ORIGINAL ARTICLE OPEN ACCESS

Genealogical Analysis of Replicate Flower Colour Hybrid Zones in *Antirrhinum*

Arka Pal[1] | Daria Shipilina[1] | Alan Le Moan[2] | Adrian J. McNairn[3] | Jennifer K. Grenier[3] | Marek Kucka[4,5] | Graham Coop[6] | Yingguang Frank Chan[4,7] | Nicholas H. Barton[1] | David L. Field[8] | Sean Stankowski[1,9]

[1]Institute of Science and Technology Austria, Klosterneuburg, Austria | [2]UMR 7144 AD2M, CNRS-Sorbonne Université, Station Biologique de Roscoff, Roscoff, France | [3]Genomics Innovation Hub and TREx Facility, Cornell University, Ithaca, New York, USA | [4]Friedrich Miescher Laboratory of the Max Planck Society, Tübingen, Germany | [5]Department of Translational Genomics, University of Cologne, Cologne, Germany | [6]Department of Evolution & Ecology and Center for Population Biology, University of California – Davis, Davis, California, USA | [7]Groningen Institute for Evolutionary Life Sciences (GELIFES), university of Groningen, Groningen, the Netherlands | [8]Applied BioSciences, Macquarie University, Sydney, New South Wales, Australia | [9]Department of Ecology and Evolution, The University of Sussex, Brighton, UK

Correspondence: Arka Pal (arka.pal@ist.ac.at)

ABSTRACT

A major goal of speciation research is identifying loci that underpin barriers to gene flow. Population genomics takes a 'bottom-up' approach, scanning the genome for molecular signatures of processes that drive or maintain divergence. However, interpreting the 'genomic landscape' of speciation is complicated, because genome scans conflate multiple processes, most of which are not informative about gene flow. However, studying replicated population contrasts, including multiple incidences of secondary contact, can strengthen inferences. In this paper, we use linked-read sequencing (haplotagging), F_{ST} scans and genealogical methods to characterise the genomic landscape associated with replicate hybrid zone formation. We studied two flower colour varieties of the common snapdragon, *Antirrhinum majus* subspecies *majus*, that form secondary hybrid zones in multiple independent valleys in the Pyrenees. Consistent with past work, we found very low differentiation at one well-studied zone (Planoles). However, at a second zone (Avellanet), we found stronger differentiation and greater heterogeneity, which we argue is due to differences in the amount of introgression following secondary contact. Topology weighting of genealogical trees identified loci where haplotype diversity was associated with the two snapdragon varieties. Two of the strongest associations were at previously identified flower colour loci: *Flavia*, that affects yellow pigmentation, and *Rosea/Eluta*, two linked loci that affect magenta pigmentation. Preliminary analysis of coalescence times provides additional evidence for selective sweeps at these loci and barriers to gene flow. Our study highlights the impact of demographic history on the differentiation landscape, emphasising the need to distinguish between historical divergence and recent introgression.

1 | Introduction

A major goal of speciation research is to identify loci underlying barriers to gene flow. Population genomic studies usually take a 'bottom-up' approach by scanning the genome for patterns of within- and between-population variation that indicate selection driving or maintaining divergence (Ravinet et al. 2017; Wolf and Ellegren 2017). For example, during speciation with gene flow, genomic regions associated with local adaptation or genetic incompatibilities are expected to show elevated genetic differentiation (usually measured by F_{ST}), with the rest of the genome homogenised through genetic exchange (Feder et al. 2012; Wu 2001). Indeed, numerous studies of the 'genomic landscape' have found highly heterogenous patterns of genetic differentiation and, in some cases, have shown that regions with high F_{ST} house genes underpinning adaptive traits that also act as reproductive barriers (Hooper et al. 2024; Martin et al. 2013; Poelstra et al. 2014; Todesco et al. 2020). However, we now know that interpreting the differentiation landscape is more challenging than some researchers once hoped (Ravinet et al. 2017; Wolf and Ellegren 2017).

The main challenge is that genome scans can conflate multiple processes, some of which are not directly relevant to current heterogenous gene flow (Ravinet et al. 2017). Consider a simple model of secondary contact, where divergence builds up over a long period (which may involve intermittent isolation), and erodes following recent contact. Genome-wide divergence builds up relatively slowly, due to both drift and selection. Divergence will inevitably be heterogenous along the genome both by chance and due to intrinsic properties of the genome, such as the local density of functional elements and local recombination rate (Burri 2017). After contact, introgression will erode divergence where the populations meet, potentially revealing the location of barrier loci (Duranton et al. 2018). While relatively fast compared with the build-up of divergence, this erosion takes some time and will be delayed if interbreeding is geographically localised (Barton and Gale 1993). Thus, genome scans reflect both initial divergence and postcontact introgression, and these may be hard to disentangle.

Inclusion of replicate hybrid zones aids the interpretation of genome scans, allowing comparison of divergence across multiple contacts (Nadeau et al. 2014; Rancilhac et al. 2024; Vijay et al. 2016; Wilding et al. 2001). Overall divergence may reflect differences in the timing of contact or rates of gene flow. Nevertheless, parallel contacts should ultimately lead to similar differentiation landscapes if large-effect outlier loci reflect barriers to gene flow that have resisted introgression in each location. In contrast, outliers found in a single zone might reflect local demographic processes (e.g., bottlenecks), evolutionary noise, sampling effects or population-specific barriers (Westram et al. 2021). Several studies have used this logic to identify loci that underpin local adaptation and speciation. The most compelling studies combine traditional site-based genome scans with tree-based methods, which make it possible to analyse more than two populations within a single framework that acknowledges their recent shared history (Poelstra et al. 2014; Rancilhac et al. 2024).

In this paper, we study genome-wide variation associated with replicate hybrid zones in the common snapdragon, *Antirrhinum majus*, a classic model for understanding phenotypic variation both in the laboratory and in nature (Hudson et al. 2008). We focus on two varieties of *A. majus* subspecies *majus*—*A.m.m* var. *pseudomajus* and *A.m.m* var. *striatum* (hereafter, var. *pseudomajus* and var. *striatum* for brevity)—that are native to France and Spain (Whibley et al. 2006). These varieties have largely nonoverlapping geographic distributions, occupy similar habitats and are pollinated by the same bee species (Tavares et al. 2018). The major difference between them is their contrasting flower colour: var. *pseudomajus* has magenta flowers with a small patch of yellow pigment on the face of the flower below the bee entry point, while var. *striatum* has yellow with restricted veins of magenta coloration above the bee entry point (Figure 1A). These differences in colour, which are thought to be alternative adaptations to attract the same bee pollinators, are caused by a small number of loci that control the production of two flavonoid pigments in floral tissue, anthocyanin (magenta) and aurone (yellow). *Rosea* (Schwinn et al. 2006; Whibley et al. 2006), *Eluta* (Tavares et al. 2018) *and Rubia* (Field et al. 2025) affect anthocyanin production, while *Sulfurea* (Bradley et al. 2017), *Flavia*, *Aurina* (Bradley et al. 2025) and *Cremosa* (Richardson et al. 2025) affect aurone production.

During the last ice age, var. *pseudomajus* and var. *striatum* are thought to have been restricted to areas of low elevation, but subsequently expanded into the Spanish Pyrenees (Vargas et al. 2004; Whibley et al. 2006). As a result, at least three separate hybrid zones have formed in separate valleys below the altitudinal limit of *A. majus* (Figure 1A). In one such zone near the town of Planoles, a transition from yellow to magenta flowers occurs over a few kilometres (Whibley et al. 2006). Scans of genome-wide sequence variation have revealed strong allele frequency differentiation and sharp geographic clines around previously identified colour loci (Field et al. 2025; Tavares et al. 2018; Surendranadh et al. 2025). In contrast, most of the surrounding genome shows low genetic differentiation, probably owing to the homogenising effects of dispersal and recombination (Ringbauer et al. 2018; Tavares et al. 2018).

Here, we expand our analysis to include individuals from the Planoles hybrid zone and a second zone near the town of Avellanet, located over 50 km to the west (Figure 1A). By comparing their respective genomic landscapes and jointly analysing two independent hybrid zones, we hoped to disentangle ancestral divergence from the effects of recent introgression. We were especially interested in whether known flower colour loci act similarly in both localities and stand out from their genomic background. To address this, we used both traditional F_{ST} scans and genealogical methods for studying the genome-wide distributions of tree topologies and coalescence times across the genome. As a secondary aim, we use this study as an opportunity to compare different methods for inferring genealogical trees along the genome. Several approaches are
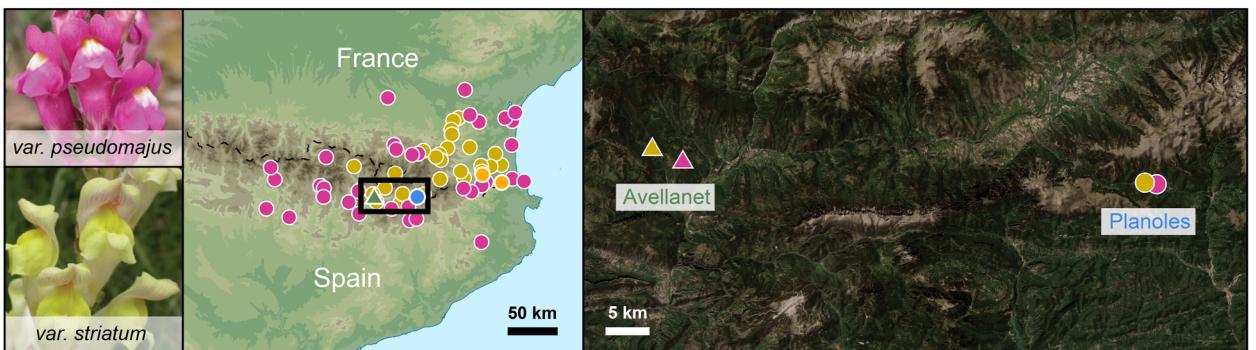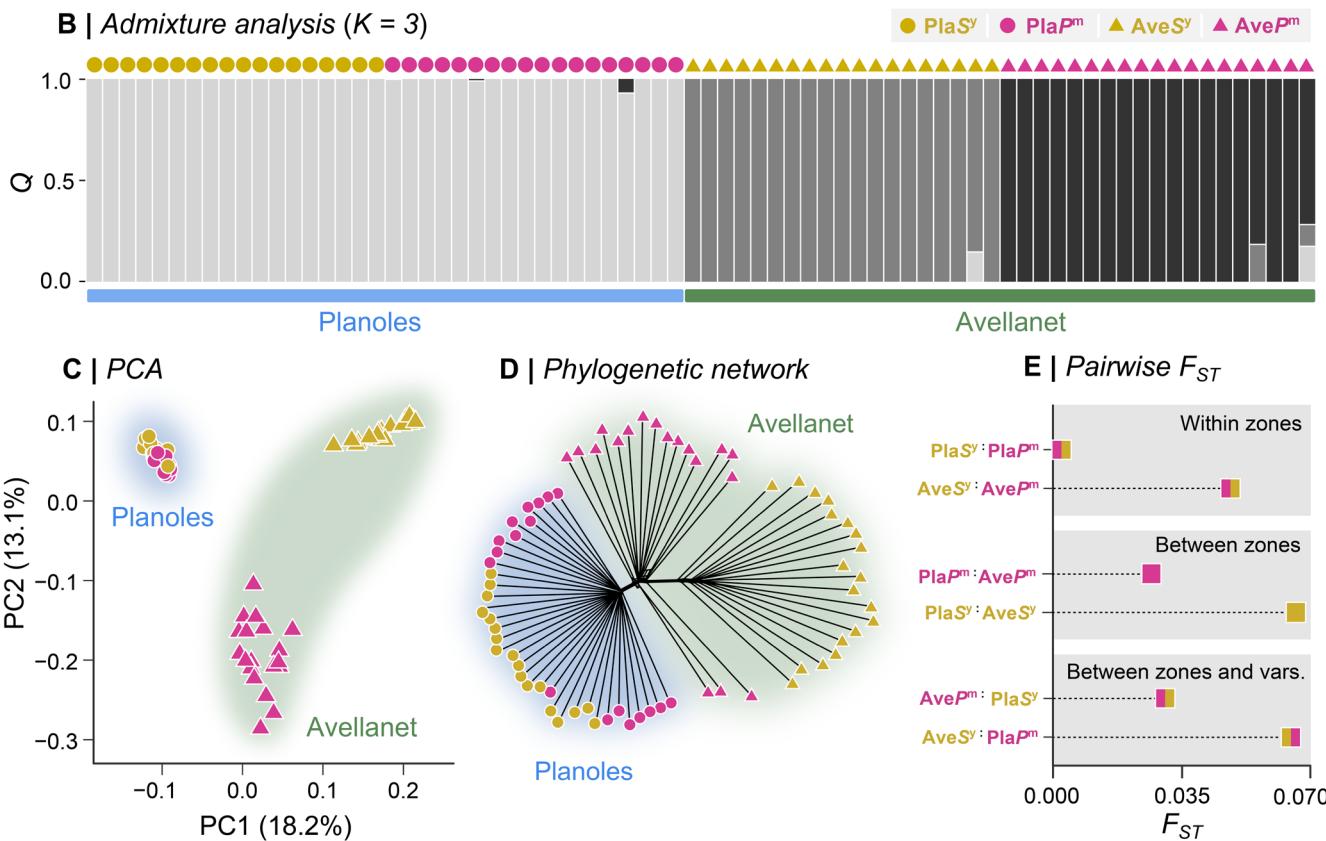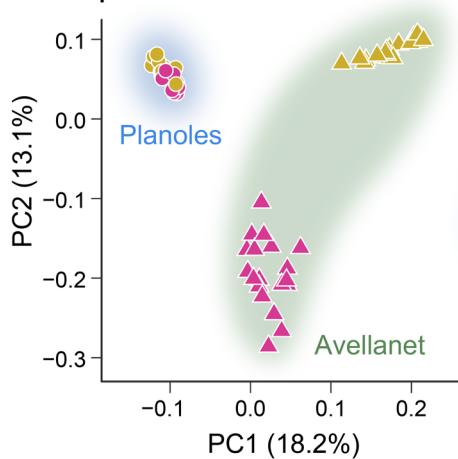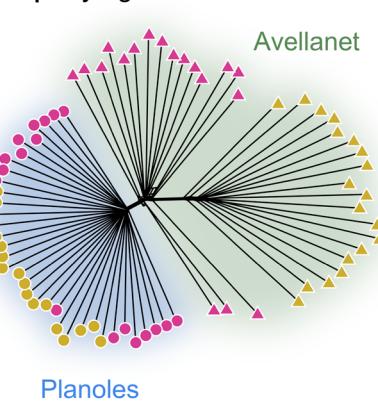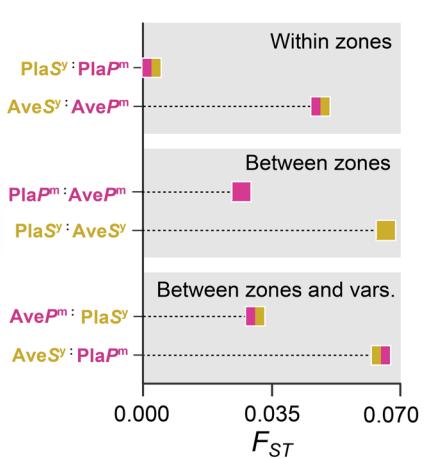
FIGURE 1 | Evolutionary relationship between *A. majus* subspecies *majus* populations from two hybrid zones. (A) Geographic distributions of magenta-coloured *A.m.m.* var. *pseudomajus* and yellow-coloured *A.m.m.* var. *striatum*. The map is based on sample locations in Whibley et al. (2006) and does not show the full distribution of either variant. Circles are coloured according to the population. Samples were collected from two hybrids zones: Avellanet (19 magenta and 19 yellow samples) and Planoles (18 magenta and 18 yellow samples). Points on the map represent the average location of each population. (B) Genetic structure, shown as admixture coefficient (Q) for $K=3$ clusters inferred by *Admixture* from 1.7 million LD-thinned SNPs. Each vertical bar is one individual. (C) The first two principal components of the same dataset. (D) Phylogenetic network (*neightbourNet*) on the same data set. (E) Estimates of per-site Weir and Cockerham's F_{ST}, averaged over all 11.5 million SNPs. F_{ST} was calculated between varieties from the same hybrid zone (PlaP^m vs. PlaS^y, AveP^m vs. AveS^y), and between the hybrid zones (PlaP^m vs. AveP^m, PlaS^y vs. AveS^y, PlaS^y vs. AveP^m, PlaP^m vs. AveS^y). Pla: Planoles, Ave: Avellanet, P^m: magenta-coloured var. *pseudomajus*, S^y: yellow-coloured var. *striatum*.

now available for inferring genealogies from phased SNP datasets (Nielsen et al. 2024). However, these methods have not yet been widely used to study adaptation and speciation. It is also unclear how they perform when applied to real data sets, and there has been limited discussion about when more sophisticated methods might be warranted over simpler ones. We hope that this study helps other researchers decide which method might be most appropriate for their data and specific goals.

2 | Results and Discussion

2.1 | Genome-Wide Analysis Reveals Different Histories of Postcontact Gene Flow Across the Two Hybrid Zones

We sampled 18 individuals of magenta-coloured var. *pseudomajus* and 18 yellow-coloured var. *striatum* from Planoles (hereafter, PlaP^m and PlaS^y), as well as 19 of each variety from Avellanet

(hereafter, AveP^m and AveS^y) (Table S1). We sequenced them using *haplotagging*, a method of linked-read sequencing (Meier et al. 2021). An advantage of *haplotagging* over standard short-read sequencing is the ability to track source haplotypes by means of molecular barcoding. After mapping sequence reads to the *A. majus* reference genome v. 3.5 (M. Li et al. 2019), we followed the variant calling and imputation pipeline outlined in Meier et al. (2021) that leverages linked-read information to identify 11,533,030 bi-allelic SNPs (22 SNPs/kbp) across all the samples. We then phased the SNPs using *SHAPEIT5* (Hofmeister et al. 2023) and used information from a closely related outgroup (*A. molle*; Durán-Castillo et al. 2022) to polarise variants as ancestral or derived.

Based on previous work that showed low genome-wide differentiation between the varieties in Planoles (Tavares et al. 2018), we expected to see similarly low differentiation at the previously unstudied hybrid zone at Avellanet. To test this hypothesis, we generated an LD-thinned dataset containing 1.71 million SNPs and performed *Admixture* (Figure 1B, and S1), principal component (Figure 1C) and phylogenetic (Figures 1D and S2) analysis to characterise the genetic structure. In contrast to our expectations, we found different patterns of genetic structure at each hybrid zone. Specifically, PlaP^m and PlaS^y always formed a single group, rather than clustering by flower colour. In contrast, the AveP^m and AveS^y always formed two distinct groups. This result was also supported by the average genome-wide F_{ST} estimated from all 11.5 million SNPs, which showed that genetic differentiation was much lower at Planoles (PlaP^m vs. PlaS^y: $F_{ST} = 0.003$) than it was at Avellanet (AveP^m vs. AveS^y: $F_{ST} = 0.048$) (Figure 1E). In fact, F_{ST} between AveP^m and AveS^y was higher than between AveP^m and PlaP^m ($F_{ST} = 0.027$), which are separated by more than 50 km, whilst AveS^y and PlaS^y showed the highest pairwise F_{ST} (= 0.066) of all.

The above results suggest a more substantial history of hybridisation and gene flow at the Planoles hybrid zone than at Avellanet. To assess this more formally, we used the program *δαδi* (Gutenkunst et al. 2010) to fit a series of demographic models to the joint site frequency spectrum separately at each hybrid zone. We first compared the fit of a model of strict isolation (SI, where two populations diverge with no gene flow) to a model of secondary contact (SC, where populations diverge in allopatry followed by gene exchange after coming back into contact). For both hybrid zones, the SC model was a far better fit to the data than the SI model (ΔAIC > 2000 for both zones), providing evidence of gene flow between the magenta and yellow populations at each zone (Figure S3, Table S2). It also suggested a more substantial history of gene flow at Planoles characterised by a much longer period since secondary contact than at Avellanet.

Together, these results suggest strikingly different histories of gene flow at each of the hybrid zones, which is largely consistent with observations made at these hybrid zones over more than a decade. At Planoles, plants are abundant every year, and hybrid individuals can be found over broad areas spanning more than 1 km (Whibley et al. 2006). In contrast, we do not always find a large number of plants at Avellanet (Stankowski, Barton & Field; personal observations). In some years, the plants are abundant, and in others, their distribution is patchy and hybrids

are uncommon. Thus, the difference in genetic structure between the zones may reflect the demographic stability of the populations, which is what ultimately provides opportunities for hybridisation and subsequent gene flow across the zone.

2.2 | Genome Scans Reveal Highly Heterogeneous Differentiation Landscapes With Varying Degrees of Parallelism

Although we observed a strong difference in the magnitude of F_{ST} at each hybrid zone, it is possible that the finer-scale pattern of differentiation along the genome is highly similar. Indeed, highly correlated F_{ST} landscapes have been observed in studies where multiple populations with varying levels of differentiation have been compared (Burri et al. 2015; Stankowski et al. 2019). The general explanation for observing correlated differentiation landscapes is that common evolutionary processes and intrinsic genomic properties have shaped variation across multiple incidences of divergence in isolation (Burri et al. 2015), local adaptation (Jones et al. 2012) or secondary contact (Nouhaud et al. 2022).

To test for correlated differentiation landscapes, we first calculated Hudson's F_{ST} in 10-kbp nonoverlapping genomic windows for each pair of populations. This revealed highly variable patterns of differentiation among the comparisons, both in the level of F_{ST} and the pattern of heterogeneity. First, comparing the genome scans between var. *pseudomajus* and var. *striatum* at each of the hybrid zones, we found little heterogeneity in the pattern of differentiation at Planoles. F_{ST} was consistently low across most of the genome (median = 0.008, SD = 0.011) (Figure 2), with the exception of a small number of localised peaks of differentiation rising above the background. In contrast, the F_{ST} landscape was highly heterogeneous at Avellanet (median = 0.028, SD = 0.064), with far greater variability across chromosomes and many areas with pronounced differentiation (Figure 2). The windowed F_{ST} estimates exhibited strong dissimilarity between the hybrid zones (Spearman's rho = 0.04; hereafter, ρ) (Figure S4A, Table S6).

The remaining comparisons showed that differentiation patterns depended heavily on the populations included. Most notably, AveS^y showed highly parallel patterns (ρ ranging from 0.63 to 0.86) between comparisons that included it (Figure S4B, Table S6). This shows that the highly heterogeneous differentiation landscape at the Avellanet hybrid zone is driven more by the history of the AveS^y population than by gene flow between AveP^m and AveS^y.

At both hybrid zones, elevated F_{ST} windows tended to coincide with reduced genetic diversity (π) in one of the two populations, and/or elevated between-population sequence divergence (d_{xy}) (Figure S5). At Planoles, the diversity landscapes were highly similar ($\rho = 0.89$), with outlier regions tending to show lower π in PlaS^y (Figure S6A, Table S6). The diversity landscapes were less similar at Avellanet ($\rho = 0.47$), with the population that showed lower π varying among the genomic regions (Table S6). For example, in the most pronounced F_{ST} island on Chr 2, AveS^y had lower π, whereas AveP^m had lower π in the outlier regions on Chr 1 (Figures S6B and S7). At Avellanet, we found a clear
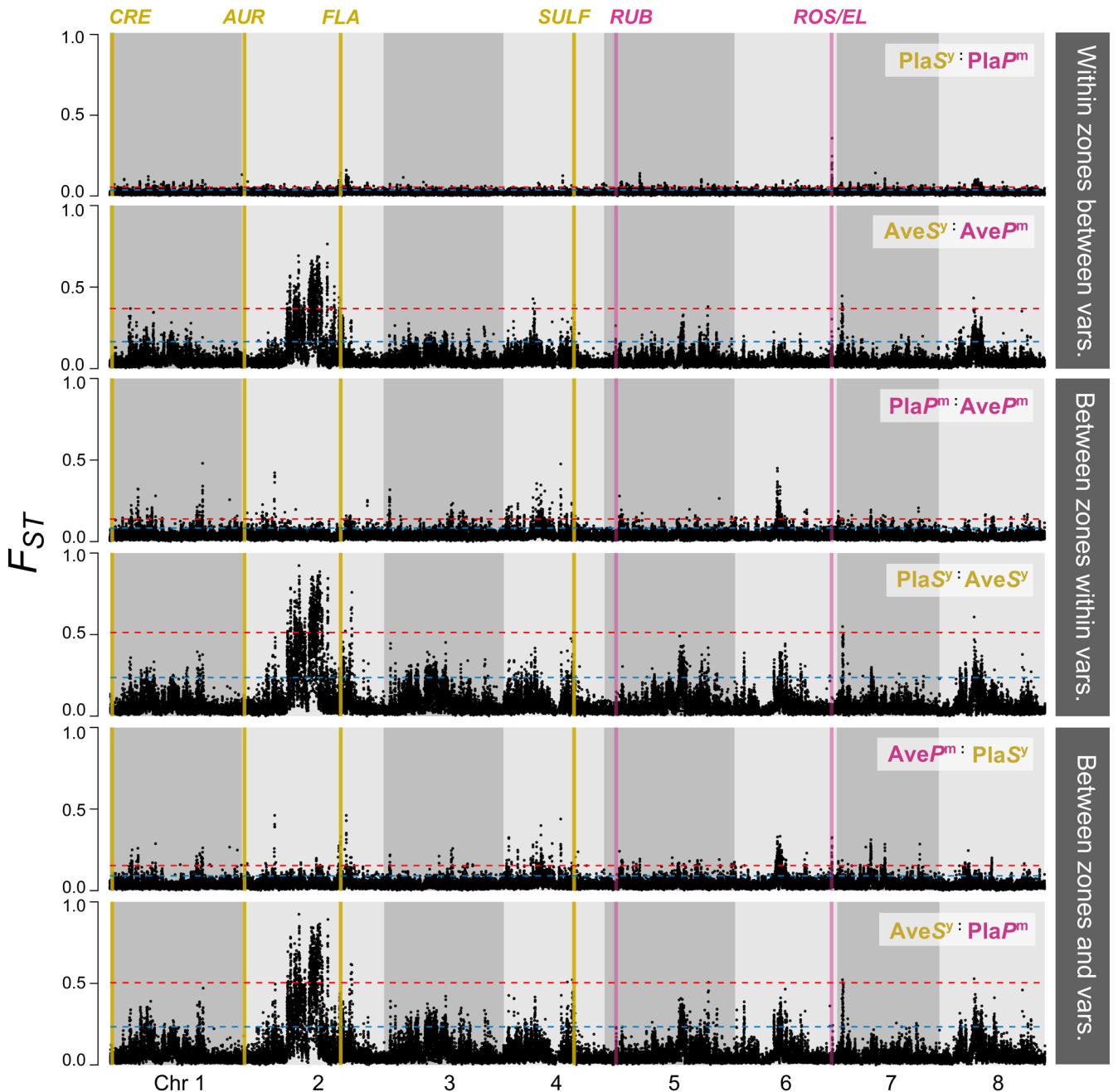
FIGURE 2 | Genome scans show heterogenous F_{ST} landscapes with varying degrees of parallelism. F_{ST} is estimated in 10 kbp nonoverlapping windows ($n=50{,}881$) for each chromosome. Dotted blue and red lines show the 95th and 99th percentile of genome-wide F_{ST} estimates. Top panel (Within zones and between varieties): Comparison between varieties at each hybrid zone (PlaP^m vs. PlaS^y, AveP^m vs. AveS^y). Middle panel (Between zones and within varieties): Comparison between hybrid zones for magenta and yellow population (PlaP^m vs. AveP^m, PlaS^y vs. AveS^y). Bottom panel (Between zones and varieties): Comparison between varieties from different hybrid zones (PlaS^y vs. AveP^m, PlaP^m vs. AveS^y). Grey shading delimits chromosomal boundaries. Pla: Planoles, Ave: Avellanet, P^m: magenta-coloured var. *pseudomajus*, S^y: yellow-coloured var. *striatum*.

negative relationship between the local recombination rate and F_{ST} ($\rho = -0.30$), indicating that highly differentiated regions at Avellanet tended to show lower recombination rates (Figure S5, Table S6). Coupled with the low diversity and strongly elevated d_{xy} in highly differentiated regions (Figure S5), this suggests that F_{ST} has been shaped primarily by widespread linked selection acting independently in the two populations. At Planoles, the relationship between recombination rate and F_{ST} is less pronounced ($\rho = 0.08$), with outlier regions showing variable rates of recombination and modest reductions in π (Figure S5).

We next examined patterns of differentiation at known colour loci to determine whether they would be detected in outlier scans, since flower colour is expected to have evolved before the hybrid zone formations (Tavares et al. 2018). At Planoles, two of the known colour regions were identified as F_{ST} outliers using both 95th and 99th percentile thresholds (Figures 2 and S4A). This included regions containing the *Flavia* locus (hereafter, *FLA*) on Chr 2 that controls the intensity of yellow pigmentation, and the two tightly linked loci *Rosea* and *Eluta* on Chr 6 (hereafter, *ROS/EL*) that have large effects on magenta colouration. At Avellanet, the *FLA* locus

TABLE 1 | Results of tree inference for four genealogical inference methods.

Inference method	Number of trees	Mean tree span (bp)	Mean no. SNPs per tree
Neighbour-joining trees in 50 SNP windows	9237	2126.44	50
Relate	198,375	99.81	2.33
Tsinfer	406,135	49.24	1.14
Singer	1,950,778	9.25	0.24

Note: The methods were applied to the same 2 Mbp region (45–65 Mbp in Chr 2) that contained 461,864 SNPs. The total number of trees inferred by each method, the mean span of trees in bp and mean number of SNPs associated with each tree are provided.

was identified at the 99th percentile threshold, while *ROS/EL* was only detected at the 95th percentile threshold (Figures 2 and S4A)

Overall, we found that the F_{ST} landscapes at the two hybrid zones are quite different, which is not unexpected given their different demographic histories and patterns of hybridisation. Although samples of var. *pseudomajus* and var. *striatum* from Avellanet are more physically distant than in Planoles (Figure 1, Table S1), similarly distant samples at Planoles show a very similar F_{ST} landscape to the one presented here (compare top panel in Figure 2–5E in Field et al. 2025). While we were able to detect two of the major-effect colour loci at Planoles, we failed to detect the other known loci at either Planoles or Avellanet. This may be due to several factors. First, this data set consists of low-coverage sequencing, so SNPs are sparser than previously analysed high-coverage pool-seq data sets (Tavares et al. 2018). Second, previous work at Planoles has shown that the detection of smaller effect loci (e.g., *Rubia* and *Aurina*) depends on the proximity of sampling to the hybrid zone (Field et al. 2025). Our samples were collected very close to the point of contact, meaning that differentiation at small effect loci may have been swamped by gene flow. Finally, the large-effect locus *Sulfurea* is a deletion polymorphism, which has made it difficult to identify with SNP markers in previous F_{ST} scans (Bradley et al. 2017; Field et al. 2025).

2.3 | Different Genealogical Inference Methods Produce Vastly Different Numbers of Trees Yet Infer Similar Genealogical Landscapes

Given the challenges of interpreting multiple pairwise F_{ST} scans, we next shifted to genealogical tools that allowed us to jointly analyse relationships among all four populations. Among the variety of tools available, we selected and compared four that were broadly representative of the main approaches in methodological implementation: (i) Neighbour-joining trees in arbitrary windows (Martin and Van Belleghem 2017); (ii) *tsinfer* (Kelleher et al. 2019); (iii) *Relate* (Speidel et al. 2019); and (iv) *Singer* (Deng et al. 2024). We restricted our comparison to a 2 Mbp region (45–65 Mbp in Chr 2) with ~462k SNPs that showed highly heterogeneous patterns of F_{ST} (Figure 2) and contains the *FLA* locus (Bradley et al. 2025).

The first method divides the genome into nonoverlapping windows containing the same number of SNPs (50 SNPs in our analysis) and infers a phylogenetic tree for each region separately. We inferred neighbour-joining trees, though other methods such as maximum-likelihood have also been applied (Fontaine et al. 2015). While simple and widely used, this approach has

a significant limitation. Arbitrarily defined genomic segments often span historical recombination events, where relationships between haplotypes cannot (and ideally should not) be accurately represented by a single bifurcating tree (Shipilina et al. 2023). As a result, important genealogical signals may be dampened by the clumping of unique trees into one.

Unlike NJ trees, *tsinfer, Relate* and *Singer* infer a sequence of trees, consistent with how historical recombination events have altered genealogical relationships across the genome. They all do this by allowing topologies to vary locally to reconcile neighbouring site patterns that cannot be represented as a single bifurcating tree, but their approaches vary significantly. *Tsinfer* reconstructs plausible ancestral sequences from sampled chromosomes and then infers the relationship between those sequences, preserving the correlation between consecutive genealogical trees. Therefore, neighbouring trees inevitably share many of the same nodes and branches. *Relate*, on the other hand, infers a completely new tree upon encountering an incompatible SNP. Therefore, consecutive trees do not share homologous nodes or branches, although this can be partly addressed by assigning the same age to nodes with identical descendant sets across adjacent trees.

Although *tsinfer* and *Relate* accommodate the effects of past recombination, they do not explicitly model recombination. In other words, incompatible SNP patterns only imply that historical recombination occurred somewhere between the boundary of neighbouring trees. *Singer* goes further and takes a Bayesian approach, attempting to infer the full ARG by fitting a model of coalescence and recombination to the SNP data. So, unlike the deterministic topology inference of *tsinfer* and *Relate* (i.e., multiple runs will always produce the same tree topologies, though inferred branch lengths may differ between runs), each MCMC iteration of *Singer* estimates a tree sequence that is drawn from the posterior distribution of possible trees. Thus, a region with no SNP may contain multiple inferred trees that are not supported by any data within that region but are a plausible outcome of the estimated model.

Examination of the resulting tree sequences shows that methods produce vastly different results (Table 1). First, we found that the number of trees varied substantially across methods. The neighbour-joining method contained the lowest number of trees at 9237 (i.e., 1 tree for each 50 SNPs window). *Relate* and *tsinfer* inferred substantially more trees, with 198,375 and 406,135, respectively. *Singer* inferred the most trees by far, with 1,950,778. The average span of an NJ tree was 2.1 kbp (SD = 2.3 kbp), compared with 100 bp (SD = 322 bp) for *Relate*, 50 bp (SD = 222 bp)
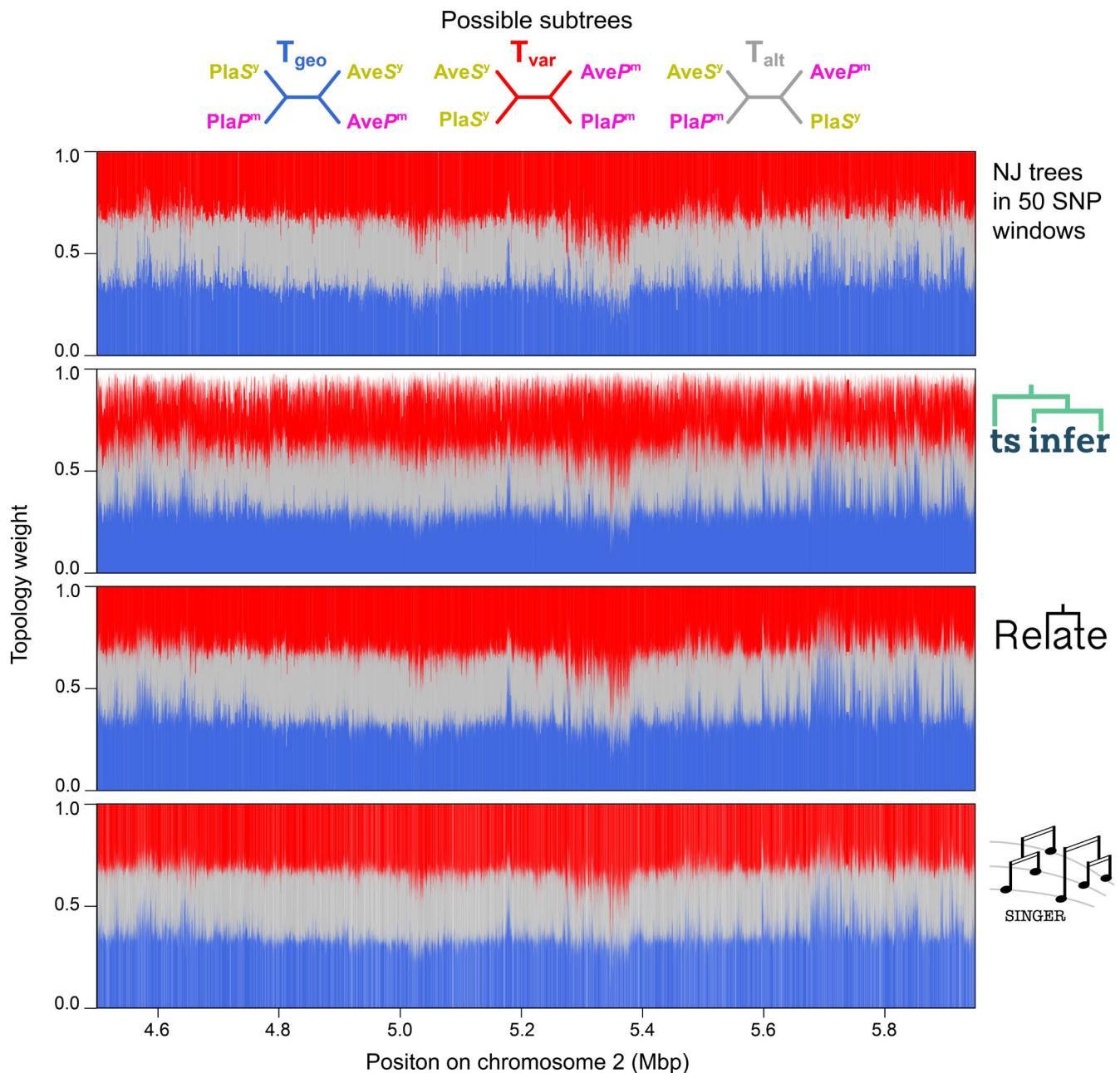
FIGURE 3 | Topology weighting of trees sequences inferred by four different methods yield broadly similar genealogical landscapes. Topology weights of the three possible subtree topologies (T_{geo}, T_{var}, T_{alt}) are plotted for each tree in the sequence along a small section of Chr 2. Each vertical bar shows the proportions of each topology in one genealogical tree. Therefore, topology weights add up to 1 except for trees inferred by *tsinfer* since it allows polytomies.

for *tsinfer* and 9 bp (SD = 24 bp) for *Singer*. Finally, we asked how many SNPs fell within the span of each marginal tree. Marginal trees in *Relate* contained an average of 2.3 SNP (SD = 1.7) compared with 1.14 (SD = 0.47) for *tsinfer*. On average, trees for *Singer* contain less than 1 SNP (0.24, SD = 0.54), with 1,950,778 (80.73%) trees containing no SNP.

Overall, the characteristics of each tree sequence align with their respective methodological approaches. The number of SNPs associated with each NJ tree is defined by the user and will ultimately reflect a trade-off between information content (i.e., number of SNPs) and tree span. In an ideal world, window size would be minimised such that trees span as few recombination events as possible. However, if we assume that the transitions between trees by *tsinfer* reflect real recombination events, this would imply that the average 50 SNP window spans 43 observable recombination events. Although *Relate* and *tsinfer* define margins between trees based on incompatible SNP patterns, *Relate* produces half the number of trees. This may reflect different levels of tolerance for incompatibilities, and the number of trees may vary depending upon the parameters chosen by the user. Finally, because *Singer* models recombination explicitly, it inevitably produces far more trees than the other methods. Although many trees in the sequence are not supported by SNP data, allowing recombination to shape the sequence in the absence of polymorphism data is more consistent with reality and
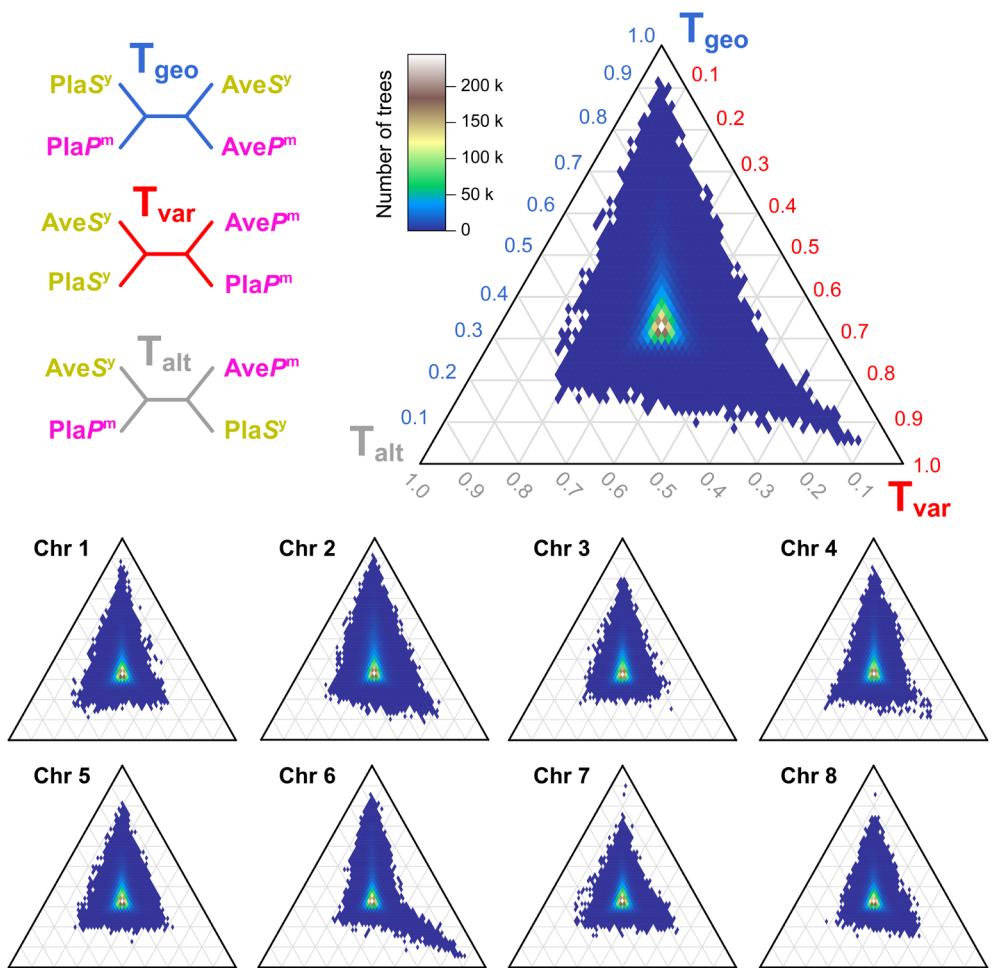
FIGURE 4 | Ternary plots showing the joint distribution of topology weights. Empirical distributions of weights for the 4,975,454 trees inferred using *Relate* (top right) and for each chromosome (bottom). Each tile in the distribution is coloured according to the density of genealogies falling in that area of the distribution, as indicated in the colour scale. The three topologies associated with each axis in the ternary plot are shown in the top left of the plot. The three corners of the ternary plot—[1,0,0], [0,1,0], [0,0,1]—correspond to trees that perfectly match the three possible group-level subtrees.

may provide additional information in some inference schemes. However, it seems reasonable to exercise caution when making detailed inferences from trees that are not supported by SNP data.

Moving beyond the summaries of tree sequences, we next used topology weighting to compare how topologies change along the genome. Topology weighting iteratively subsamples one haplotype from each population and estimates the proportion of each subtree topology. In this data set of four populations: AveP^m, AveS^y, PlaP^m and PlaS^y, it weighs the contribution of each of the three possible topologies (Figure 3)—the geography tree (T_{geo}), where samples cluster by hybrid zone ((AveS^y, AveP^m) (PlaS^y, PlaP^m)), the variety topology (T_{var}), where samples cluster by the variety, ((AveS^y, PlaS^y) (AveP^m, PlaP^m)), and an alternative topology (T_{alt}), where samples neither cluster by geography or variety ((AveP^m, PlaS^y) (AveS^y, PlaP^m)). By iteratively sampling many subtrees (in our case 10,000), we can obtain their relative frequencies (i.e., topology weights), which provide a measure of the weight (or bias) of the full tree to each group-level topology (Figure 3).

Although the characteristics of the tree sequences vary among the four methods, there is striking similarity in genomic distributions of the topology weights inferred from them. Figure 3 shows the weights for the three group-level topologies. From visual inspection alone, the topology weights are highly similar among the methods, increasing and decreasing in a coordinated way along the chromosome. Correlation analysis of the weights, performed on the topologies that coincide with SNP positions, shows the similarity is indeed quite strong among the methods (ρ value ranges for T_{geo}: 0.57–0.77, T_{var}: 0.57–0.66, T_{alt}: 0.45–0.71; Figure S8, Table S7).

However, there are also some clear distinctions. First, the change in amplitude of the weights is not as extreme in the NJ method compared with the other methods. This is not surprising, as the 50 SNP windows span many distinct marginal trees, which we would expect to have a smoothing effect. Another major difference is that topology weights from *tsinfer* do not add up to 1, implying that some subtrees cannot be classified as one of the three possible topologies. The reason for this is that *tsinfer* infers polytomies, while *Relate* and *Singer* force all branches to bifurcate.
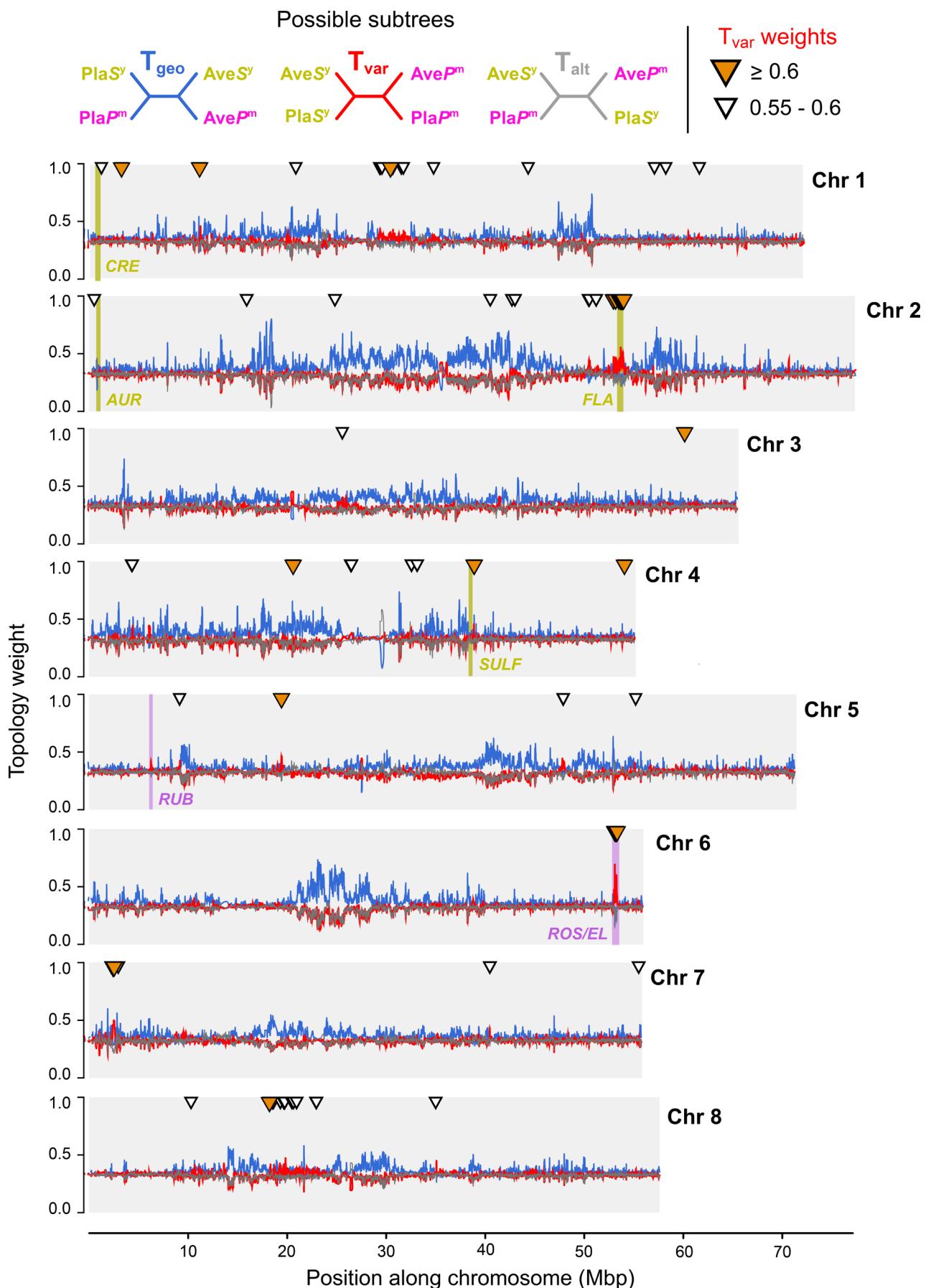
FIGURE 5 | Legend on next page.

FIGURE 5 | Genealogical landscape of parallel hybrid zone formation revealed by topology weighting. Topology weights (loess smoothed, span = 50 kbp) for the 4,975,454 trees inferred by *Relate* plotted along each chromosome. 7 loci controlling flower colour are highlighted in yellow or magenta. White triangles indicate trees with raw T_{var} weight between 0.55 and 0.6. Orange triangles indicate trees with raw T_{var} weights ≥ 0.60.

In summary, the results of our comparisons show that the different methods produce vastly different tree sequences, yet largely agree on how group-level relationships change along the genome for our snapdragon data set. It is good to know that the crudest and most sophisticated approaches give a similar picture, if only from a topological standpoint. Deciding which method to use will depend on the size of the dataset and the goals of the study. However, we see little reason to use the window-based approach given that more computationally efficient and precise methods are now available. *Tsinfer* and *Relate* are far better options but have different strengths. For example, *tsinfer* retains nodes and branches among trees, meaning that they can be represented as an ARG and used in analysis that leverage homology of tree features (Shipilina et al. 2023). In contrast, *Relate* has been shown to be more accurate than *tsinfer* (+*tsdate*) when it comes to estimating deeper coalescence times (Brandt et al. 2022). *Singer* is far more computationally demanding than the other methods, so it is difficult to scale to large data sets. However, for smaller data sets, and in more defined genomic regions of interest, *Singer* allows for extremely fine-scale genealogical inference along with estimates of uncertainty.

2.4 | Topology Weighting Reveals Regions Associated With Flower Colour

For our purpose, *Relate* seemed to be the ideal choice to infer genome-wide genealogies, due to its scalability, efficiency, resolution of polytomies and accuracy of inferring deeper coalescence times. Applying *Relate* to our genome-wide dataset yielded a sequence that contained 4,975,454 trees, with an average span of 101.2 bp (SD = 739.4 bp). We again used topology weighting to quantify bias towards the three group-level relationships (T_{geo}, T_{var}, T_{alt}) for each marginal tree.

We first analysed the distribution of all topology weights in a ternary framework using the program *TwisstNtern* (Stankowski et al. 2024). The ternary plot is a natural framework for analysing the joint distribution of weights in a tree with four groups because it is possible to graphically represent each tree as a single point in an equilateral triangle based on the three weights. The three corners of the ternary plot—[1, 0, 0], [0, 1, 0], [0, 0, 1]—correspond to all trees where the sampled subtrees match only one of the three possible group-level subtrees. In contrast, the centre of the ternary plot—[1/3, 1/3, 1/3]—corresponds to trees where all three of the possible subtrees are found at equal frequency. Any other location in the ternary plot indicates an enrichment of one particular subtree topology. Previous simulations have shown that the ternary distribution of weights can be shaped by a range of factors, including population split times and effective population sizes, as well as processes that lead to haplotype sharing between nonsister groups (e.g., introgression) (Stankowski et al. 2024).

In our analysis, we expected the ternary distribution of weights to be skewed toward the geography topology (T_{geo}, top of the triangle in Figure 4), because this topology matches the genome-wide relationships observed between the populations (i.e., topology weighting of the genome-wide neighbour-joining tree (Figure S2) yields weights of $T_{geo} = 1.0$, Tc = 0.0, $T_{alt} = 0.0$). While we did observe this skew, the bias towards T_{geo} was relatively weak (mean $T_{geo} = 0.367$, $T_{var} = 0.317$, $T_{alt} = 0.316$), and similar across the 8 chromosomes. Although some of the trees showed high T_{geo} weights (max $T_{geo} = 0.93$ with 5% of trees showing weights above 0.49), none of the 4,975,454 trees perfectly matched T_{geo}. Rather, most of the genealogies clustered near the centre of the ternary plot (i.e., most weights were near 0.33 for all three topologies), indicating that haplotype diversity is broadly shared across the four groups.

Notably, we found striking left–right asymmetry in the distribution of topology weights between the left and right halves of the ternary plot (Figure 4). Specifically, we observed a long tail of topology weights extending towards the right-hand corner of the plot, resulting in a 1% bias in the distribution towards the variety topology (T_{var}). Such a bias is unexpected when sharing is due to the random sorting of ancestral polymorphism, as there is an equal chance that any given tree will be biased towards either one of the discordant topologies, leading to a symmetrical distribution of weights (Stankowski et al. 2024). This asymmetry is similar to what is measured by the site-based statistic Patterson's *D*.

Indeed, roughly symmetrical distributions were observed on several chromosomes, including Chr 1, 3 and 5 (Figure 4). The remaining chromosomes showed substantial asymmetry towards the variety topology, driven by a relatively small number of genealogical trees (1490 or 0.03%) with T_{var} weights that exceeded 0.55 (Figure S9). This indicates a bias of haplotype sharing between populations of the same variety (Figure S9). The most striking bias was observed on Chr 6, where weights approached 0.9.

To explore regions associated with the genetic differentiation of varieties, we plotted the genomic positions of detected T_{var} outliers (Figure 5). T_{var} outliers were spread across multiple points along each of the chromosomes rather than clustering at a single site. Most of the known colour genes were observed near T_{var} outliers, but we also observed a bias toward T_{var} in regions of the genome that have no known effect on flower colour, including regions of Chr 1, 5, 7 and 8.

2.5 | Coalescence Times at *Flavia* and *Rosea/Eluta* Differ From Surrounding Background

The two genomic regions that showed the clearest association with the colour topology were on Chr 2 and Chr 6, together accounting for 66% of all T_{var} outliers (or 84% using the 0.6 cutoff). The outlier region on Chr 2 includes the recently discovered *Flavia* (*FLA*) locus, which affects the patterning of

yellow colouration in the face of the flower (Bradley et al. 2025) (Figure 6). This signal of T_{var} enrichment extends over roughly 2 Mbp of the chromosome, interrupting bias towards T_{geo} on either side of it. Within the *FLA* locus, weights for some genealogies exceed 0.7 (Figure 6). The other region, *ROS/EL* located on Chr

6, contains the two linked colour loci, *Rosea* and *Eluta*. *Rosea* activates anthocyanin biosynthesis across the corolla, while *Eluta* modifies its distribution (Tavares et al. 2018). Within the *ROS/ EL* region, T_{var} weights are strongly elevated and characterised by local peaks and troughs spanning about 1 Mbp. On either side
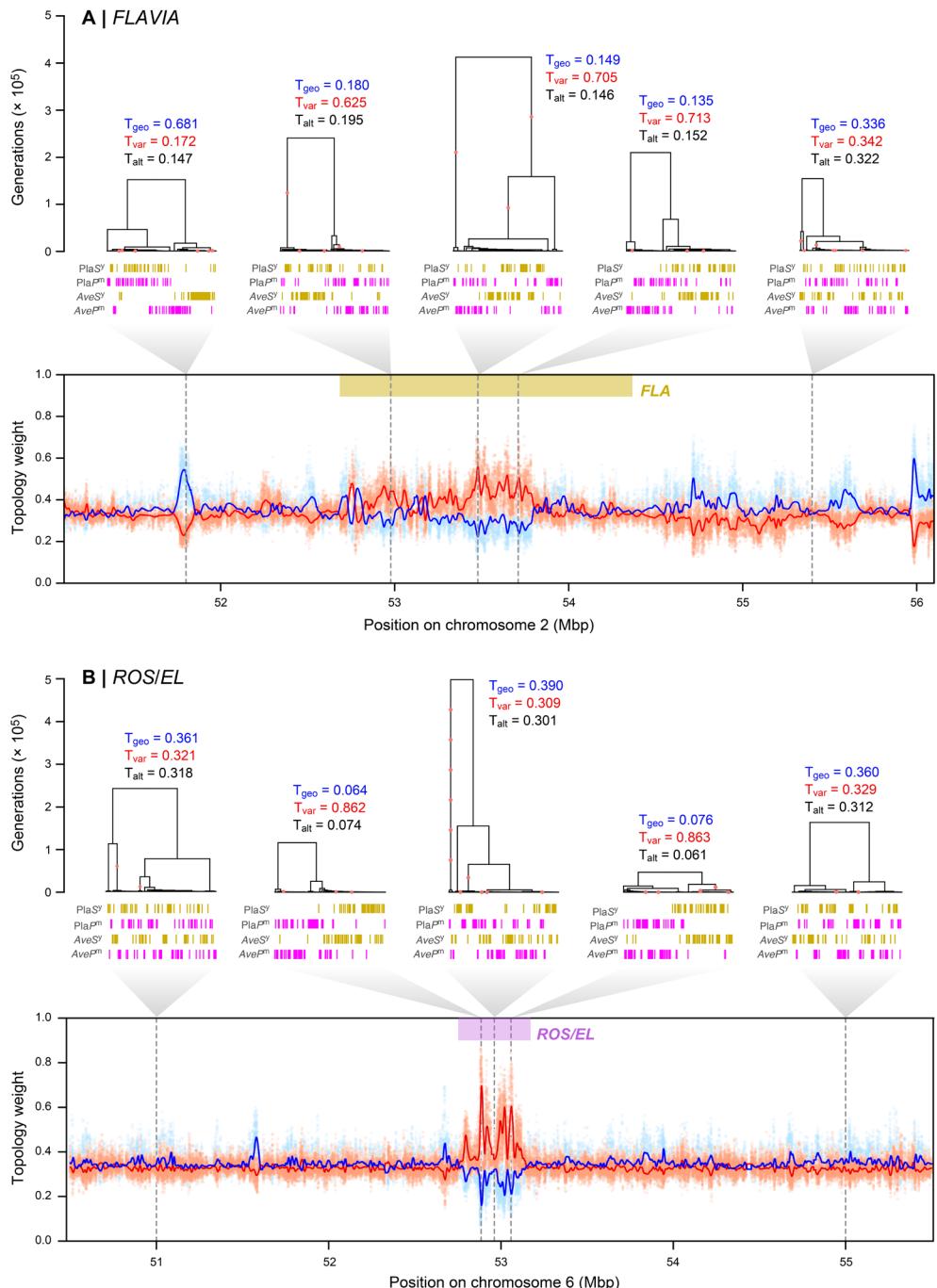

FIGURE 6 | Fine-scale genealogical landscape at the *FLAVIA* and *ROS/EL* loci. (A) 5 Mbp genomic region centred around *Flavia*. Trees show relationships at various points along the sequence, with red circles indicating mutations associated with each tree. Vertical bars represent haplotypes, coloured according to the populations (PlaS^y, PlaP^m, AveS^y, AveP^m). From left to right (1–5), trees 1 and 5 are chosen arbitrarily, but equally distant from the locus. Trees 3 and 4 are trees have the highest smoothed and raw T_{var} weights, respectively. Bottom panel shows topology weights (T_{geo}, blue; T_{var}, red; T_{alt}, black) through the region. Solid lines are loess smoothed weights (span = 50 kbp), while dots are raw weights. *FLA* locus (Chr2:52560000–54,050,000) is marked in a yellow bar, while the rest is considered as flank in TMRCA calculations. (B) Same as (A), but for *ROS/ EL* locus. From left to right, trees 1 and 5 are equally distant from the colour locus, while trees 2–4 are within in. Tree 2 and 4 have the highest raw T_{var} weights at the *ROS1* and *EL* loci. Tree 3 shows a low T_{var} likely due to recombination between the two linked loci. *ROS/EL* locus (Chr6:52775000–53,150,000) is marked in a magenta bar, while the rest is considered as flank in TMRCA calculations.
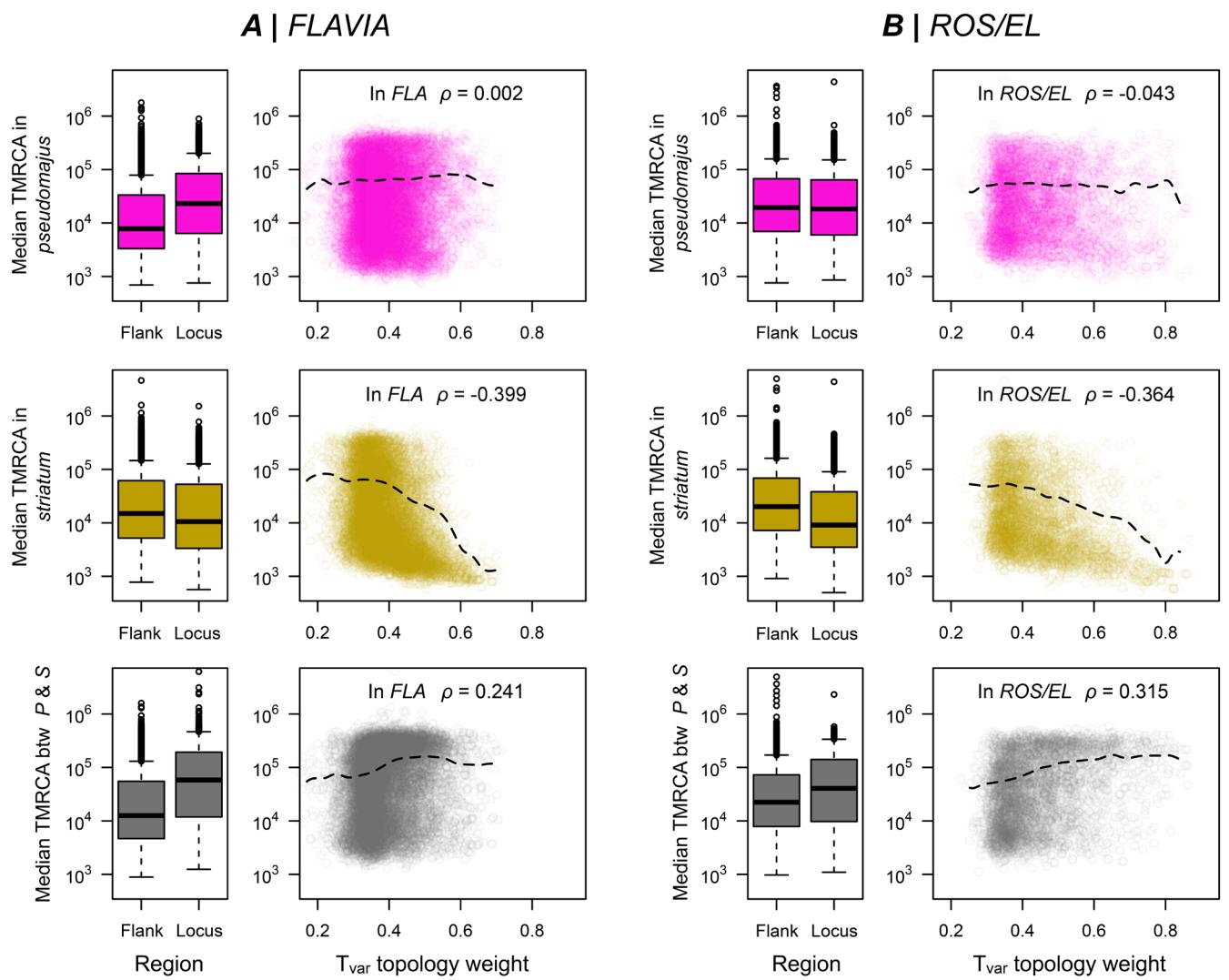
FIGURE 7 | Coalescence times at *FLAVIA* and *ROS/EL*. Boxplots for (A) *FLAVIA* and (B) *ROS/EL* show the median time to most recent common ancestor (TMRCA) within var. *pseudomajus* (top row), within var. *striatum* (middle row), and between var. *pseudomajus* and var. *striatum* (bottom row), all on a log scale. The left boxes ('Flank') show the TMRCAs in the flanking regions around the locus, while the right boxes ('Locus') show the values inside the locus. The scatterplots and dashed black lines show the full distribution and the overall smoothed trend between the T_{var} weight and median TMRCA within the locus, (rows are as indicated for the boxplots). ρ is the correlation coefficient from Spearman's rank correlation.

of *ROS/EL*, all three topology weights hover around 0.33, indicating that haplotype variation is broadly distributed among the groups.

Given existing evidence for selection on *FLA* (Bradley et al. 2025) and *ROS/EL* (Tavares et al. 2018), we next examined the coalescence times for genealogies in and around these colour loci. Since positive selection purges haplotype diversity from the population, we may expect to find shallower coalescence times within each variety (often measured using π) reflecting the historical sweep of causal alleles (Hejase et al. 2020). In addition, these loci can also generate local barrier effects in the genome, which we would expect to increase coalescence times between the varieties (often measured using d_{xy}) (Hejase et al. 2020; Wakeley 2009).

To test for these patterns, we first compared the median time to the most recent common ancestor (TMRCA) for genealogies inside each locus to those in the flanking regions of

the loci where there was no obvious association with colour (Figure 6). For *FLA*, the median TMRCAs for var. *pseudomajus* (i.e., PlaP^m + AveP^m) were higher inside the locus than in the flanking regions; whereas, for *ROS/EL*, there was no obvious difference (Figure 7). For var. *striatum* (i.e., PlaS^y + AveS^y), we observed a similar result in *FLA* and *ROS/EL*, with median TMRCAs being lower in the loci than in the flanking regions. We also compared TMRCAs between the varieties, finding higher median TMRCAs inside *FLA* and *ROS/EL* loci in comparison to the flanking regions.

We also examined the relationship between the TMRCAs and T_{var} weights inside each locus, as we expected signatures to be most pronounced for genealogies that more closely resembled the variety topology. In *FLA*, we found no relationship between the median TMRCA and T_{var} weight within var. *pseudomajus* ($\rho = 0.002$, Figure 7). For *ROS/EL*, we only observed a weak negative relationship, with higher T_{var} trees showing a broad range of median TMRCAs within var. *pseudomajus* (Figure 7).

In contrast, we found a very clear negative relationship between median TMRCA and T_{var} weight within var. *striatum* for both loci, such that genealogies with high T_{var} weights tended to have shallower median TMRCAs ($\rho = -0.399$ for *FLA* and -0.364 for *ROS/EL*, Figure 7). We also observed a clear positive relationship between T_{var} weights and median TMRCAs between the two varieties, such that genealogies with a high T_{var} tended to have higher TMRCA or deeper coalescence times. Similar patterns are observed when each hybrid zone was analysed separately (Figures S10 and S11).

While preliminary, these results are consistent with (i) selection having acted on haplotypes associated with colour and (ii) suggest that these loci have a local barrier effect. For *FLA*, evidence for selection mainly comes from sharp allele frequency clines at the Planoles hybrid zone (Bradley et al. 2025; Field et al. 2025, Surendranadh et al. 2025). Our results provide preliminary evidence that a selective sweep has occurred on the background of var. *striatum* and the allele is present in both Avellanet and Planoles. The lack of a signature in var. *pseudomajus* is consistent with the phenotypic effect of *FLA*, as it only affects yellow colouration. The *ROS/EL* locus is more complex, with previous evidence suggesting that there have been multiple independent sweeps at the two linked loci (Tavares et al. 2018). Here, we only found evidence for selection in the yellow group, while previous work has suggested sweeps on both the yellow and magenta backgrounds. Our results do not preclude such a sweep, as footprints of selection are transient and fade with time. However, it might indicate a more recent sweep in the yellow population. Also, our analysis is coarse-grained and does not consider how fine-scale genealogical relationships change across the region. In future, we plan to use genealogical tools and a much larger dataset to dissect this region in fine detail.

2.6 | Conclusion and Implications for Genomic Studies of Speciation

In this paper, we studied the genomic landscape associated with replicate hybrid zones in *Antirrhinum majus*. Our study highlights many of the known challenges in interpreting genome scans in the context of adaptation and speciation, as differentiation landscapes can be shaped by a multitude of factors and processes that have nothing to do with speciation per se (Ravinet et al. 2017; Wolf and Ellegren 2017). Comparing two parallel hybrid zones, we show that genome scans can be dominated by signals of historical demography, a factor less widely discussed but that must be accounted for when interpreting patterns of genome-wide variation. At Planoles, var. *pseudomajus* and var. *striatum* show very little differentiation, as we would expect between taxa that were described as varieties based only on a difference in flower colour. At a second previously unstudied hybrid zone at Avallenet, differentiation was far more striking and characteristic of more divergent taxa (Bolnick et al. 2023; Stankowski and Ravinet 2021). This shows that differentiation landscapes can be extremely variable within species, highlighting the dangers of generalising broader processes from a single pair of samples.

Our analyses suggest that different levels of divergence at the two hybrid zones are primarily due to variation in the timing and/or rate of gene flow following secondary contact. This raises important questions about what has caused this difference and, more broadly, the long-term dynamics of gene flow between *Antirrhinum* varieties in the Pyrenees. The strongest patterns of differentiation were always observed in comparisons that included var. *striatum* from Avellanet. This result can be explained in several ways. For example, varieties at Avellanet may exhibit historical divergence that was typical of allopatric var. *striatum* and var. *pseudomajus*. This divergence may have been maintained at Avellanet either by a lower rate of gene flow than at Planoles or by secondary contact being much more recent than at Planoles. The current spatial distribution of var. *pseudomajus* hints at the first possibility, as the continuous populations around Planoles contrast with the patchier distribution at Avellanet. Moreover, a slightly distant population of var. *striatum*, isolated from Planoles by a mountain pass, shows a similar pattern of differentiation when compared to populations at Planoles (Field et al. 2025). Another possibility is that the var. *striatum* population from Avellanet was historically isolated, causing its demographic history to be distinct from other populations. These questions are beyond the scope of our current dataset, and more detailed work is needed to understand the biogeographic and evolutionary history of *A. majus*.

Topology weighting of marginal trees inferred from genealogies allowed us to identify loci associated with the two snapdragon varieties. Flower colour is the only trait that consistently differs between the varieties, suggesting that the T_{var} outliers may underpin this variation. Two of the loci known to cause differences in pigmentation, *Flavia* (Bradley et al. 2025) and *Rosea/Eluta* (Schwinn et al. 2006; Tavares et al. 2018; Whibley et al. 2006), were identified as T_{var} outliers; whilst others, *Cremosa* (Richardson et al. 2025), *Rubia* (Field et al. 2025), *Sufurea* (Bradley et al. 2017) and *Aurina* (Bradley et al. 2025) did not show clear associations. This could be due to several reasons, including proximity of samples to the core of the hybrid zones, effect size of the loci and sequencing coverage. The remaining T_{var} outliers have not been previously associated with colour. It is possible that these outliers underpin some other trait (floral or nonfloral) that differs between the varieties, or they may simply be spurious associations reflecting the highly stochastic nature of the coalescent process. This highlights a more general limitation of *all* genome scans: They detect regions of elevated differentiation between populations, and more detailed mapping studies and functional work are needed to demonstrate causality.

Finally, we encourage others to explore and critically evaluate the utility of genealogical methods in their research. Several recent studies, mostly focusing on human populations, suggest that genealogical tools can lead to more accurate inferences about past evolutionary processes (Fan et al. 2022, 2023; Speidel et al. 2019; Stern et al. 2019; Wohns et al. 2022). However, relatively few studies have used genealogical methods to study adaptation and speciation (Campagna et al. 2017; Hejase et al. 2020, 2022; Hooper et al. 2024; Meyer et al. 2024; Rueda-M et al. 2024; Stankowski et al. 2024; Wang and Coop 2022). Our preliminary genealogical comparison of known adaptive loci with the surrounding genomic background further highlights the potential of these tools for

studying the interplay between selective sweeps and barriers to gene flow. ARGs and tree sequences are very rich structures that are complex and challenging to interpret (Shipilina et al. 2023). However, paired with new linked-read sequencing methods, we think there is tremendous scope for creativity around how we can best visualise local genealogical relationships, account for uncertainty, and identify signatures that are associated with the speciation process.

3 | Materials and Methods

3.1 | Sample Collection, DNA Extraction and Sequencing

Leaf material was collected from individuals of *A. majus* at two hybrid zones near the towns of Planoles (42.3162° N, 2.1039° E) and Avellanet (42.3503° N, 1.3288° E) (Table S1). Several leaves were collected from each individual and refrigerated at 4°C before further processing. DNA was preserved by placing leaf tissue in a paper envelope, and the envelopes into an air-tight plastic bag with silica gel. DNA was extracted using a custom protocol optimised for isolating high molecular weight DNA (Methods S1).

3.2 | Library Preparation and Sequencing

Sequencing libraries were constructed by mixing genomic DNA with a pool of haplotagging beads with a different set of A and C barcode oligos (see Supplementary Table 1 for oligonucleotide sequences). This modification shifts the barcode position for the A/C segment from the original i7 index position into Read 2, followed by a mutated Tn5-17A/18G-MEmut sequence—ACTTGTGTATAAGAGACAG (Steiniger-White et al. 2002). The mutated Tn5-MEmut sequence allows tagmentation but does not otherwise interfere with Illumina sequencing. An additional standard 8-bp i7 Illumina index barcode was added during the final PCR amplification to introduce a fifth barcode segment to allow multiplexing of more than 384 samples. Amplified libraries were cleaned up and size-selected using Ampure magnetic beads (Beckman Coulter), Qubit quantified and adjusted with 10 mM Tris, pH 8, 0.1 mM EDTA to 2.5 nM concentration for sequencing. Libraries were sequenced aiming for 2× coverage with Illumina paired-end sequencing (2×150 bp) across a lane of Novaseq 6000 S4 by Azenta Life Sciences (Leipzig Germany). The sequences were then demultiplexed by recognising and trimming away the Tn5-MEmut sequence from R2 and the remaining B/D and A/C along with the Plate barcodes. The remaining sequences were processed as previously described in Meier et al. (2021).

3.3 | Processing of Raw Reads and Read Mapping

Raw reads were mapped to the *A. majus* reference genome v3.5 (Li et al. 2019) using *EMA v0.7.0* (Shajii et al. 2018), a BX-tag-aware modification of *BWA* (Li 2013). First, haplotag barcodes with BX tags were converted to 16-basepair barcodes using 16BaseBCGen (https://tinyurl.com/SamHaplotag). Reads with

correct BX tags (98.14%) were then mapped with *EMA*, which favours alignments where reads with the same barcode group together. Reads with faulty BX tags (1.86%) were mapped to the genome using *BWA v0.7.17*. The resulting BAM files were combined and checked for quality using the *multi-bamqc* command in *qualimap v2.2.1* (Okonechnikov et al. 2016) (Table S1). PCR and optical duplicates were marked and removed using the *markdup* tool in *sambamba* (Tarasov et al. 2015).

3.4 | Variant Discovery Imputation, Phasing and Allele Polarisation

We used the *mpileup* and *call* commands in *bcftools v1.18* (Danecek et al. 2021) to identify candidate sites that were then used for final genotype inference and imputation by *STITCH v1.6.10* (Davies et al. 2016). Variant calling was performed with the *bcftools* multiallelic calling program using the flags -*m* and --*annotate AD, ADF, ADR, DP, QS*, SP. The resulting VCF was filtered to remove low-quality and potentially erroneous variant sites (Table S2). We first removed all INDELs (*bcftools view -V indels*), all SNPs within five basepairs of INDELs (*bcftools filter –SnpGap 5*), all monomorphic REF or ALT sites (*bcftools view -m2 -e "AC==AN || AC==0"*), and all sites with more than two alleles (*bcftools view -M2*). Next, we removed all sites with g 2.5 times the mean coverage across all samples (130×), sites with a genotype quality score < 20 and a mapping quality score < 30 (*bcftools filter -e "INFO/DP> 130 | QUAL< 20 | MQ< 30"*). Finally, bi-alleleic sites with > 0.8 of missing genotypes were removed (*bcftools view -e 'F_MISSING> 0.80'*), producing a set 11,574,426 candidate sites (Table S2).

We applied *STITCH* to impute variants for the 11 million sites described above. *STITCH* models each chromosome as a mosaic of *K* founding haplotypes using both the underlying sequence reads and the linked-read information encoded in the BX-tag. Unlike traditional callers, *STITCH* imputes genotypes in the presence of missing data based on haplotype information from all sequenced individuals. Following guidelines and informed by pilot *STITCH* runs, we used the following parameters: --*K= 75,--nGen= 100,--niter= 40,--expRate= 0.5,--downsampleToCov   10 --use_bx_tag TRUE*. To optimise computational resources and runtime, we performed *STITCH* with the above parameters on 1 Mbp regions with an overlap of 100 kb overhang allowing them to be combined afterwards. Out of the 11 million sites, 41,396 (0.4%) sites were deemed invariant by *STITCH* (i.e., the *bcftools* and *STITCH* calls disagreed) and were removed, leaving a final set of 11,533,030 SNPs of which 93.9% had an INFO score ≥ 0.8, computed by *STITCH* as a proxy for imputation confidence (Table S3). Moreover, the observed and imputed allele frequencies were highly correlated ($R^2= 0.87$).

Finally, we used the *phase_common_static* from *SHAPEIT5 v 5.1.1* (Hofmeister et al. 2023) to statistically phase genotypes without a reference panel. We polarised alleles in *A. majus* as ancestral or derived using high-coverage PoolSeq sequence data (mean coverage= 89.97×) from multiple populations of the closely related outgroup species *A. molle* (Durán-Castillo et al. 2022). Detailed information on the logic used can be found in the supplementary methods.

3.5 | Genome-Wide Evolutionary Relationships and Demographic Inference

We used three methods to infer genome-wide evolutionary relationships among the sequenced samples. First, we estimated principal components of the genotype matrix. Prior to analysis, we pruned the data set to reduce linkage disequilibrium (LD) between neighbouring SNPs (r^2 threshold of 0.1, window size = 50 SNPs, step size = 10 SNPs). This was done using *Plink v2.0* (Chang et al. 2015) using the command *--indep-pairwise 50 10 0.1*, yielding 1,710,010 SNPs.

We used the model-based clustering program *Admixture v1.3* (Alexander et al. 2009) to assess the genetic structure. *Plink v2.0* was first used to produce BED files from the original VCF file. We ran *Admixture* on the LD-pruned dataset using the unsupervised model for all values of *K* ranging from 2 to 6.

We also inferred a phylogenetic network using the R package *phangorn v2.12* (Schliep 2011). The LD-pruned data set was converted to PHYLIP format using the script *vcf2phylip*. We then calculated a distance matrix from all aligned SNPs using the *dist.ml* function with *model = "JC69"*. The phylogenetic network was then inferred using the *neighborNet* function and drawn with *Splitstree v4.19.1* (Huson and Bryant 2006).

We calculated per-site F_{ST} between each pair of populations on the full SNP data set, using the approach described by Weir and Cockerham (1984), implemented in *vcftools v0.1.16* (Danecek et al. 2011) using the *--weir-fst-pop* flag. Site-based estimates were averaged to obtain a genome-wide estimate.

We estimated gene flow between the two varieties independently in each locality using diffusion approximation, implemented in the software program *δaδi* (Gutenkunst et al. 2010). Since secondary contact is considered the most likely explanation for the current distribution of the two varieties (Tavares et al. 2018), we focused on comparing secondary contact models (SC) with strict isolation models (SI). All models included variation in the ancestral population's effective size prior to population split, following Momigliano et al. (2021). In their basic form, both SI and SC models represent a population split into two populations with specific effective population sizes (N1 and N2) that diverge for a period without gene flow (Ts). In the SC model, these populations then begin exchanging migrants during a secondary contact phase (Tsc), with potentially asymmetric migration (M1 and M2). We expanded these models to account for recent population growth (p1 and p2) and/or Hill–Robertson interference by fitting a genome fraction (P) where the effective population is only a fraction (hrf) of what is found in the rest of the genome. In total, we tested eight distinct models, including four modifications of the SI and SC models: (1) standard model, (2) model with population growth in the daughter populations, (3) a standard model with Hill–Robertson interference and (4) a combined model that included both population growth and Hill–Robertson interference. Each model was fitted 30 times to the data to ensure convergence, and model comparison was performed using the Akaike information criterion (AIC). The importance of gene flow in each locality was then compared by calculating the ratio between Tsc/Ts.

3.6 | Genome-Wide Differentiation, Diversity and Recombination Rate

We calculated F_{ST} in 10 kb windows for each pair of populations using the script *popgenWindows.py* (https://github.com/simonhmartin/genomics_general). Genetic diversity was measured for each site using the *-site-pi* function in *vcftools*.

We used *LDhat v2.2* (Auton and McVean 2007) to calculate the population-scaled recombination rate (ρ) between each SNP, separately for each population. We first used the *lkgen* function in *LDhat* to generate a log-likelihood lookup table for the number of haplotypes in each population, with $\theta = 0.009$ calculated from average genome-wide π in 10 kb windows from a previously published study (Tavares et al. 2018). We then used the *interval* function with the parameters: *-its 10,000,000 -samp 5000 -bpen 5* to estimate variable recombination rates. Finally, we summarised results from the MCMC iterations to estimate mean ρ between each SNP using the *stat* function with the parameters: *--burnin 1000*. *LDhat* was performed on windows of 2000 variants with an overlap of 100 variants at each end and combined afterwards.

3.7 | Genealogical Inference

We used four methods to infer trees from our data. First, we inferred neighbour-joining trees for 50 SNP nonoverlapping windows using the script *phyml_sliding_windows.py* (https://github.com/simonhmartin/genomics_general) with *–minPerInd = 15*.

The second method used was *tsinfer* (Kelleher et al. 2019). We used a custom script to convert phased, polarised SNPs into the *tskit.samples* format, which was then used to infer tree topologies with the *tsinfer.infer* function in the *tsinfer v0.3.2* library (https://github.com/tskit-dev/tsinfer), followed by the *TreeSequence.simplify* function in the *tskit v0.5.8* library (https://github.com/tskit-dev/tskit) to remove unary nodes.

Third, we inferred a tree sequence using *Relate v1.1.8* (Speidel et al. 2019). We assumed $\mu = 5.7 \times 10^{-9}$/bp/generation and a uniform recombination rate of 1 cM. We initially ran Relate separately on each chromosome, setting the haploid N_e to 813,388, as derived earlier from $\pi = 4N_e\mu$ where $\pi = 0.009$. We then used the *EstimatePopulationSize.sh* script to jointly infer a time-varying population size history and branch lengths under that history. For this step, we used a *--threshold 0* to ensure that no trees were excluded in the joint-fitting and *--num_iter = 10*. We also included each population in the argument. Finally, we converted the genealogical trees stored in *.anc* and *.mut* format to *.newick* format with the *RelateExtract –mode AncToNewick* function. We focused our analysis on a five Mbp region around two flower colour loci: *FLAVIA* (locus—Chr2:52650000–54050000; region including locus and the flanking sequence on either side—Chr2:51100000–56100000) and *ROS/EL* (locus—Chr6:52775000–53150000; region including locus and the flanking sequence on either side—Chr2:50500000–55500000). Specific genealogical trees were plotted using a custom script modified from *Treeview.sh* in the *Relate* library. Time to the most recent ancestor (TMRCA) is computed using a custom modified script from the *tskit* library for three subsampled groups: Within all var. *pseudomajus* individuals, within

all var. *striatum* individuals, and between var. *pseudomajus* and var. *striatum* individuals. For each case, TMRCA is first computed for all pairwise combinations of individuals, followed by calculating the median.

Finally, we ran *Singer v0.1.7* (Deng et al. 2024) on 500 kbp genomic windows. For each window, we calculated average π with *VCFtools*, which was then used to calculate N_e from $\pi = 4N_e\mu$. We ran *singer_master* with the parameters: *-m = 5.7e-9,-ratio = 1,-mcmc_iter = 100,-thin = 20,-polar = 0.9*. We then used the function *convert_to_tskit* to convert the last MCMC iteration to *tskit* format and to extract trees in *newick* format.

3.8 | Topology Weighting and Ternary Analysis

Topology weighting was performed on sequences of trees derived from the various genealogical inference methods using *Twisst* (Martin and Van Belleghem 2017). Due to the large number of trees and haplotypes, we followed standard *Twisst* guidelines and limited the topology sampling to 10,000 subtrees using the flag *--method fixed*. Genome-wide topology weights were plotted with loess smoothing (span = 50 kbp). We used the *TwisstNTern* framework (Stankowski et al. 2024) to visualise and calculate asymmetry in the distribution of topology weights for the whole genome and for each chromosome separately using the *--superfine* granularity.

Author Contributions

Conceptualization: A.P., D.L.F., D.S., G.C., N.H.B., S.S., Y.F.C.; Formal analyses: A.L.M., A.P., D.S., S.S., Y.F.C.; Funding acquisition: D.L.F., N.H.B.; Resources: A.J.M., J.K.G., M.K., N.H.B., Y.F.C.; Supervision: G.C., N.H.B., S.S.; Visualisation, writing – original draft preparation: A.P., S.S.; Writing – review and editing: A.L.M., A.P., D.L.F., D.S., G.C., N.H.B., S.S., Y.F.C.; Data curation: A.P.

Ethics Statement

All samples were collected in accordance with the Nagoya Protocol on Access and Benefit sharing (Permit No. 170512ESNC3/ABSCH-IRCC-ES-237652-1). Benefits from this research accrue from the sharing of our data and results on public databases as described above.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Genomic sequence data produced and analysed in this study are deposited at the European Nucleotide Archive (ENA) under the BioProject accession number PRJEB88592. Original code for analyses can be accessed at Github (https://github.com/arka-pal/snapdragon-replicateHZ).

References

Alexander, D. H., J. Novembre, and K. Lange. 2009. "Fast Model-Based Estimation of Ancestry in Unrelated Individuals." *Genome Research* 19, no. 9: 1655–1664. https://doi.org/10.1101/gr.094052.109.

Auton, A., and G. McVean. 2007. "Recombination Rate Estimation in the Presence of Hotspots." *Genome Research* 17, no. 8: 1219–1227. https://doi.org/10.1101/gr.6386707.

Barton, N. H., and K. S. Gale. 1993. "Genetic Analysis of Hybrid Zones." In *Hybrid Zones and the Evolutionary Process*, edited by R. G. Harrison, 13–45. Oxford University Press. https://www.ucl.ac.uk/taxome/lit/Barton&Gale%201993.pdf.

Bolnick, D. I., A. K. Hund, P. Nosil, et al. 2023. "A Multivariate View of the Speciation Continuum." *Evolution* 77, no. 1: 318–328. https://doi.org/10.1093/evolut/qpac004.

Bradley, D., L. Boell, D. Richardson, et al. 2025. "Shaping of Developmental Gradients Through Selection on Multiple Loci in *Antirrhinum*." bioRxiv. https://doi.org/10.1101/2025.02.12.637738.

Bradley, D., P. Xu, I.-I. Mohorianu, et al. 2017. "Evolution of Flower Colour Pattern Through Selection on Regulatory Small RNAs." *Science* 358, no. 6365: 925–928. https://doi.org/10.1126/science.aao3526.

Brandt, D. Y. C., X. Wei, Y. Deng, A. H. Vaughn, and R. Nielsen. 2022. "Evaluation of Methods for Estimating Coalescence Times Using Ancestral Recombination Graphs." *Genetics* 221, no. 1: iyac044. https://doi.org/10.1093/genetics/iyac044.

Burri, R. 2017. "Interpreting Differentiation Landscapes in the Light of Long-Term Linked Selection." *Evolution Letters* 1, no. 3: 118–131. https://doi.org/10.1002/evl3.14.

Burri, R., A. Nater, T. Kawakami, et al. 2015. "Linked Selection and Recombination Rate Variation Drive the Evolution of the Genomic Landscape of Differentiation Across the Speciation Continuum of *Ficedula* Flycatchers." *Genome Research* 25, no. 11: 1656–1665. https://doi.org/10.1101/gr.196485.115.

Campagna, L., M. Repenning, L. F. Silveira, C. S. Fontana, P. L. Tubaro, and I. J. Lovette. 2017. "Repeated Divergent Selection on Pigmentation Genes in a Rapid Finch Radiation." *Science Advances* 3, no. 5: e1602404. https://doi.org/10.1126/sciadv.1602404.

Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee. 2015. "Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets." *GigaScience* 4, no. 1: s13742-015-0047–0048. https://doi.org/10.1186/s13742-015-0047-8.

Danecek, P., A. Auton, G. Abecasis, et al. 2011. "The Variant Call Format and VCFtools." *Bioinformatics* 27, no. 15: 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Danecek, P., J. K. Bonfield, J. Liddle, et al. 2021. "Twelve Years of SAMtools and BCFtools." *GigaScience* 10, no. 2: giab008. https://doi.org/10.1093/gigascience/giab008.

Davies, R. W., J. Flint, S. Myers, and R. Mott. 2016. "Rapid Genotype Imputation From Sequence Without Reference Panels." *Nature Genetics* 48, no. 8: 8. https://doi.org/10.1038/ng.3594.

Deng, Y., R. Nielsen, and Y. S. Song. 2024. "Robust and Accurate Bayesian Inference of Genome-Wide Genealogies for Large Samples." *bioRxiv.* 2024.03.16.585351. https://doi.org/10.1101/2024.03.16.585351.

Durán-Castillo, M., A. Hudson, Y. Wilson, D. L. Field, and A. D. Twyford. 2022. "A Phylogeny of Antirrhinum Reveals Parallel Evolution of Alpine Morphology." *New Phytologist* 233, no. 3: 1426–1439. https://doi.org/10.1111/nph.17581.

Duranton, M., F. Allal, C. Fraïsse, N. Bierne, F. Bonhomme, and P.-A. Gagnaire. 2018. "The Origin and Remoulding of Genomic Islands of Differentiation in the European Sea Bass." *Nature Communications* 9, no. 1: 2518. https://doi.org/10.1038/s41467-018-04963-6.

Fan, C., J. L. Cahoon, B. L. Dinh, et al. 2023. "A Likelihood-Based Framework for Demographic Inference From Genealogical Trees." *bioRxiv.* https://doi.org/10.1101/2023.10.10.561787.

Fan, C., N. Mancuso, and C. W. K. Chiang. 2022. "A Genealogical Estimate of Genetic Relationships." *American Journal of Human Genetics* 109, no. 5: 812–824. https://doi.org/10.1016/j.ajhg.2022.03.016.

Feder, J. L., S. P. Egan, and P. Nosil. 2012. "The Genomics of Speciation-With-Gene-Flow." *Trends in Genetics* 28, no. 7: 342–350. https://doi.org/10.1016/j.tig.2012.03.009.

Field, D. L., S. Stankowski, T. Reiter, et al. 2025. "Genome-Wide Cline Analysis Identifies New Locus Contributing to a Barrier to Gene Flow Across an *Antirrhinum* Hybrid Zone." *bioRxiv.* https://doi.org/10.1101/2025.02.17.638607.

Fontaine, M. C., J. B. Pease, A. Steele, et al. 2015. "Extensive Introgression in a Malaria Vector Species Complex Revealed by Phylogenomics." *Science* 347, no. 6217: 1258524. https://doi.org/10.1126/science.1258524.

Gutenkunst, R., R. Hernandez, S. Williamson, and C. Bustamante. 2010. "Diffusion Approximations for Demographic Inference: DaDi." *Nature Precedings*: 1. https://doi.org/10.1038/npre.2010.4594.1.

Hejase, H. A., Z. Mo, L. Campagna, and A. Siepel. 2022. "A Deep-Learning Approach for Inference of Selective Sweeps From the Ancestral Recombination Graph." *Molecular Biology and Evolution* 39, no. 1: msab332. https://doi.org/10.1093/molbev/msab332.

Hejase, H. A., A. Salman-Minkov, L. Campagna, et al. 2020. "Genomic Islands of Differentiation in a Rapid Avian Radiation Have Been Driven by Recent Selective Sweeps." *Proceedings of the National Academy of Sciences* 117, no. 48: 30554–30565. https://doi.org/10.1073/pnas.2015987117.

Hofmeister, R. J., D. M. Ribeiro, S. Rubinacci, and O. Delaneau. 2023. "Accurate Rare Variant Phasing of Whole-Genome and Whole-Exome Sequencing Data in the UK Biobank." *Nature Genetics* 55, no. 7: 7. https://doi.org/10.1038/s41588-023-01415-w.

Hooper, D. M., C. S. McDiarmid, M. J. Powers, et al. 2024. "Spread of Yellow-Bill-Colour Alleles Favoured by Selection in the Long-Tailed Finch Hybrid System." *Current Biology* 34: 5444–5456.e8. https://doi.org/10.1016/j.cub.2024.10.019.

Hudson, A., J. Critchley, and Y. Erasmus. 2008. "The Genus *Antirrhinum* (Snapdragon): A Flowering Plant Model for Evolution and Development." *Cold Spring Harbour Protocols* 2008, no. 10: pdb.emo100. https://doi.org/10.1101/pdb.emo100.

Huson, D. H., and D. Bryant. 2006. "Application of Phylogenetic Networks in Evolutionary Studies." *Molecular Biology and Evolution* 23, no. 2: 254–267. https://doi.org/10.1093/molbev/msj030.

Jones, F. C., M. G. Grabherr, Y. F. Chan, et al. 2012. "The Genomic Basis of Adaptive Evolution in Threespine Sticklebacks." *Nature* 484, no. 7392: 55–61. https://doi.org/10.1038/nature10944.

Kelleher, J., Y. Wong, A. W. Wohns, C. Fadil, P. K. Albers, and G. McVean. 2019. "Inferring Whole-Genome Histories in Large Population Datasets." *Nature Genetics* 51, no. 9: 9. https://doi.org/10.1038/s41588-019-0483-y.

Li, H. 2013. "Aligning Sequence Reads, Clone Sequences and Assembly Contigs With BWA-MEM." arXiv. https://doi.org/10.48550/ARXIV.1303.3997.

Li, M., D. Zhang, Q. Gao, et al. 2019. "Genome Structure and Evolution of *Antirrhinum Majus* L." *Nature Plants* 5, no. 2: 174–183. https://doi.org/10.1038/s41477-018-0349-9.

Martin, S. H., K. K. Dasmahapatra, N. J. Nadeau, et al. 2013. "Genome-Wide Evidence for Speciation With Gene Flow in Heliconius Butterflies." *Genome Research* 23, no. 11: 1817–1828. https://doi.org/10.1101/gr.159426.113.

Martin, S. H., and S. M. Van Belleghem. 2017. "Exploring Evolutionary Relationships Across the Genome Using Topology Weighting." *Genetics* 206, no. 1: 429–438. https://doi.org/10.1534/genetics.116.194720.

Meier, J. I., P. A. Salazar, M. Kučka, et al. 2021. "Haplotype Tagging Reveals Parallel Formation of Hybrid Races in Two Butterfly Species." *Proceedings of the National Academy of Sciences of the United States of America* 118, no. 25: e2015005118. https://doi.org/10.1073/pnas.2015005118.

Meyer, L., P. Barry, F. Riquet, et al. 2024. "Divergence and Gene Flow History at Two Large Chromosomal Inversions Underlying Ecotype Differentiation in the Long-Snouted Seahorse." *Molecular Ecology* 33, no. 24: e17277. https://doi.org/10.1111/mec.17277.

Momigliano, P., A.-B. Florin, and J. Merilä. 2021. "Biases in Demographic Modelling Affect Our Understanding of Recent Divergence." *Molecular Biology and Evolution*: 38, no. 7: 2967–2985. msab047. https://doi.org/10.1093/molbev/msab047.

Nadeau, N. J., M. Ruiz, P. Salazar, et al. 2014. "Population Genomics of Parallel Hybrid Zones in the Mimetic Butterflies, H. Melpomene and *H. erato.*" *Genome Research* 24, no. 8: 1316–1333. https://doi.org/10.1101/gr.169292.113.

Nielsen, R., A. H. Vaughn, and Y. Deng. 2024. "Inference and Applications of Ancestral Recombination Graphs." *Nature Reviews Genetics* 26: 1–58. https://doi.org/10.1038/s41576-024-00772-4.

Nouhaud, P., S. H. Martin, B. Portinha, V. C. Sousa, and J. Kulmuni. 2022. "Rapid and Predictable Genome Evolution Across Three Hybrid Ant Populations." *PLoS Biology* 20, no. 12: e3001914. https://doi.org/10.1371/journal.pbio.3001914.

Okonechnikov, K., A. Conesa, and F. García-Alcalde. 2016. "Qualimap 2: Advanced Multi-Sample Quality Control for High-Throughput Sequencing Data." *Bioinformatics* 32, no. 2: 292–294. https://doi.org/10.1093/bioinformatics/btv566.

Poelstra, J. W., N. Vijay, C. M. Bossu, et al. 2014. "The Genomic Landscape Underlying Phenotypic Integrity in the Face of Gene Flow in Crows." *Science* 344, no. 6190: 1410–1414. https://doi.org/10.1126/science.1253226.

Rancilhac, L., S. G. de Souza, S. M. Lukhele, et al. 2024. "Introgression Across Narrow Contact Zones Shapes the Genomic Landscape of Phylogenetic Variation in an African Bird Clade." *bioRxiv*, 2024.08.13.607717. https://doi.org/10.1101/2024.08.13.607717.

Ravinet, M., R. Faria, R. K. Butlin, et al. 2017. "Interpreting the Genomic Landscape of Speciation: A Road Map for Finding Barriers to Gene Flow." *Journal of Evolutionary Biology* 30, no. 8: 1450–1477. https://doi.org/10.1111/jeb.13047.

Richardson, D. M., D. Bradley, L. Copsey, et al. 2025. "Genomic Tree Scans Identify Loci Underlying Adaptive Peaks in *Antirrhinum.*" *bioRxiv.* https://doi.org/10.1101/2025.02.12.637406.

Ringbauer, H., A. Kolesnikov, D. L. Field, and N. H. Barton. 2018. "Estimating Barriers to Gene Flow From Distorted Isolation-By-Distance Patterns." *Genetics* 208, no. 3: 1231–1245. https://doi.org/10.1534/genetics.117.300638.

Rueda-M, N., C. Pardo-Diaz, G. Montejo-Kovacevich, et al. 2024. "Genomic Evidence Reveals Three W-Autosome Fusions in Heliconius

Butterflies." *PLoS Genetics* 20, no. 7: e1011318. https://doi.org/10.1371/journal.pgen.1011318.

Schliep, K. P. 2011. "Phangorn: Phylogenetic Analysis in R." *Bioinformatics* 27, no. 4: 592–593. https://doi.org/10.1093/bioinformatics/btq706.

Schwinn, K., J. Venail, Y. Shang, et al. 2006. "A Small Family of MYB-Regulatory Genes Controls Floral Pigmentation Intensity and Patterning in the Genus Antirrhinum." *Plant Cell* 18, no. 4: 831–851. https://doi.org/10.1105/tpc.105.039255.

Shajii, A., I. Numanagić, C. Whelan, and B. Berger. 2018. "Statistical Binning for Barcoded Reads Improves Downstream Analyses." *Cell Systems* 7, no. 2: 219–226.e5. https://doi.org/10.1016/j.cels.2018.07.005.

Shipilina, D., A. Pal, S. Stankowski, Y. F. Chan, and N. H. Barton. 2023. "On the Origin and Structure of Haplotype Blocks." *Molecular Ecology* 32, no. 6: 1441–1457. https://doi.org/10.1111/mec.16793.

Speidel, L., M. Forest, S. Shi, and S. R. Myers. 2019. "A Method for Genome-Wide Genealogy Estimation for Thousands of Samples." *Nature Genetics* 51, no. 9: 1321–1329. https://doi.org/10.1038/s41588-019-0484-x.

Stankowski, S., M. A. Chase, A. M. Fuiten, M. F. Rodrigues, P. L. Ralph, and M. A. Streisfeld. 2019. "Widespread Selection and Gene Flow Shape the Genomic Landscape During a Radiation of Monkeyflowers." *PLoS Biology* 17, no. 7: e3000391. https://doi.org/10.1371/journal.pbio.3000391.

Stankowski, S., and M. Ravinet. 2021. "Defining the Speciation Continuum." *Evolution* 75, no. 6: 1256–1273. https://doi.org/10.1111/evo.14215.

Stankowski, S., Z. B. Zagrodzka, M. D. Garlovsky, et al. 2024. "The Genetic Basis of a Recent Transition to Live-Bearing in Marine Snails." *Science* 383, no. 6678: 114–119. https://doi.org/10.1126/science.adi2982.

Steiniger-White, M., A. Bhasin, S. Lovell, I. Rayment, and W. S. Reznikoff. 2002. "Evidence for "Unseen" Transposase–DNA Contacts." *Journal of Molecular Biology* 322, no. 5: 971–982. https://doi.org/10.1016/s0022-2836(02)00877-x.

Stern, A. J., P. R. Wilton, and R. Nielsen. 2019. "An Approximate Full-Likelihood Method for Inferring Selection and Allele Frequency Trajectories From DNA Sequence Data." *PLoS Genetics* 15, no. 9: e1008384. https://doi.org/10.1371/journal.pgen.1008384.

Surendranadh, P., S. Stankowski, D. L. Field, and N. H. Barton. 2025. "Extreme Long-Range Migration Distorts Flower Colour Clines in an *Antirrhinum* hybrid zone." *bioRxiv.* https://doi.org/10.1101/2025.06.25.661303.

Tarasov, A., A. J. Vilella, E. Cuppen, I. J. Nijman, and P. Prins. 2015. "Sambamba: Fast Processing of NGS Alignment Formats." *Bioinformatics* 31, no. 12: 2032–2034. https://doi.org/10.1093/bioinformatics/btv098.

Tavares, H., A. Whibley, D. L. Field, et al. 2018. "Selection and Gene Flow Shape Genomic Islands That Control Floral Guides." *Proceedings of the National Academy of Sciences* 115, no. 43: 11006–11011. https://doi.org/10.1073/pnas.1801832115.

Todesco, M., G. L. Owens, N. Bercovich, et al. 2020. "Massive Haplotypes Underlie Ecotypic Differentiation in Sunflowers." *Nature* 584, no. 7822: 7822. https://doi.org/10.1038/s41586-020-2467-6.

Vargas, P., J. A. Rosselló, R. Oyama, and J. Güemes. 2004. "Molecular Evidence for Naturalness of Genera in the Tribe Antirrhineae (Scrophulariaceae) and Three Independent Evolutionary Lineages From the New World and the Old." *Plant Systematics and Evolution* 249, no. 3: 151–172. https://doi.org/10.1007/s00606-004-0216-1.

Vijay, N., C. M. Bossu, J. W. Poelstra, et al. 2016. "Evolution of Heterogeneous Genome Differentiation Across Multiple Contact Zones in a Crow Species Complex." *Nature Communications* 7, no. 1: 13195. https://doi.org/10.1038/ncomms13195.

Wakeley, J. 2009. *Coalescent Theory: An Introduction.* Roberts & Co. Publishers.

Wang, S., and G. Coop. 2022. "A Complex Evolutionary History of Genetic Barriers to Gene Flow in Hybridising Warblers." *bioRxiv.* 2022.11.14.516535. https://doi.org/10.1101/2022.11.14.516535.

Weir, B. S., and C. C. Cockerham. 1984. "Estimating F-Statistics for the Analysis of Population Structure." *Evolution* 38, no. 6: 1358. https://doi.org/10.2307/2408641.

Westram, A. M., R. Faria, K. Johannesson, and R. Butlin. 2021. "Using Replicate Hybrid Zones to Understand the Genomic Basis of Adaptive Divergence." *Molecular Ecology* 30, no. 15: 3797–3814. https://doi.org/10.1111/mec.15861.

Whibley, A. C., N. B. Langlade, C. Andalo, et al. 2006. "Evolutionary Paths Underlying Flower Colour Variation in Antirrhinum." *Science* 313, no. 5789: 963–966. https://doi.org/10.1126/science.1129161.

Wilding, C. S., R. K. Butlin, and J. Grahame. 2001. "Differential Gene Exchange Between Parapatric Morphs of *Littorina Saxatilis* Detected Using AFLP Markers." *Journal of Evolutionary Biology* 14, no. 4: 611–619. https://doi.org/10.1046/j.1420-9101.2001.00304.x.

Wohns, A. W., Y. Wong, B. Jeffery, et al. 2022. "A Unified Genealogy of Modern and Ancient Genomes." *Science* 375, no. 6583: eabi8264. https://doi.org/10.1126/science.abi8264.

Wolf, J. B. W., and H. Ellegren. 2017. "Making Sense of Genomic Islands of Differentiation in Light of Speciation." *Nature Reviews Genetics* 18, no. 2: 2. https://doi.org/10.1038/nrg.2016.133.

Wu, C. 2001. "The Genic View of the Process of Speciation." *Journal of Evolutionary Biology* 14, no. 6: 851–865. https://doi.org/10.1046/j.1420-9101.2001.00335.x.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** mec70067-sup-0001-Supinfo.docx.