermore, by measuring sCF/sDF1/sDF2 values, rampant introgression and high levels of ILS were detected, which resulted in strong incongruency among the relationships inferred from different datasets and methods (ML vs. MSC). Reasonable explanations to account for most of these conflicts were found through reticulate evolutionary analyses and polytomy tests. However, we still failed to reconstruct with confidence the evolutionary history among *Amana*, *Erythronium*, and *Tulipa* due to pervasive ILS and introgression. For this reason, we offer two hypotheses: (1) under high-level ILS, introgression appears to be strong enough to have created a “tyranny of the majority” in the ASTRAL analysis, which has been misled to favor a sister relationship between *Amana* and *Tulipa*; or (2) ancient plastid

capture occurred among these genera, leading to support for *Amana* being sister to *Erythronium*, in conflict with nuclear analyses. Further investigations into plastid genome or chromosome structure will be required to verify which of these two topologies best represents the history of the majority of the genome of these plants.

We argue that the Eurasian group of *Erythronium* species, in particular, and tribal patterns of introgression that may involve *Gagea* are worthy of further study. Likewise, patterns of polyploidization, which were not considered in our study, likely played an important role in the speciation of *Tulipa* [15]; it should also be the topic of further investigation within a phylogenomic context.

In light of our findings, we encourage others to explore the potential of transcriptomics as an alternative method for studying plant evolution, especially in taxa with unusually large genomes. Transcriptomics offers a cost-effective and efficient means of obtaining high-quality genetic data, making it a valuable tool for studying plant evolution.

Methods

Plant material and sequencing

Fifty transcriptomes were newly sequenced (Additional file 2: Table S1), including 46 species of tribe Tulipeae and representing all four genera, i.e., *Gagea* (2 spp.), *Amana* (3 spp., both of the two major clades [5]), *Erythronium* (3 spp., all three major clades [12]), and *Tulipa* (38 spp., 41 accessions, all four subgenera [1]), plus one outgroup species of tribe Lilieae (*Notholirion campanulatum*). Twenty of those species were field-collected in China (17 spp.) and the USA (2 spp.) in early spring during anthesis, while the others were collected in the Botanical Garden of the University of Fribourg (BGUF) and the Royal Botanic Garden Edinburgh (RBGE).

We extracted total RNA from silica gel-dried or fresh leaves, using an improved CTAB protocol [43]. Sequencing was done by the Beijing Genomics Institute (BGI, Shenzhen, China), Novogene (Beijing, China), and Berry genomics (Beijing, China) on an Illumina HiSeq2500 platform with raw reads-lengths of 150 bp. Although we used both dry and fresh leaf material for RNA isolation, our results and a recent study [16] showed that silica gel-dried leaf samples did not affect the quality of the transcriptome when compared to fresh materials.

Transcriptome data processing and orthology inference

First, we used BOWTIE2 v.2.44 [44] to filter organelle genes (just plastid genes here, as no mitochondrial genes were found in our data) in the quality-controlled transcriptome data of all 50 newly sequenced samples. The reference index (database used by BOWTIE2 for comparison) was generated from 36 whole plastomes of

tribe Tulipeae downloaded from NCBI (Additional file 2: Table S4). Finally, 50 transcriptomes were divided into plastid vs. nuclear datasets.

Because our new data did not yield complete plastomes for some species, we also downloaded 14 complete plastomes of Tulipeae (*Gagea*: 1 sp., *Amana*: 3 spp., *Erythronium*: 2 spp. and *Tulipa*: 7 spp.) and Liliaeae (only *Notholirion campanulatum*) from the National Center for Biotechnology Information (NCBI; <http://www.ncbi.nlm.nih.gov>) to assess whether the phylogenetic signals provided by these incomplete plastomes were acceptable. In addition, we obtained nuclear transcriptome data of *Erythronium japonicum* (SRR14576607) from NCBI to complement our phylogenetic analyses with a species representing the genus' Eurasian clade [12] (Additional file 2: Table S1).

All reads in the plastid dataset were assembled into contigs by GENEIOUS PRIME v.2020.2.4 [45] using default settings, and the contigs were then mapped to an annotated reference plastome (*Tulipa thianschanica*, MW077738). After realigning the regions with low reliability and obvious mismatch manually, we extracted protein-coding genes (PCGs) from the 50 samples together with 14 supplementary plastomes. After being aligned by translation method in GENEIOUS to avoid frameshift mutations, all the PCGs were further filtered of terminator and initiator errors, large deletions, and low sample coverage (< 50%). A total of 74 PCGs were finally selected as the plastid dataset for subsequent analysis (Additional file 2: Table S4).

All 51 nuclear transcriptomes were assembled de novo in TRINITY v.2.5.1 [46] (default parameters). Open reading frames (ORFs) were predicted by TRANSDCODER v.5.5.0 [47]. After removing redundant sequences of ORFs by CD-HIT v.4.8.1 [48] (parameters -c 0.99 and -n 10), ORTHOFINDER v.2.5.4 [49] was used with the parameters -d and -S blast_nucl (calling nucleotide BLAST) to detect orthologous genes (OGs). Finally, 10,851 candidate OGs were selected from 45,763 OGs by applying the criteria of mode, mean, and median copy numbers ≤ 3 , and variance ≤ 9 , while allowing a species missing rate of approximately 30% (i.e., the number of OGs with 0 copies is ≤ 16) (Additional file 2: Table S5). We found that species with multiple copies of an OG often contained several fragments of one transcript, which may be caused by RNA degradation or alternative splicing. We therefore used a script (Additional file 3: File S6) to concatenate these fragments, so that all candidate OGs will become single-copy OGs corresponding to the specific samples. All 10,851 OGs were aligned in MAFFT v.7.508 [50] (L-INS-i [51]) and trimmed by TRIMAL v.1.4.1 [52], which removed suspicious sequences (parameters -recoverap 0.75, -seqoverlap 80 and -strictplus). This

dataset was further reduced by filtering out OGs with low identical sites (%) in GENEIOUS PRIME. After removing all gaps by GBLOCKS v.0.91b [53], all the OGs were improved by manually deleting the regions containing obvious errors. Finally, OGs with less than 35 sequences (< 70% accessions) were removed once again, resulting in a dataset of 2594 OGs. Then, according to 2594 OGs' average copy number (before script processing), two sub-datasets, consisting of 1594 OGs (relatively low copy) and 1000 OGs (relatively high copy), were built up respectively. In addition, 2032 of 2594 OGs were also selected to build a dataset, as being considered suitable for phylogenetic analysis through the matched-pair tests of symmetry [54] performed by IQ-TREE2 [55]. In short, we created four nuclear datasets of 2594/2032/1594/1000 OGs.

Phylogenetic tree reconstruction

We used the R package PHYTOOLS 0.2.2 [56] to generate partition files (74 genes $\times$ three codon positions, 222 partitions in total) for the plastid dataset, and 2594/1594/1000 gene positions (without using three codon positions) for three nuclear datasets. ML trees were inferred by IQ-TREE2 v.2.03 [55] with 1000 bootstrap replicates for the plastid dataset and 1000 ultrafast bootstrap replicates for four nuclear datasets under the optimal substitution models as selected by the parameter -m MFP + MERGE. Moreover, SH-like approximate likelihood ratio test (SH-aLRT) was also performed with 1000 replicates for every ML trees.

RAXML-NG v.1.1 [41] was used to infer gene trees for 2594 orthogroups (OGs) under the GTR + I + G model with 200 bootstrap replicates. These gene trees were then resampled into four datasets based on subsets of 2594, 2032, 1594, and 1000 OGs. Clades with bootstrap values (BS) lower than 10 (or 20/30/40/50) were pruned using NEWICK UTILITIES v.1.6 [57]. After pruning, the resulting gene tree datasets were used as input for reconstructing coalescent-based species trees in ASTRAL v.5.7.8 [23]. The final trees were visualized using FIGTREE v.1.4.4 (Available from: <http://tree.bio.ed.ac.uk/software/figtree/>)

Problematic taxa identification and polytomy test

After comparing the phylogenetic trees reconstructed from different nuclear datasets, two representative trees, a ML tree (2594 OGs) and a coalescent-based tree (1594 gene trees with BS value ≥ 10 following ASTRAL documentation), were selected for subsequent analyses. IQ-TREE2 was used for calculating "site con/discordance factors" (sCF and sDF1/sDF2) [58] based on parsimony-informative sites in the dataset of 2594 OGs, each node

of the two trees was sampled 1000 times with parameter “–scf 1000.”

As sDF1/2 actually represent the proportion of two incongruent gene trees (IGTs) in a quartet (detailed below, “Distinguishing between ILS and introgression”), the nodes which may be affected by ILS and introgression were then determined from the high/imbalanced values of sDF1/2 respectively. We focused on selecting samples near these nodes as being problematic taxa with complex evolutionary histories.

In addition, polytomy tests were carried out in ASTRAL using the coalescent-based tree to verify whether polytomies are a reasonable explanation for the problematic taxa. Following Sayyari and Mirarab [42] and Dong et al. [59], the nodes with P -value > 0.05 were considered to form polytomies.

Distinguishing between ILS and introgression

We used a coalescent simulation approach based on ILS expectations to ask whether cytonuclear discord is best explained by ILS or hybridization. We first generated gene tree simulations under the multispecies coalescent under the species tree in order to generate gene tree distributions, which quantify expected incomplete lineage sorting (ILS). The simulations used the ASTRAL trees generated with BS10-filtered gene trees and both the 1000OG and 1594OG gene sets; species trees were scaled by 2 to account for expected plastid genome N_e (given matrilineal inheritance but both parents yielding offspring [60]). Simulations used Dendropy (https://github.com/ryanafolk/tree_utilities/blob/master/simulate_gene_trees.py). We then compared the simulated gene tree sets with the empirical plastid tree. The comparison was based on calculating pairwise Robinson-Foulds distances for all simulated gene trees, which provided a null distribution of discord. We then compared this null distribution to the actual distances between the observed plastid tree and the simulated trees.

Beyond cytonuclear discordance, notable disparities emerge in phylogenetic trees constructed through coalescent analyses and concatenation analyses based on nuclear gene datasets. In order to determine whether ILS or introgression is the main cause of taxa identified as problematic, we first grouped the problematic taxa into five different groups according to clade (viz. tr. Tulipeae, subg. *Clusianae*, subg. *Eriostemon* and two clades of subg. *Tulipa*) and inferred instances of reticulate evolutionary relationships by using the maximum pseudolikelihood (MPL) method in PHYLONETWORKS V0.15.3 [31]. As a starting tree for network searches, we subsetted samples according to the five groups and deleted OGs with more than 30% of sample missing rate as a result of subsampling, after which the remaining OGs were used

for a focused re-inference of the coalescent species tree following methods above. We then performed network searches by allowing zero to five maximum number of hybridizations and 10 independent searches for each reticulation setting to avoid local optima. After the first run, some networks with low scores (proportional to the network's pseudo-deviance: the lower, the better) were selected as “guide trees” to replace the original ASTRAL coalescent tree and start a new run with a larger maximum number of hybridizations. All networks were visualized using PHYLOPLOTS.

Because the reconstructed phylogenetic tree is based on different datasets and our results were inconsistent regarding the sister group of *Amana* (either *Tulipa* or *Erythronium*, Figs. 2, 3, and 4 and Additional file 2: File S1), and also because neither *Erythronium* nor *Tulipa* was strongly supported as the sister of *Amana* in reticulate evolution analyses (Figs. 8a, b), we employed D-statistics and internal branch lengths test implemented in the programs DFOIL [61] and QuIBL [30], respectively, to further explore the major cause of this unclear phylogenetic relationship.

D-statistics are a site-based hybridization detection method that rely on specific patterns of discordant topologies created by ILS, which are distorted by imbalances when introgression is present. To explain, when considering a species tree, either in the form of quartets (unrooted trees over four taxa contained in the species tree) or rooted triplets (rooted tree over three taxa contained in the species tree), there are three possible gene trees in total: one congruent gene tree (CGT, the topology consistent with species tree) and two incongruent gene trees (IGTs). The presence of ILS will lead to an equal frequency of the two IGTs, while introgression always break this balance because of its directionality [62], but also creating the basis for a probabilistic test for hybridization. D-statistics in DFOIL uses chi-square tests of significance, taking P -value < 0.01 (instead of Z -score > 3) as the cut-off value to reject the null hypothesis (no introgression) [62].

QuIBL takes advantage of branch length expectations under ILS rather than topological expectations. Gene trees evolving under ILS are expected to have branch lengths conforming to an exponential distribution, whereas a subset of gene trees subject to introgression will violate this distribution. QuIBL detects whether the distribution of IGTs' internal branch lengths conforms to the one-parameter (i.e., effective population size, N_e) exponential distribution, with rejection of the hypothesis of exponential branch length distribution suggesting that introgression may be present. The Bayesian information criterion (BIC) was calculated for ILS-distribution (BIC1Dist) and ILS

+ introgression-distribution (BIC2Dist), following the author, we used ΔBIC ($\text{BIC2Dist} - \text{BIC1Dist}$) > 10 as cut-off to accept introgression [30]. As DFOIL and QuIBL can only process a rooted triplet in each run, we prepared a set of subsampled analyses, including (1) *Erythronium* (4 spp.); (2) *Amana* (3 spp.); and (3) six species representing all four subgenera of *Tulipa* (i.e., *T. heteropetala*, *T. heterophylla*, *T. montana*, *T. patens*, *T. sinkiangensis*, and *T. schrenkii*) for individual runs ($4 \times 3 \times 6 = 72$ in total). For each subsampled analysis, *Notholirion campanulatum* was used a fixed outgroup for rooting, while one *Erythronium* sp. and *Amana* sp. was taken as a first differentiated clade, and one species of *Tulipa* was taken as a later differentiated clade, these four samples forming a rooted triple. A sequence file comprising 2594 concatenated OGs was used directly as input for DFOIL; gene tree files were inferred by IQTREE2 as input for QuIBL.

Abbreviations

AGTs	Anomalous gene trees
BGI	Beijing Genomics Institute
BGUF	Botanical Garden of the University of Fribourg
BIC	Bayesian information criterion
CGT	Congruent gene trees
IGTs	Incongruent gene trees
ILS	Incomplete lineage sorting
LPP	Local posterior probabilities
ML	Maximum likelihood
MP	Maximum parsimony
MPL	Maximum pseudolikelihood
MSC	Multi-species coalescent
nrITS	Nuclear ribosomal internal transcribed spacer
OGs	Orthologous genes
ORFs	Open reading frames
PCGs	Plastid protein-coding genes
QTP	Qinghai-Tibet Plateau
RBGE	Royal Botanic Garden Edinburgh
sCF	Site concordance factors
sDF	Site discordance factors
UF	Ultrafast bootstrap
WGS	Whole-genome-sequencing

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-025-02204-z>.

Supplementary Material 1. Phylogenetic trees: (1) File S1-all nuclear and plastid trees; (2) File S2-sCF gCF trees; (3) File S3-Polytomy_15940G.tre; (4) File S4-Quartet trees. tre; (5) File S5-Different combinations of E1234_trees.

Supplementary Material 2. Tables. (1) Table S1_samples.xlsx; (2) Table S2_DFOIL.xlsx; (3) Table S3_QuIBL.xlsx; (4) Table S4 plastid reference and 74PCGs.xlsx; (5) Table S5_nuclear orthologous genes.xlsx.

Supplementary Material 3. File S6_script. A script to concatenate OGs' fragments.

Acknowledgements

We sincerely thank Botanical Garden of the University of Fribourg (BGUF) and the Royal Botanic Garden Edinburgh (RBGE) for providing research materials. We would also like to express our sincere gratitude to Xiang-Yu Zhou for granting us permission to use their plant photographs in this paper.

Author contributions

PL, ZPY, XZ, YGS and DAY conducted the sampling. ZHZ, MZW and RAF analyzed the data. ZHZ, MZW and HPC wrote the manuscript. XZ prepared the photo plate. MZW, HPC, RAF and KMC revised the manuscript. All the authors approved the final manuscript.

Funding

This research was supported by the National Natural Science Foundation of China (Grant No. 31970225) and Key Technology Research and Development Program of Zhejiang Province (Grant No. 2023 C03138).

Data availability

All data generated or analysed during this study are included in this published article, its supplementary information files and publicly available repositories. Fifty newly sequenced transcriptomic data have been deposited in the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under the primary accession number PRJNA1160175. All other datasets used in this study are publicly available on NCBI, with their respective accession numbers provided in Additional file 2: Table S1.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Systematic & Evolutionary Botany and Biodiversity Group, State Key Laboratory for Vegetation Structure, Function and Construction, College of Life Sciences, Zhejiang University, Hangzhou 310058, China. ²College of Biology and Environmental Engineering, Zhejiang Shuren University, Hangzhou, Zhejiang, China. ³College of Life Sciences and Technologies, Tarim University, Alar 843300, China. ⁴Department of Environment & Biodiversity, University of Salzburg, Salzburg, Austria. ⁵Shanghai Chenshan Botanical Garden, Shanghai 201602, China. ⁶Department of Biological Sciences, Mississippi State University, Starkville, MS, USA. ⁷California Academy of Sciences, San Francisco, CA, USA. ⁸Department of Botany, University of Wisconsin, Madison, WI 53706, USA.

Received: 31 August 2024 Accepted: 2 April 2025

Published online: 28 April 2025

References

- Christenhusz MJM, Govaerts R, David JC, Hall T, Borland K, Roberts PS, Tuomisto A, Buerki S, Chase MW, Fay MF. Tiptoe through the tulips - cultural history, molecular phylogenetics and classification of *Tulipa* (Liliaceae). Bot J Linn Soc. 2013;172(3):280–328.
- Tan D. The systematics of *Tulipa* L.(SL) from China[in Chinese]. Chinese Academy of Sciences. 2005.
- Chao C, Tzeng H, Tseng Y. On the Review of Taxonomic Status of Liliaceae. Quarterly Journal of Forest Research. 2015;37:143–60.

4. Chase MW, Christenhusz MJM, Fay MF, Byng JW, Judd WS, Soltis DE, Mabblerley DJ, Sennikov AN, Soltis PS, Stevens PF, et al. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: Apg IV. Bot J Linn Soc. 2016;181(1):1–20.
5. Li P, Lu R, Xu W, Ohi-Toma T, Cai M, Qiu Y, Cameron KM, Fu C. Comparative Genomics and Phylogenomics of East Asian Tulips (*Amana*, Liliaceae). Front Plant Sci. 2017;8:451.
6. Kim JS, Kim J. Updated molecular phylogenetic analysis, dating and biogeographical history of the lily family (Liliaceae: Liliales). Bot J Linn Soc. 2018;187(4):579–93.
7. Lu R, Yang T, Chen Y, Wang S, Cai M, Cameron KM, Li P, Fu C. Comparative plastome genomics and phylogenetic analyses of Liliaceae. Bot J Linn Soc. 2021;196(3):279–93.
8. Hayashi K, Kawano S. 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 Spec Biol. 2000;15(1):73–93.
9. Allen GA, Soltis DE, Soltis PS. Phylogeny and Biogeography of *Erythronium* (Liliaceae) Inferred from Chloroplast matK and Nuclear rDNA ITS Sequences. Syst Bot. 2003;28(3):512–23.
10. Rønsted N, Law S, Thornton H, Fay MF, Chase MW. Molecular phylogenetic evidence for the monophyly of *Fritillaria* and *Lilium* (Liliaceae; Liliales) and the infrageneric classification of *Fritillaria*. Mol Phylogenet Evol. 2005;35(3):509–27.
11. Zarrei M, Wilkin P, Fay MF, Ingrouille MJ, Zarre S, Chase MW. Molecular systematics of *Gagea* and *Lloydia* (Liliaceae; Liliales): implications of analyses of nuclear ribosomal and plastid DNA sequences for infrageneric classification. Ann Bot-London. 2009;104(1):125–42.
12. Clennett JCB, Chase MW, Forest F, Maurin O, Wilkin P. Phylogenetic systematics of *Erythronium* (Liliaceae): morphological and molecular analyses. Bot J Linn Soc. 2012;170(4):504–28.
13. Kim JS, Hong J, Chase MW, Fay MF, Kim J. Familial relationships of the monocot order Liliales based on a molecular phylogenetic analysis using four plastid loci: *matK*, *rbcL*, *atpB* and *atpF-H*. BOT J LINN SOC. 2013;172(1):5–21.
14. Veldkamp JF, Zonneveld BJM. The infrageneric nomenclature of *Tulipa* (Liliaceae). Plant Syst Evol. 2012;298(1):87–92.
15. Zonneveld BJM. The systematic value of nuclear genome size for “all” species of *Tulipa* L. (Liliaceae). Plant Syst Evol. 2009;281(1):217–45.
16. He J, Lyu R, Luo Y, Xiao J, Xie L, Wen J, Li W, Pei L, Cheng J. A phylo-transcriptome study using silica gel-dried leaf tissues produces an updated robust phylogeny of Ranunculaceae. Mol Phylogenet Evol. 2022;174:107545.
17. McKain MR, Johnson MG, Uribe-Convers S, Eaton D, Yang Y. Practical considerations for plant phylogenomics. APPL PLANT SCI. 2018;6(3): e1038.
18. Yang Y, Moore MJ, Brockington SF, Soltis DE, Wong GK, Carpenter EJ, Zhang Y, Chen L, Yan Z, Xie Y, et al. Dissecting Molecular Evolution in the Highly Diverse Plant Clade Caryophyllales Using Transcriptome Sequencing. Mol Biol Evol. 2015;32(8):2001–14.
19. Shi C, Wang S, Xia E, Jiang J, Zeng F, Gao L. Full transcription of the chloroplast genome in photosynthetic eukaryotes. Sci Rep-UK. 2016;6(1):30135.
20. Shen X, Steenwyk JL, Rokas A. Dissecting Incongruence between Concatenation- and Quartet-Based Approaches in Phylogenomic Data. Syst Biol. 2021;70(5):997–1014.
21. Mallet J, Besansky N, Hahn MW. How reticulated are species? BioEssays. 2016;38(2):140–9.
22. Li J, Cai J, Qin H, Price M, Zhang Z, Yu Y, Xie D, He X, Zhou S, Gao X: Phylogeny, Age, and Evolution of Tribe Liliae (Liliaceae) Based on Whole Plastid Genomes. Front Plant Sci. 2022;12:699226.
23. Mirarab S, Warnow T. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics. 2015;31(12):i44–52.
24. Solís-Lemus C, Yang M, Ané C. Inconsistency of Species Tree Methods under Gene Flow. Syst Biol. 2016;65(5):843–51.
25. Kubatko LS, Degnan JH. Inconsistency of phylogenetic estimates from concatenated data under coalescence. SYST BIOL. 2007;56(1):17–24.
26. Ewing GB, Ebersberger I, Schmidt HA, von Haeseler A. Rooted triple consensus and anomalous gene trees. BMC EVOL BIOL. 2008;8:118.
27. Mallet J. Hybridization as an invasion of the genome. Trends in ecology & evolution (Amsterdam). 2005;20(5):229–37.
28. Hibbins MS, Hahn MW. Phylogenomic approaches to detecting and characterizing introgression. Genetics. 2022;220(2):iyab173.
29. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, et al. A Draft Sequence of the Neandertal Genome. Science. 2010;328(5979):710–22.
30. Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, García-Accinelli G, Van Belleghem SM, Patterson N, Neafsey DE, et al. Genomic architecture and introgression shape a butterfly radiation. Science. 2019;366(6465):594–9.
31. Solís-Lemus C, Bastide P, Ané C. PhyloNetworks: A Package for Phylogenetic Networks. MOL BIOL EVOL. 2017;34(12):3292–8.
32. Rabier C, Berry V, Stoltz M, Santos JD, Wang W, Glaszmann J, Pardi F, Scornavacca C, Kosakovsky Pond SL. On the inference of complex phylogenetic networks by Markov Chain Monte-Carlo. PLOS COMPUT BIOL. 2021;17(9): e1008380.
33. Hibbins MS, Hahn MW. Distinguishing between histories of speciation and introgression using genomic data. BSSB. 2024;3(1).
34. Pang X, Zhang D. Impact of Ghost Introgression on Coalescent-Based Species Tree Inference and Estimation of Divergence Time. Syst Biol. 2023;72(1):35–49.
35. Yu Y, Nakhleh L. A maximum pseudo-likelihood approach for phylogenetic networks. BMC Genomics. 2015;16:1–10.
36. Wilson B. *Tulipa*: The taxonomy and evolutionary history of the genus and its impact on conservation priorities in Central Asia. University of Cambridge. 2023.
37. Yang Y, Qu X, Zhang R, Stull GW, Yi T. Plastid phylogenomic analyses of Fagales reveal signatures of conflict and ancient chloroplast capture. MOL PHYLOGENET EVOL. 2021;163: 107232.
38. Hirota SK, Yasumoto AA, Nitta K, Tagane M, Miki N, Suyama Y, Yahara T. Evolutionary history of *Hemerocallis* in Japan inferred from chloroplast and nuclear phylogenies and levels of interspecific gene flow. Mol Phylogenet Evol. 2021;164:107264.
39. Folk RA, Soltis PS, Soltis DE, Guralnick R. New prospects in the detection and comparative analysis of hybridization in the tree of life. AM J BOT. 2018;105(3):364–75.
40. Wu J, Wang M, Zhu Z, Cai M, Lee J, Li P. Cytogeography of the East Asian Tulips (*Amana*, Liliaceae). Taxonomy (Basel, Switzerland). 2022;2(1):145–59.
41. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019;35(21):4453–5.
42. Sayyari E, Mirarab S. Testing for Polytomies in Phylogenetic Species Trees Using Quartet Frequencies. GENES-BASEL. 2018;9(3):132.
43. Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. NUCLEIC ACIDS RES. 1980;8(19):4321–6.
44. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. NAT METHODS. 2012;9(4):357–9.
45. Kearsse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647–9.
46. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. NAT PROTOC. 2013;8(8):1494–512.
47. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. NAT BIOTECHNOL. 2011;29(7):644–52.
48. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22(13):1658–9.
49. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015;16:1–14.
50. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
51. Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. BRIEF BIOINFORM. 2008;9(4):286–98.

52. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25(15):1972–3.
53. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol*. 2007;56(4):564–77.
54. Naser-Khdour S, Minh BQ, Zhang W, Stone EA, Lanfear R, Bryant D. The prevalence and impact of model violations in phylogenetic analysis. *Genome Biol Evol*. 2019;11(12):3341–52.
55. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020;37(5):1530–4.
56. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). *METHODS ECOL EVOL*. 2012;3(2):217–23.
57. Junier T, Zdobnov EM. The Newick utilities: high-throughput phylogenetic tree processing in the Unix shell. *Bioinformatics*. 2010;26(13):1669–70.
58. Minh BQ, Hahn MW, Lanfear R. New methods to calculate concordance factors for phylogenomic datasets. *Mol Biol Evol*. 2020;37(9):2727–33.
59. Dong W, Li E, Liu Y, Xu C, Wang Y, Liu K, Cui X, Sun J, Suo Z, Zhang Z, et al. Phylogenomic approaches untangle early divergences and complex diversifications of the olive plant family. *BMC BIOL*. 2022;20(1):92.
60. Joly S. JML: testing hybridization from species trees. *Mol Ecol Resour*. 2012;12(1):179–84.
61. Pease JB, Hahn MW. Detection and polarization of introgression in a five-taxon phylogeny. *Syst Biol*. 2015;64(4):651–62.
62. Zheng Y, Janke A. Gene flow analysis method, the D-statistic, is robust in a wide parameter space. *BMC Bioinformatics*. 2018;19(1):10.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.