Supplementary Materials for

Rapid establishment of species barriers in plants compared with that in animals

François Monnet *et al*.

Corresponding author: Camille Roux, camille.roux@cnrs.fr

Science **389**, 1147 (2025)
DOI: 10.1126/science.adl2356

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

Contents

A Materials and Methods

A.1 Animal dataset

The animal data come from the Roux et *al.* (2016) study (*14*). They consist essentially of non-model animal populations/species, initially selected without any particular knowledge about the demographic history, and were sampled from natural populations. These data were produced by RNA sequencing, and only synonymous positions were retained for statistical inferences.

A.2 Plant dataset

Raw data used in this work comes from previously published studies (*43–68*). The following criteria were applied to identify datasets in plants:

i) Currently diploid genomes.

ii) High-throughput sequencing, i.e, RNA-seq, RAD-seq or whole genome sequencing (WGS).

iii) Freely available from `NCBI`.

iv) Individuals sampled from natural populations (geographic distribution represented in Fig. S2).

v) A minimum of two sampled populations/species per genus.

vi) A minimum of two sequenced individuals per sampled population/species.

Datasets fitting these criteria were examined through exploration of literature found *via* Google Scholar (`https://scholar.google.com`), `NCBI` (`https://www.ncbi.nlm.nih.gov/Traces/study/`) and `DDBJ` (`https://ddbj.nig.ac.jp/search`).

Finally, 118 different plant species/populations from 25 different genera were retained for the demographic analysis according to our criteria (Table S1), allowing 280 pairwise demographic analyses to be carried out. These comparisons cover all possible pairs within each genus. No comparisons are made between different genera, with the exception of comparisons within the *Laccospadicinae* (*Howea* and *Linospadix*) due to their relatively small genetic distance.

A.3 Assembly, read mapping and genotype calling

For the plant datasets: reads and metadata were downloaded using SRA-Toolkit, version 2.11.0 (`https://github.com/ncbi/sra-tools/wiki/01.-Downloading-SRA-Toolkit`). Here we separate plant projects for which we worked with synonymous positions (from RNA-seq: n=7 genera and WGS: n=4) from those for which we could not (from RAD sequencing: n=13):

A.3.1 Reads from RNA-seq and WGS.

In line with the animal dataset (*14*), the bioinformatic strategy applied to the plant data is to retain synonymous positions. Reads for a given population/species pair were therefore mapped to a reference transcriptome with the bowtie2 program version 2.4.2 (*69*): either taken from the `1KP project` (*70*) if a species of the same genus is represented there (`https://db.cngb.org/onekp/search/`), or taken from the data associated with the original articles when available (Table SS1). Every position (variants and invariants) were called with a minimum of 8 reads using Reads2SNP 2.0, the uncalled low-quality positions were then coded as "N". The resulting fasta file was used for each population/species as the input file for the demographic inferences.

A.3.2 Reads from RAD-seq.

Loci were assembled for each RAD-seq dataset using Stacks 2.6 (*71, 72*). Combinations of parameters were explored following Paris et *al.* 2017 (*73*) to maximise the amount of biological information retained. Using the two or four samples with the highest amount of available data per lineage, assemblies were built using *denovo_map.pl* (Stacks) with different combinations of parameters: the minimum depth for a stack to be valid (-*m*, ranging from 3 to 5), the number of mismatches allowed between stacks within individuals (-*M*, ranging from 1 to 6) and the number of mismatches allowed between stacks between individuals (-*n*, set to M or $M + 1$), for a total of 36 combinations. In addition, loci that were missing in at least 20% of the samples per population were withdrawn with the argument *–min-samples-per-pop 0.80* (i.e. only loci with the information for all samples were kept, as populations were composed of two or four samples). The number of polymorphic loci was plotted as a function of the different combinations of parameters using a homemade `R` script. For each dataset, a combination was selected in function of the trade-off between maximising the

number of polymorphic loci and minimising the parameter values to produce a reference set of loci for each species/population pair. Reads were mapped on this reference with bowtie2 version 2.5.1, and variants were called with Reads2SNP 2.0 in the same way as "RNA-seq and WGS" datasets.

A.4 Demographic inferences

A.4.1 Choice of the inferential method

Several methods exist to test introgression between evolutionary lineages, which can broadly be categorized into population genomic approaches (e.g., FastSimCoal (*74*), $\partial a \partial i$ (*75*), Moments (*76*), DILS (*11*)) and phylogenetic methods (e.g., QuIBL (*22*), Dsuite (*77*)). All these methods are effective within the contexts for which they were designed. The goal of this section is not to demonstrate the superiority of one approach over another but to justify why DILS was chosen as a relevant tool for addressing the specific question posed in this study: whether there is ongoing gene flow between two populations or whether they are currently isolated. The choice of method also depends on the properties of the sampling (i.e., number of sampled individuals, number of sampled populations in the genus, assumptions about demography for certain population pairs, etc.) as well as on the properties of the molecular data (i.e., locus length, presence or absence of intra-locus recombination, etc.). Finally, we also aimed to reduce biases that arise when linked selection is not taken into account, whether from background selection (*78*) or selection against species barriers (*79*). In the next section, we focus on a comparison between QuIBL and DILS because the literature has shown that FastSimCoal and $\partial a \partial i$ produce results largely comparable to DILS when applied to the same datasets (*80, 81*).

A.4.2 Key methodological differences between DILS and QuIBL

In a given phylogenetic tree with more than three lineages (species or populations), QuIBL extracts information from individual gene trees and tests, for a given triplet of lineages, whether the distribution of internal branch lengths can be explained solely by incomplete lineage sorting or whether introgression processes must also be invoked. For a tree of the topology $((1, 2), 3)$, QuIBL is particularly powerful in detecting deviations from strict allopatry between lineages 1 (or 2) and 3. Such deviations are interpreted as evidence of historical introgression events, which may vary in

recency.

In contrast, DILS leverages population genomic summary statistics, particularly those derived from the Site Frequency Spectrum (SFS), to evaluate whether observed patterns are better explained by a model with ongoing migration or one without, using an Approximate Bayesian Computation (ABC) framework. Given that the question addressed here concerns the extent of current reproductive isolation between specific pairs of lineages (in plants and animals), it was important to discriminate between different temporal patterns of introgression. Deviations from strict allopatry caused by ancient migration do not convey the same implications for current reproductive isolation as those caused by secondary contact. This distinction is explicitly addressed by DILS through the analysis of intra-specific polymorphism and interspecific divergence data for a given pair of lineages.

A.4.3 Rationale for Choosing DILS

Our decision to use DILS was motivated by the following considerations:

- DILS accounts for variations in effective population size over time, whereas phylogenetic approaches assume a constant N_e.
- DILS accounts for genome-wide variations in effective population size caused by background selection, as well as variations in effective migration rates driven by linked selection against species barriers.
- DILS accounts for intra-locus recombination, while the distribution of internal branch lengths is derived under the assumption of a simple tree per locus, with no recombination occurring within individual loci.
- DILS does not require a specific sampling scheme to test introgression, whereas phylogenetic methods cannot detect gene flow between lineages 1 and 2 in a tree of the topology $((1, 2), 3)$.
- DILS works directly with measurable quantities derived from sequences (e.g., F_{ST}, θ_W, π, D_a, D_{xy}, etc.) without requiring a preliminary step of inferring phylogenetic trees for each locus.

- DILS can handle multiple individuals within each species, reducing the risk of overlooking shared polymorphisms between species.
- DILS does not require long loci for accurate inference, making it less sensitive to biases introduced by RADseq or RNAseq data. In contrast, phylogenetic methods may perform poorly with short loci and are subject to violations of no-recombination assumptions with long loci (*82, 83*). In addition, very long loci may buffer local genomic variations for N_e and $N_e.m$.
- DILS works directly on pairs of lineages and does not require data from more than three species, allowing its application to genera where only two species have been sequenced.

A.4.4 Comparison between DILS and QuIBL on simulated data

To further illustrate the relative strengths of these approaches, we conducted a comparative analysis of DILS and QuIBL using coalescent simulations under a four populations model (Fig. S9). Data were simulated under various temporal patterns of introgression using `msnsam` (*84*), a modified version of `ms` (*85*). The simulations were designed to satisfy all assumptions made by QuIBL (e.g., long loci, no intra-locus recombination, no migration between lineages 1 and 2 in a tree of the topology $((1, 2), 3)$). For QuIBL, the input consisted of exact, simulated gene trees rather than inferred trees, as would be the case with real biological data.

Simulated Scenarios Four demographic scenarios were simulated (Fig. S9):

- **Strict Isolation ($\mathbf{SI_{4pop}}$)**: No migration.
- **Isolation-Migration ($\mathbf{IM_{4pop}}$)**: Migration between lineages 2 and 3 from the present to T_4.
- **Ancient Migration ($\mathbf{AM_{4pop}}$)**: Migration between lineages 2 and 3 between T_{dem} and T_4.
- **Secondary Contact ($\mathbf{SC_{4pop}}$)**: Migration between lineages 2 and 3 from the present to T_{dem}.

Migration was modeled symmetrically at a rate $M = N_e.m$, where N_e is the effective population size and m is the proportion of migrants in each generation.

The parameters explored included:

- T_4: Speciation time between lineages 1 and 2 ($0.5, 2, 4\, N_e$ generations).
- T_3: Speciation time between $(1, 2)$ and 3 ($T_4 + 0.5, 2, 4\, N_e$ generations).
- T_2: Speciation time between $((1, 2), 3)$ and 4 ($40\, N_e$ generations).
- T_{dem}: Migration onset/end time ($T_4 \cdot [0.1, 0.25, 0.5]$ for $\text{SC}_{4\text{pop}}$; $T_4 \cdot [0.5, 0.75, 0.9]$ for $\text{AM}_{4\text{pop}}$).
- $M (= N_e.m)$: Number of migrants per generation (0.25, 10 migrants).
- L_{mig}: Number of loci affected by migration (100, 500, 1000 loci out of 1,000 total loci).

Migration is assumed to be symmetric, occurring at a rate M corresponding to $N_e.m$, where N_e is the effective population size and m is the proportion of individuals in a population consisting of migrants in each generation. We simulate 1,000 independent loci, with no intra-locus recombination. Among these loci, a subset L_{mig} is influenced by migration, while the remaining loci represent genetic isolation between the species.

It is important to note that the assumptions of a constant N_e over time and across the genome, as well as the absence of intra-locus recombination, are not required by DILS. These constraints are therefore relaxed in the empirical application to plants *versus* animals. However, they are implemented in the simulation study to enable a direct comparison between DILS and QuIBL, as they align with the assumptions required by QuIBL.

Simulations for QuIBL and DILS were performed using the same parameter combinations. Differences lay in sampling schemes: for QuIBL, a single copy per locus was sampled from lineages 1, 2, 3, and 4, with exact gene trees simulated and rooted using lineage 4 (Fig. S9). For DILS, two diploids were sampled from lineages 2 and 3, reflecting the minimal sampling scheme when DILS was applied to the empirical plants *versus* animals dataset.

Simulations were generated using the script `simulations_for_QuIBL_DILS_full.py`. The command lines for simulating data using `msnsam` under different demographic models are as follows:

SI_{4pop}:

- **QuIBL:** `msnsam 4 1000 -T -I 4 1 1 1 1 0 -ej tbs 2 1 -ej tbs 3 1 -ej 10 4 1`

- **DILS:** `msnsam 8 1000 -t 40 -I 4 0 4 4 0 0 -ej tbs 2 1 -ej tbs 3 1 -ej 10 4 1`

AM$_{4pop}$**:**

- **QuIBL:** `msnsam 4 1000 -T -I 4 1 1 1 1 0 -em tbs 2 3 tbs -em tbs 3 2 tbs -eM tbs 0 -ej tbs 2 1 -ej tbs 3 1 -ej 10 4 1`
- **DILS:** `msnsam 8 1000 -t 40 -I 4 0 4 4 0 0 -em tbs 2 3 tbs -em tbs 3 2 tbs -eM tbs 0 -ej tbs 2 1 -ej tbs 3 1 -ej 10 4 1`

IM$_{4pop}$**:**

- **QuIBL:** `msnsam 4 1000 -T -I 4 1 1 1 1 0 -m 2 3 tbs -m 3 2 tbs -eM tbs 0 -ej tbs 2 1 -ej tbs 3 1 -ej 10 4 1`
- **DILS:** `msnsam 8 1000 -t 40 -I 4 0 4 4 0 0 -m 2 3 tbs -m 3 2 tbs -eM tbs 0 -ej tbs 2 1 -ej tbs 3 1 -ej 10 4 1`

SC$_{4pop}$**:**

- **QuIBL:** `msnsam 4 1000 -T -I 4 1 1 1 1 0 -m 2 3 tbs -m 3 2 tbs -eM tbs 0 -ej tbs 2 1 -ej tbs 3 1 -ej 10 4 1`
- **DILS:** `msnsam 8 1000 -t 40 -I 4 0 4 4 0 0 -m 2 3 tbs -m 3 2 tbs -eM tbs 0 -ej tbs 2 1 -ej tbs 3 1 -ej 10 4 1`

The QuIBL analyses were performed using the following options:

- `-numdistributions: 2`
- `-likelihoodthresh: 0.01`
- `-numsteps: 10`
- `-gradascentscalar: 0.5`
- `-multiproc: False`

- `-maxcores: 1000`
- `-totaloutgroup: 4`

For the DILS inferences on the simulated datasets, we used the following parameters: assuming a constant population size (`population_growth: constant`), a mutation rate μ of 2×10^{-8} mutations per generation per nucleotide (`mu: 0.00000002`), no intra-locus recombination (`rho_over_theta: 0`), and an effective population size N_e uniformly explored between 0 and 200,000 individuals. The split time was uniformly explored between 0 and 1,500,000 generations, and migration ($N_e.m$) was uniformly explored between 0.2 and 10 migrants per generation. No outgroup was used to polarize mutations (`nameOutgroup: NA`).

The final output file, `table_QuIBL_DILS.txt`, summarizes the demographic parameters and the inference results obtained from DILS and QuIBL for each simulated dataset.

Results of inferences performed with DILS and QuIBL on simulated datasets When the simulated model represents current isolation (SI_{4pop} and AM_{4pop}), both methods converge on the correct model in approximately 75% of simulations across all explored parameter combinations (Fig. S10-A, left panel). DILS performs slightly better than QuIBL, with around 14% of datasets where it is the only method to correctly recover the isolation model, compared to approximately 6.5% for QuIBL. However, these performances in supporting isolation are of the same general magnitude for both methods. For the parameter combinations used in these simulations, both methods fail to recover the isolation model in about 4.5% of cases.

The differences between the two methods become more pronounced when the true model involves ongoing migration (IM_{4pop} and SC_{4pop}). In this scenario, both methods converge on the correct model in approximately 36% of simulations under the explored parameter combinations (Fig. S10-A, right panel). QuIBL alone identifies migration in only 0.001% of simulations, while DILS is the sole method to correctly infer migration in about 29% of cases. This analysis suggests that both methods are highly conservative, failing to detect migration in 35% of simulations with ongoing gene flow. However, it is important to note that these results stem from discrete parameter combinations (T_4, T_3, T_{dem}, M, and L_{mig}), where migration can be both low ($N_e.m = 0.1$) and affecting only a small portion of the genome (10% of loci).

When we focus on simulations with higher levels of introgression (Fig. S10-B, right panel), the performance of both methods improves significantly. In these cases, the two methods converge on the correct model with ongoing migration in approximately 72% of simulations. The remaining simulations with ongoing migration are correctly recovered solely by DILS, and no simulations are simultaneously missed by both methods.

Our choice to use DILS in this study was primarily driven by its demonstrated reliability in previous analyses and its ease of use for interpreting temporal patterns of introgression via explicit scenario testing. Unlike QuIBL, which excels in handling large phylogenies, DILS is particularly well-suited for datasets with simpler sampling schemes, such as cases where only two species are available for analysis. Our comparison confirms that our initial choice of DILS did not come at the expense of reduced performance. This validation, combined with DILS's interpretability and suitability for the scope of our study, highlights its relevance for addressing the research questions we aimed to explore.

A.4.5 DILS

Model comparisons were carried out using the approximate Bayesian computation (ABC) framework applied in the animal study (*14*) and distributed under the name DILS (Demographic Inferences with Linked Selection (*11*)). DILS aims to test whether the studied species are connected by gene flow. It incorporates the effects of linked selection by modeling background selection as heterogeneous effective population sizes (hetero-N_e) across the genome and selection associated with barriers to gene flow as genomic heterogeneity in effective introgression rates (hetero-$N_e.m$). Here we describe how DILS works.

A.4.6 Compared models

The primary objective of our demographic analysis is to determine which two populations historical scenario explain the best a given dataset. The term dataset here refers to a pair of populations/species (comprising either two animal or two plant lineages), for which genomic data are described by an array of summary statistics (see section A.4.7). In our ABC methodology, we discern two categories of models.

Demographic Models: Each of the four models models describes the subdivision of an ancestral

population into two daughter populations (Fig. S5-A). The three populations have independently assigned effective population sizes. The differences between these four models concern the historical patterns of gene flow between two divergent populations, as depicted in figure S5. These models encompass continuous migration (CM), and secondary contact (SC), strict isolation (SI) and ancient migration (AM) :

- models with ongoing migration
  - continuous migration (CM)
  - secondary contact (SC)
- models with current isolation
  - strict isolation (SI)
  - ancient migration (AM)

Notably, the former two models entail ongoing gene flow between the two populations, while the latter two do not. Models with past (AM) or recent (CM and SC) migration assume gene flow between sister populations/species in both directions, at two independently assigned rates.

Models of Linked Selection: Effects of linked selection have been taken into account using a genomic model that encompasses: (a) heterogeneous effective population size across the genome (*hetero.* N_e), which closely approximates the influence of background selection by down-scaling N_e (*86*); and/or (b) heterogeneous migration rate across the genome (*hetero. M*) to account for the effects of selection against hybrids (*87*). The modeling framework employed in this study does not consider the effects of positive selection on linked loci (i.e., genetic hitchhiking).

Within the *hetero.* N_e genomic model, the variable effective size among loci is assumed to conform to a re-scaled Beta distribution. In essence, all populations share a common Beta distribution with two shape parameters drawn from uniform distributions. However, each population is independently re-scaled by distinct N_e values, which are drawn from uniform distributions. Conversely, the *homo.* N_e genomic model assumes that all loci from the same genome share the same effective population size, and this parameter is independently estimated in all populations. This homogeneous model implies that the genomic landscape remains unaffected (or is uniformly affected) by background selection.

The *hetero. M* genomic model implements local reduction of gene flow in the genome. Variation in migration rates among loci is thus modeled by employing a bimodal distribution where a proportion of loci, drawn from a uniform distribution in]0-1[, is linked to barriers (i.e., $N_e.m = 0$), while the loci unaffected by species barriers are associated to an effective migration rate $N_e.m$ drawn from a uniform distribution. In the *homo. M* model, a single migration rate $N_e.m$ per direction is universally shared by all loci in the genome.

Subdivisions of the four demographic models (CM, SC, SI and AM) into various genomic submodels were made to accommodate for the effect of linked selection. Heterogeneity in effective population size was a universal consideration across all four models, while heterogeneity in migration rate was specifically accounted for in models exhibiting gene flow (i.e., CM, AM, and SC). Therefore, the SI model was divided into two submodels (*homo. N* or *hetero. N*), while the AM, CM, and SC models were divided into four submodels:

i) *homo.* N_e and *homo. M*

ii) *homo.* N_e and *hetero. M*

iii) *hetero.* N_e and *homo. M*

iv) *hetero.* N_e and *hetero. M*

For a comprehensive description of all prior distributions employed in this study, please refer to Section A.4.8.

A.4.7 Summary statistics

ABC is a statistical inferential approach based on the comparison of summary statistics derived from simulated and observed datasets (*88*). We present a comprehensive description of the statistics computed within our framework. Previous studies have demonstrated the effectiveness of these statistics in statistically distinguishing demographic models with and without ongoing migration (*14, 26, 79*). The following summary statistics are calculated for each locus:

- The number of bi-allelic polymorphisms in the alignment including all sequenced copies in the 2 species/populations

- Pairwise nucleotide diversity π (*89*)
- Watterson's θ (*90*)
- Tajima's D (*91*)
- The proportion of sites displaying fixed differences between the populations/species (S_f)
- The proportion of sites featuring polymorphisms exclusive to a specific population/species (S_{xA} and S_{xB})
- The fraction of sites with polymorphisms shared between the two populations/species (S_s)
- The number of successive shared polymorphic sites
- Raw divergence D_{xy} between the two populations/species (*92*)
- Net divergence D_a between the two populations/species (*92*)
- Relative genetic differentiation between the two populations/species quantified by F_{ST} (*93*)

For the ABC analysis, we used the means and variances of these statistics calculated over all the available loci. Additionally, we utilize the joint Site Frequency Spectrum (jSFS (*94*)) to summarize the data, specifically capturing the count of single-nucleotide polymorphisms (SNPs) where the minor allele occurs in each bin covering the jSFS. Because of the absence of outgroup lineages, jSFS were folded. Singletons are deliberately excluded from the jSFS to mitigate potential inference biases arising from sequencing errors. Each of the non-excluded bin of the jSFS is used as a descriptive statistics in the ABC analysis.

We supplement this set of summary statistics with measures taken on all the loci:

- Pearson's correlation coefficient for π between species
- Pearson's correlation coefficient for θ between species
- Pearson's correlation coefficient between D_{xy} and D_a
- Pearson's correlation coefficient between D_{xy} and F_{ST}

- Pearson's correlation coefficient between D_a and F_{ST}
- Proportion of loci with both S_s and S_f sites
- Proportion of loci with S_s sites but no S_f
- Proportion of loci without S_s sites but with S_f
- Proportion of loci with neither S_s nor S_f sites

The summary statistics obtained from both the empirical data sets (i.e., plants and animals) and the data sets simulated under the demographic models (Fig. S5) were calculated with the same scripts implemented in DILS.

A.4.8 Configuration file

DILS was run using the following parameter values:

- mu = 7.31×10^{-9}
- useSFS = 1
- barrier = bimodal
- $max\_N\_tolerated = 0.25$
- $L_{min} = 10$
- $n_{min} = 4$
- $rho\_over\_theta = 0.2$
- uniform prior for N_e between 0 and $N_{e,max}$ individuals
- uniform prior for T_{split} between 0 and T_{max} generations
- uniform prior for migration rate $N_e.m$ between 0 and 10 migrants per generations

Where:

$$N_{e,max} = 5 \times max\left(\frac{\pi_A}{4\mu}; \frac{\pi_B}{4\mu}\right)$$

π_A and π_B being the Tajima's θ (*89*) for species A and B respectively (for a given pair).

$$T_{max} = 5 \times \frac{D_a}{2\mu}$$

D_a being the net divergence (*92*).

A.4.9 Returned quantities

At the end of the analysis, DILS returns the posterior probability of ongoing migration *versus* of current isolation. The probability of ongoing migration corresponds to the relative probability of all models including ongoing migration (Secondary Contact, Continuous Migration) and their sub-models (heterogeneity and genomic homogeneity for migration and effective size); while the probability of current isolation corresponds to all models and sub-models with current isolation (Strict Isolation, Ancient Migration). These quantities are used to produce the relationships between the net divergence and the posterior probability of migration (Fig. S11). For each pair of populations/species, three statuses are then assigned:

i) Strong support for genetic isolation: we identify strong statistical support for genetic isolation when our ABC framework yields a posterior probability $P_{\text{mig}} < 0.1304$. This threshold was empirically determined by the robustness test conducted in (*14*), where the robustness R_{mig} was calculated as:

$$R_{\text{mig}} = \frac{P(P_{\text{mig}} = P_i \mid \text{mig})}{P(P_{\text{mig}} = P_i \mid \text{mig}) + P(P_{\text{mig}} = P_i \mid \text{iso})} \quad \text{(S1)}$$

where:

- P_i: the posterior probability attributed by our ABC framework to migration to a given simulated dataset.
- $P(P_{\text{mig}} = P_i \mid \text{mig})$ is the probability that a dataset simulated under a model with ongoing migration is correctly inferred as a model with migration by our ABC approach, given a posterior probability for migration of P_i.

- $P(P_{\text{mig}} = P_i \mid \text{iso})$ is the probability that a dataset simulated under a model with current isolation is wrongly inferred as a model with migration by our ABC approach, given a posterior probability for migration of P_i.

ii) Strong support for ongoing migration: strong statistical support for ongoing migration is indicated when the posterior probability $P_{\text{mig}} > 0.6419$, also empirically determined in (*14*).

iii) Ambiguity: statistical ambiguity, denoting situations where our ABC framework does not strongly support either migration or isolation, i.e, when the risk of assigning an analysed pair to a wrong status is greater than 5%.

Pairs for which support was inconclusive were excluded from further analysis. The remaining pairs were categorized either as exhibiting 'migration' or 'isolation,' as illustrated in Fig. 1-A, allowing the 'migration' status to be treated in a logistic regression (see section A.5).

A.5 Logistic regression

To study speciation dynamics, we examine reduction in the proportion of plant or animal pairs receiving strong support for models with migration as a function of time (measured here by the net molecular divergence). For this purpose, we modeled $\mathbf{Y}_i$ (the binary status 'isolation' or 'migration' best fitting the data) as a function of $\mathbf{X}_i$ (the average net genomic divergence) by using a generalized linear model (GLM) *via* a linked binomial function:

$$g\left(\mathbb{E}(Y_i|\mathbf{X}_i)\right) = g(\mu_i) = \mathbf{X}_i\boldsymbol{\beta} = \beta_0 + \beta_1 X_{1,i}$$

where β_0 represents the intercept and β_1 the coefficient reflecting the effect of genomic divergence on the isolation/migration status coded as 0 and 1, respectively. The fitted model is used to predict p_i, the proportion of pairs of populations/species that are currently connected by gene flow (migration status) for a given level of divergence $\mathbf{X}_i$.

$$p_i = \frac{\exp\left(\mathbf{X}_i\boldsymbol{\beta}\right)}{1 + \exp\left(\mathbf{X}_i\boldsymbol{\beta}\right)} = \frac{1}{1 + \exp\left(-\mathbf{X}_i\boldsymbol{\beta}\right)}$$

Reversely, we can determine the divergence level $\mathbf{X}$ for which a given proportion p_i of pairs are connected by gene flow:

$$X = -\frac{1}{2\beta_1}\left(\beta_0 + \sqrt{\beta_0^2 + 4\beta_1 \log\left(\frac{p_i}{1-p_i}\right)}\right)$$

We are interested in comparing the inflection point, i.e, the level of divergence above which more than 50% of species pairs are genetically isolated, between plants and animals. Thus, for a given fitted model, this point corresponds to a divergence level $X_{p=0.5} = -\frac{\beta_0}{\beta_1}$.
The log-likelihood function ℓ of the migration/isolation status $\mathbf{Y}$ given the average net molecular divergence $\mathbf{X}$ is then obtained to evaluate the fit of a model to the observed data:

$$\begin{aligned}
\ell(\boldsymbol{\beta}|\mathbf{Y}=\mathbf{y},\mathbf{X}=\mathbf{x}) &= \log\left(\mathcal{L}(\boldsymbol{\beta}|\mathbf{Y}=\mathbf{y},\mathbf{X}=\mathbf{x})\right) \\
&= \sum_{i=1}^{N}\left[y_i \log(p_i) + (1-y_i)\log(1-p_i)\right] \\
&= \sum_{i=1}^{N}\left[y_i \log\left(\frac{p_i}{1-p_i}\right) + \log(1-p_i)\right] \\
&= \sum_{i=1}^{N}\left[y_i \cdot \mathbf{x}_i\boldsymbol{\beta} - \log(1+\exp\left(\mathbf{x}_i\boldsymbol{\beta}\right))\right] \\
&= \sum_{i=1}^{N}\left[y_i \cdot (\beta_0 + \beta_1 X_{1,i}) - \log(1+\exp\left(\beta_0 + \beta_1 X_{1,i}\right))\right] \qquad (1)
\end{aligned}$$

The sigmoid of plants can then be tested against that of animals to determine whether plants and animals share the same dynamics of reproductive isolation accumulation. For this purpose, three models are fitted and associated to log-likelihood ℓ:

i) M_0: both plants and animals share the same logistic relationship between $\mathbf{X}_i$ and $\mathbf{Y}_i$.

ii) M_{plants}: model fitted to the plants data only.

iii) $M_{animals}$: model fitted to the animals data only.

Thus, for M_0 we fitted a GLM to the entire dataset comprising both plants and animals, after having retained only demographic inferences for which the ABC analysis produced strong statistical support for ongoing migration or current isolation, following the test of robustness applied in Roux et al. (*14*). In that sense, pairs of plants and animals with ambiguous support for isolation or

migration were excluded from all GLM regressions. The log-likelihood $\ell(M_0)$ was then estimated for the whole dataset comprising both plants and animals by using formula **1** where:

- $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$ represent for M_0 the coefficient of the model fitted to the whole plants and animals dataset by using the `glm` function (family = 'binomial') implemented in `R`.
- $\mathbf{X}_{1,i}$ represents the series of observed net divergence values.
- $\mathbf{y}_i$ represents the series of inferred isolation/migration status.

For M_{plants} and $M_{animals}$, we fitted a GLM model only to data from the corresponding kingdom. We then estimated the log-likelihoods $\ell(M_{plants})$ and $\ell(M_{animals})$ as for M_0.

Finally, we conducted a comparison between the log-likelihood $\ell(M_0)$ and the combined log-likelihood $\ell(M_{plants}) + \ell(M_{animals})$, which is derived from the summation of log-likelihoods obtained by fitting independent models to each respective kingdom. The significance of the difference between $\ell(M_0)$ and $\ell(M_{plants}) + \ell(M_{animals})$ was evaluated using a permutation-based approach. Specifically, the absolute difference between $\ell(M_0)$ and $\ell(M_{plants}) + \ell(M_{animals})$ was compared to a null distribution generated from 10,000 random permutations of the data. The *P*-value corresponds to the proportion of permutations in which the absolute difference exceeded the observed value (Table S2).

A.6 Controls for methodological and biological biases

To ensure the robustness of our main finding—that plants experience a faster cessation of gene flow compared to animals—we performed additional control analyses addressing three potential sources of bias: sequencing methodology, unequal representation of genera, and geographic distances between lineages. Each of these analyses confirmed that our conclusions are unlikely to result from methodological artifacts or sampling biases, as detailed in the sections below.

A.6.1 Testing for a sequencing technology effect

Out of the total dataset comprising 280 pairs of plants and 61 pairs of animals, 183 plant pairs and 54 animal pairs exhibited high robustness in model comparison and passed the goodness-of-fit test. These retained datasets encompass a diversity of sequencing methodologies. Specifically,

within plants, among the 183 retained pairs: 81 pairs were acquired through RAD-sequencing, 70 pairs through RNA-sequencing, and 32 pairs through whole genome sequencing. In the case of animals: 46 pairs were derived from RNA-sequencing, while 8 pairs were the result of Sanger sequencing. To assess the potential influence of sequencing techniques, we determined whether the observed differences in dynamics between plants and animals, as previously reported for the entire dataset, remained consistent when considering only the data generated exclusively through RNA sequencing. This choice was motivated by the fact that RNA-sequencing is the sole sequencing technique shared by both biological kingdoms under study. By retaining only the data from RNAseq, we maintain a statistically significant support for a more rapid cessation of gene flow in plants than in animals (P-value$<$ 0.0001; Table S3).

A.6.2 Testing for a phylogenetic effect

To control for the variation in the number of pairs between genera, we carried out 1,000 animal-plant comparisons as for Fig. 1 but by randomly selecting a single pair per animal and plant genus (Fig. S1-A). Over these 1,000 sub-samples, the relative positions of the sigmoids were compared *via* the inflection points ($X_{p=0.5} = -\frac{\beta_0}{\beta_1}$) of the models fitted to the plant *versus* animal sub-samples. We consistently find that the inflection point occurs at lower levels of divergence in plants than in animals (Fig. S1-B). However, the difference in likelihood between the global model and the combined likelihoods of the plant and animal models is not significant in 7% of the permutations (Fig. S1-C; blue bars). This lack of significance is attributed to the sub-sampling process, which substantially reduces the number of data points contributing to the likelihood calculations, thereby decreasing the power to discriminate effectively between models. Nevertheless, the distance between the plant and animal inflection points was significantly greater than expected by random chance in 100% of the permutations (Fig. S1-C; orange bars).

A.6.3 Testing for a geographic effect

Geographical (geodesic) distance in meters was measured using GPS coordinates provided in the metadata when available, using the `distGeo` function in the `R` package *geosphere*. For a given pair of populations/species A and B, this distance corresponds to the distance between the two geographically closest individuals. In the case of sampled sympatric pairs, and if a single coordinate

was provided by the authors for all individuals A and B, we consider a distance of 10m in line with current sampling practices to reduce relatedness. Among the 25 plant genera under examination, our review of the literature has not yielded information pertaining to the geographical origins of specimens from *Gossypium*.

To test whether the observed pattern of reduced gene flow in plants compared to animals could be merely explained by greater geographic distances between plant pairs, we performed logistic regression analyses separately for plants and animals. Using the GPS coordinates of sampled individuals, we calculated the minimum geographic distance (`distances_meters`) between taxa within each pair.

We then modeled the probability of ongoing migration (`P_ongoing_migration`) as a function of geographic distance and net divergence (`netdiv_avg`) using binomial logistic regression with a logit link function. The logistic regression results showed that geographic distance (`distances_meters`) was not a significant predictor of ongoing migration for either animals ($P = 0.688$) or plants ($P = 0.937$), confirming that our conclusions are not driven by geographic factors.

A.7 Testing factors influencing speciation dynamics within plants

To investigate factors that may influence the dynamics of reproductive isolation in plants, we focused on two life-history traits: (i) growth form, specifically comparing herbs and trees (Section A.7.1), and (ii) mating systems by comparing different selfing rates (Supplementary Materials A.7.2). The first comparison is motivated by the typically greater pollen dispersal observed in trees compared to annual plants, leading to reduced nuclear genetic differentiation within tree populations (*95, 96*). For the effect of the mating system, we limited our analysis to a comparison between species in our sample with the lowest selfing rates and those with the highest. However, it is important to note that our sample is not representative of the full diversity of selfing rates found in plants (*97*), as it is biased towards high-outcrossing species (Fig. S13). The theoretical effect of selfing rate on the evolution of reproductive isolation remains unclear and appears contradictory (*98*). On one hand, higher selfing rates reduce gene flow, decrease effective recombination, and increase the strength of genetic drift, which can facilitate the fixation of deleterious mutations that may act as barriers when compensated. On the other hand, reduced efficacy of selection in high selfing species can mitigate intragenomic conflicts, potentially leading to fewer interspecific incompatibilities.

For these analyses, we focused on the plant dataset, subdivided into different groups based on the specific comparisons performed for each tested factor. We first conducted a comparison based on life form (herbs *versus* trees), followed by three comparisons to assess the effect of selfing rate (pairs of selfers *versus* pairs of outcrossers, pairs of selfers *versus* pairs with different systems, and pairs of outcrossers *versus* pairs with systems).

Each comparison was performed similarly to the plant-*versus*-animal comparison: we first fitted a global model on the entire dataset and then tested whether the sum of the likelihoods of the models fitted on each subgroup was significantly greater than that of the global model.

A.7.1 Effects of plant life forms

The plant dataset was divided into four categories based on their life form:

- **Liana** (e.g., *Actinidia*, *Nepenthes*)
- **Herb** (e.g., *Arabis*, *Dactylorhiza*, *Helianthus*, *Hibiscus*, *Isoetes*, *Lupinus*, *Pitcairnia*, *Pulmonaria*, *Rhodanthemum*, *Senecio*, *Silene*, *Phlox*)
- **Tree** (e.g., *Ficus*, *Laccospadicinae*, *Phoebe*, *Picea*, *Populus*, *Quercus*, *Yucca*)
- **Shrub** (e.g., *Gossypium*, *Salix*, *Stachyurus*)

We performed two comparisons: first, between herbs and trees (Fig. S3-A), which showed no significant reduction in gene flow among herbs compared to trees (P = 0.261; Table S4). When herbs, lianas, and shrubs were grouped together (Fig. S3-B), we still did not observe a reduction in gene flow relative to trees (P = 0.1937; Table S4).

A.7.2 Effects of plant mating systems

In line with (*99*), we aimed to test whether the mating system influences the dynamics of introgression. More specifically, we tested whether selfing species exhibit a faster emergence of reproductive isolation compared to outcrossing species, as hypothesized due to reduced dispersal, decreased effective recombination, and a higher rate of accumulation of genetic incompatibilities. Alternatively, selfing species might experience a slower development of reproductive isolation compared to outcrossers, driven by reduced intragenomic conflicts that could otherwise act as barriers.

To investigate whether the extent of inbreeding influences the rate of reproductive barrier establishment, we categorized plant species according to their selfing rates using three independent sources of information: two quantitative and one qualitative.

i) First, we developed a custom estimator, hereafter referred to as `selfing_ML`, which infers the selfing rate s using a maximum likelihood approach. This method models the probability of observing a genotype given the allele frequency at each polymorphic position and a specified selfing rate, under the assumption of Hardy–Weinberg equilibrium modified for self-fertilization.

ii) Second, we used the R package `inbreedR` (*23*), which estimates inbreeding based on multilocus identity disequilibrium (g_2), a signal of correlated heterozygosity among loci (Fig. S13-A). However, for 11 plant species, `inbreedR` failed to produce estimates.

iii) Finally, we incorporated qualitative data on plant mating systems extracted from an established plant trait database (*25*). Based on this source, we identified in our dataset 18 dioecious species, 19 self-incompatible (SI), 5 distylous, 10 self-compatible (SC), and 4 gynodioecious species. Information was not available for the remaining species.

Genomic data yielded similar results regarding the shape of the selfing rate distribution across species. Both methods produced a unimodal distribution with a peak at $s = 0$ (Fig. S13-A), whereas the literature suggests that, across a broader range of species, the distribution is typically bimodal, with a second peak at $s = 1$ (Fig. S13-B; (*97*)). This indicates that our sampling is likely biased toward a deficit of highly selfing species. A difference between `selfing_ML` and `inbreedR` is that more species were estimated to have $s = 0$ using `selfing_ML`. Qualitative analysis of mating systems reveals that these species are either dioecious or possess a genetic self-incompatibility (SI) system that prevents self-fertilization (Fig. S14), suggesting that this pattern has a biological basis rather than resulting from a methodological artifact. This qualitative analysis also supports the observed deficit of clearly selfing species: among the 10 species (over a total of 56 species with known mating system) identified as self-compatible (SC), the median selfing rate estimate was below 0.2 (Fig. S14). The gynodioecious taxa in our study correspond to *Silene nutans* lineages in which hermaphrodites are not self-incompatible, resulting in a reproductive system that combines obligatory outcrossing via females and partial selfing from hermaphroditic individuals.

To assess whether selfing influences the dynamics of reproductive isolation, we categorized species separately based on selfing rates estimated by the `selfing_ML` and `inbreedR` methods. For each method, we classified species into two categories based on whether their selfing rate was above (classified as 'highest') or below (classified as 'lowest') the method-specific median across the 105 plant species analyzed (94 species for which `inbreedR` provided estimates). The median selfing rate was 0 for `selfing_ML` and 0.026 for `inbreedR`. Based on these classifications, we grouped pairs of species into three categories: *highest–highest*, *lowest–lowest*, and *mixed* (one 'highest', one 'lowest' selfer). For `selfing_ML`, this resulted in 30 highest–highest pairs, 108 lowest–lowest, and 45 mixed pairs; for `inbreedR`, we obtained 45, 41, and 57 pairs, respectively (Fig. S4-A and B).

We then applied the same analytical framework used to compare reproductive isolation dynamics between plants and animals, testing whether the extent of gene flow differed significantly between the three pair categories. For both selfing rate estimation methods, and across all pairwise comparisons, we found mostly non-significant effect (Table S4). This lack of detectable signal may partly reflect the bias in our dataset toward outcrossing species, as previously noted.

Definitions and Notations concerning `selfing_ML`.

In this paragraph, we detail the rationale and functioning of the `selfing_ML` method.

- s is the equilibrium selfing rate, where $s \in [0, 1]$.
- p is the allele frequency of the alternative allele at a given site.
- $q = 1 - p$ is the frequency of the reference allele.
- F is the inbreeding coefficient, which is related to the selfing rate s at equilibrium by:

$$F = \frac{s}{2 - s}$$

Genotype Probabilities.

For a given site, the probabilities of observing the genotypes are modeled using p, q, and F:

- Probability of observing genotype "11" (homozygous for the alternative allele):

$$P_{11} = p^2 + p \cdot q \cdot F \tag{2}$$

- Probability of observing heterozygous genotype ”10” or ”01”:

$$P_{10} = 2 \cdot p \cdot q \cdot (1 - F) \tag{3}$$

- Probability of observing genotype ”00” (homozygous for the reference allele):

$$P_{00} = q^2 + p \cdot q \cdot F \tag{4}$$

Log-Likelihood Calculation.

For an individual i, the log-likelihood of the equilibrium selfing rate s explaining F is computed by summing the log-probabilities of the observed genotypes across all polymorphic sites:

$$\log \mathcal{L}(s \mid \text{data}) = \sum_{l=1}^{L} \sum_{j=1}^{n} \log P_{\text{geno}_{ij}}(f_A, s)$$

Where:

- L is the number of polymorphic loci (i.e, loci containing at least one polymorphic site).
- n is the number of polymorphic sites within each locus.
- $P_{\text{geno}_{ij}}$ is the probability of the observed genotype at site j for individual i, calculated based on equations (2-4).

Selfing Rate Estimation.

We evaluate the log-likelihood for a range of selfing rates s (from 0 to 1 in steps of 0.05). The estimated selfing rate $\hat{s}_i$ for individual i is the value of s that maximizes the log-likelihood:

$$\hat{s}_i = \arg \max_{s \in [0,1]} \log \mathcal{L}(s \mid \text{data})$$

Numerical Stability.

To avoid issues with numerical underflow when computing log-probabilities, a small constant $\epsilon = 10^{-10}$ is added:

$$\log(P_{11}) = \log(p^2 + p \cdot q \cdot F + \epsilon)$$

This ensures that the probabilities are always strictly positive, preventing errors in the log calculations.

Implementation.

The entire pipeline for estimating selfing rates using the maximum likelihood approach described above has been implemented in Python. This method, along with detailed documentation and example datasets, is freely available as an open-source project on `Zenodo` (*24*).

Accuracy of the estimate.

We evaluated the performance of our method across a range of 11 selfing rates (s) from 0 to 1. Genotype data were simulated using the R package `hierfstat` (*100*), which generates diploid genotypes for populations with a specified inbreeding coefficient (F), given a defined number of loci and individuals. The simulated datasets were then analyzed with our custom tool to estimate the selfing rate (s), allowing us to assess both the accuracy and the limitations of the method. We investigated the impact of varying sample sizes (2, 3, 4, 5, and 10 diploid individuals) and explored datasets consisting of 1,000 and 5,000 SNPs. Each combination of selfing rate, SNP count, and sample size was replicated 20 times.

Overall, selfing rate estimates (s) are substantially underestimated when the sample consists of only 2 diploid individuals (Fig. S12). Thus, for simulations with a sample size of 2 diploids and low true selfing rates (below 0.3), the inferred selfing rate is frequently estimated as zero. This bias diminishes rapidly with an increase in sample size. However, even with small sample sizes, our method reliably detects high selfing rates when the true value is elevated, although the estimates tend to be slightly underestimated. This underestimation in small samples can be attributed to the overestimation of true allele frequencies for rare alleles, which are the most common type of alleles in a site frequency spectrum. In a sample of only 2 diploid individuals, a rare allele would have a minimum frequency of 25%, which does not accurately reflect its true population frequency. This inflated allele frequency in small samples biases the selfing rate estimate downward by making the absence of homozygotes for such alleles unlikely under a non-zero selfing rate. This effect diminishes both with a moderate increase in sample size (beyond 2 individuals) and at higher true selfing rates.

A.8 Data availability

All assembled datasets, the reference list used for mapping, the results of demographic inference, and the R scripts for statistical analyses and figure generation are available on Zenodo (*24, 41, 42*).

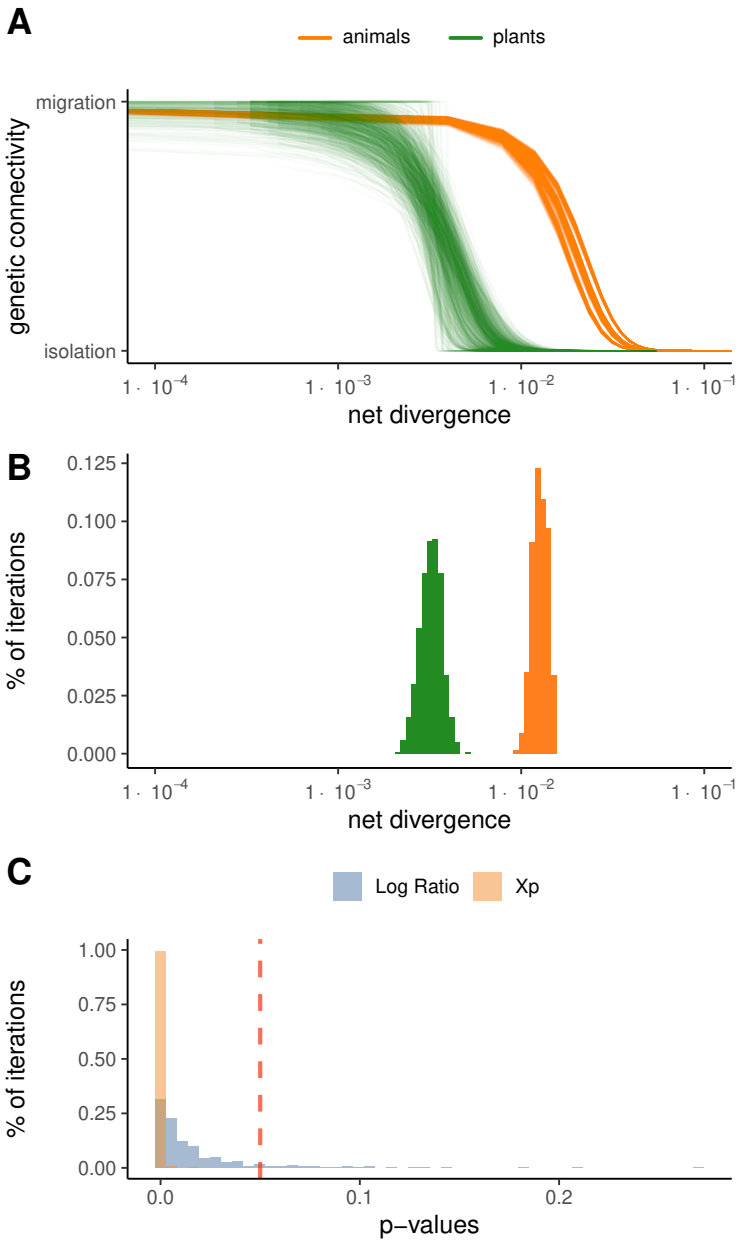

Figure S1: **Test of the robustness of differences between plants and animals to genus effects through random subsampling.** (**A**) Relationship between net divergence and migration/isolation status based on random sub-sampling of one population/species pair per genus (plants: green; animals: orange), repeated 1,000 times. Each line represents one sub-sampling. (**B**) Distribution of inflection points ($X_{p=0.5}$) for plants and animals across 1,000 sub-samplings. (**C**) *P*-value distributions from 1,000 sub-samplings: blue bars represent the log-likelihood ratio tests, and orange bars represent the tests on difference between inflection points ($X_{p=0.5;\text{animals}} - X_{p=0.5;\text{plants}}$). The red dashed line indicates the 0.05 significance threshold.

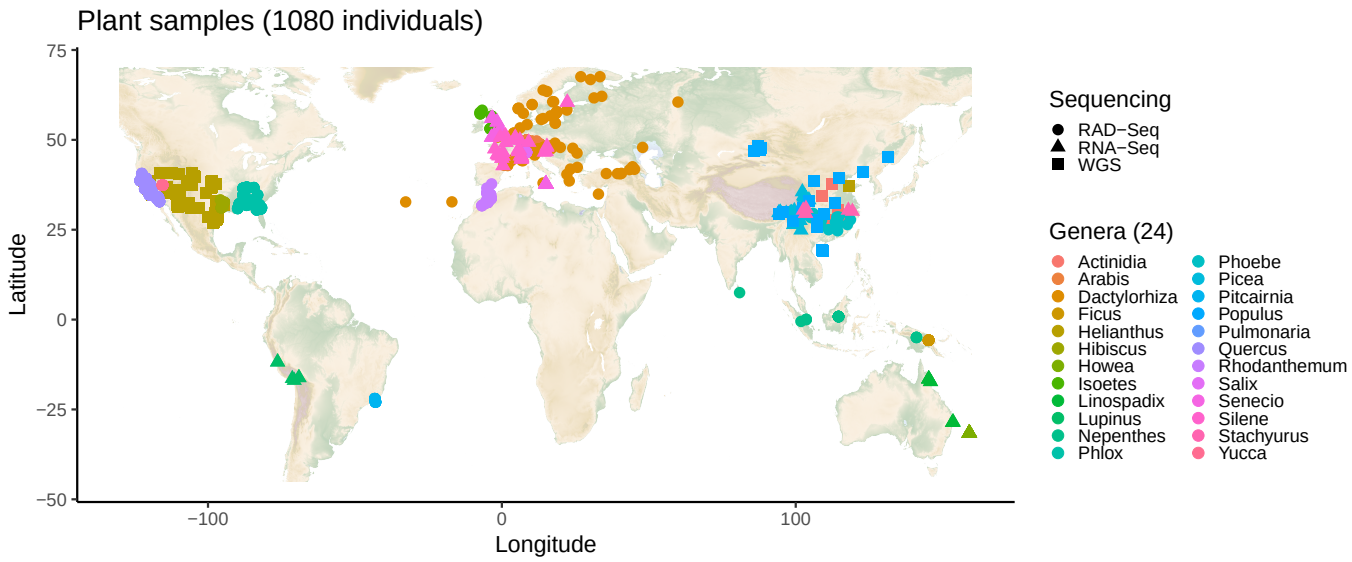

Figure S2: **Geographical location of plant samples and sequencing methods.** Each symbol represents a sequenced plant individual. The shapes correspond to the sequencing technologies used, while the colors indicate the genera.

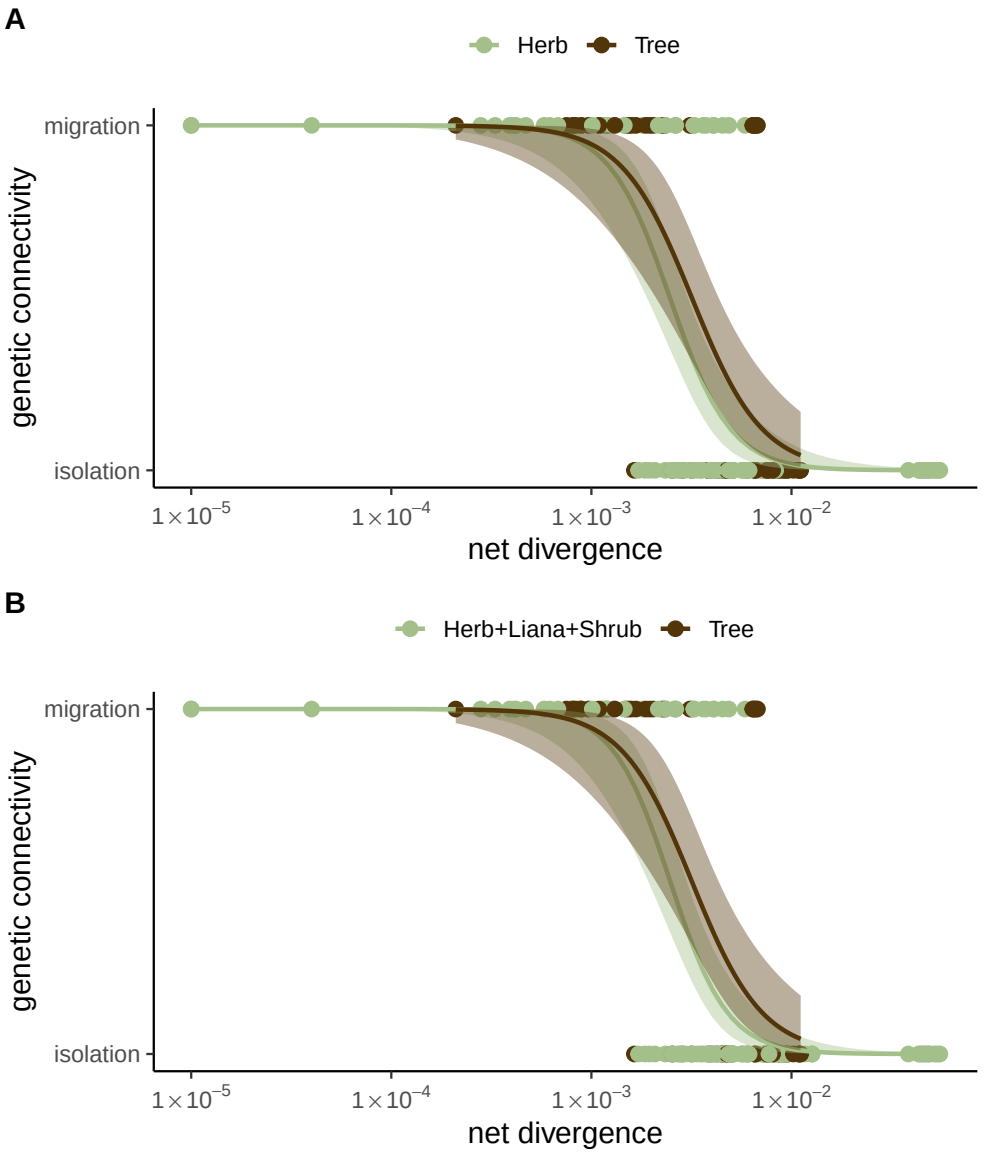

Figure S3: **Relationship between mean net divergence and migration/isolation status across plant life forms.** (**A**) Comparison between herbs and trees. (**B**) Comparison between herbs, lianas, and shrubs combined *versus* trees.

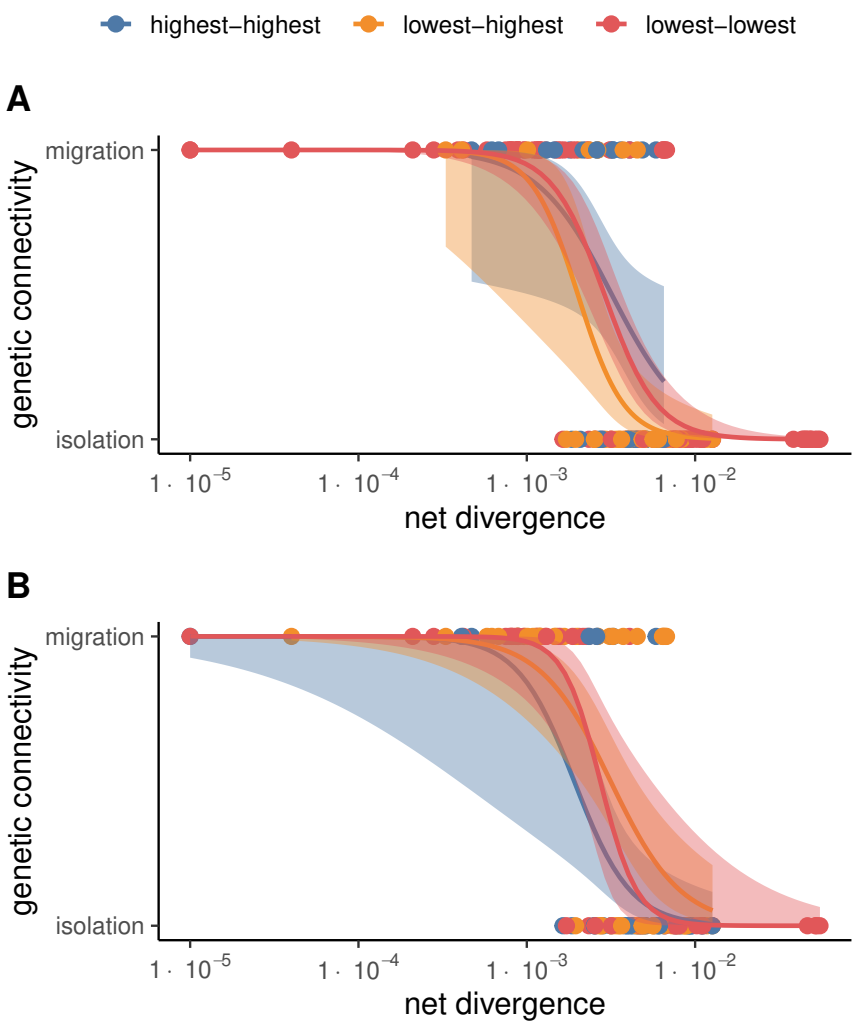

Figure S4: **Relationship between net divergence and genetic connectivity across plant pairs grouped by selfing rates or reproductive systems.** (**A**) Grouping based on species-averaged selfing rates estimated with `selfing_ML` (median across species = 0). Pairs are classified as *highest-highest* (both species with $s > 0$, $n = 30$), *lowest-lowest* (both with $s = 0$, $n = 108$), or mixed (one species with $s > 0$, the other with $s = 0$, $n = 45$). (**B**) Same classification using estimates from `inbreedR` (median = 0.026). Pairs are grouped as *highest-highest* ($s > 0.026$, $n = 45$), *lowest-lowest* ($s \leq 0.026$, $n = 41$), or mixed ($n = 57$). Only pairs with a significant signal of migration or isolation are included as for Figure 1.

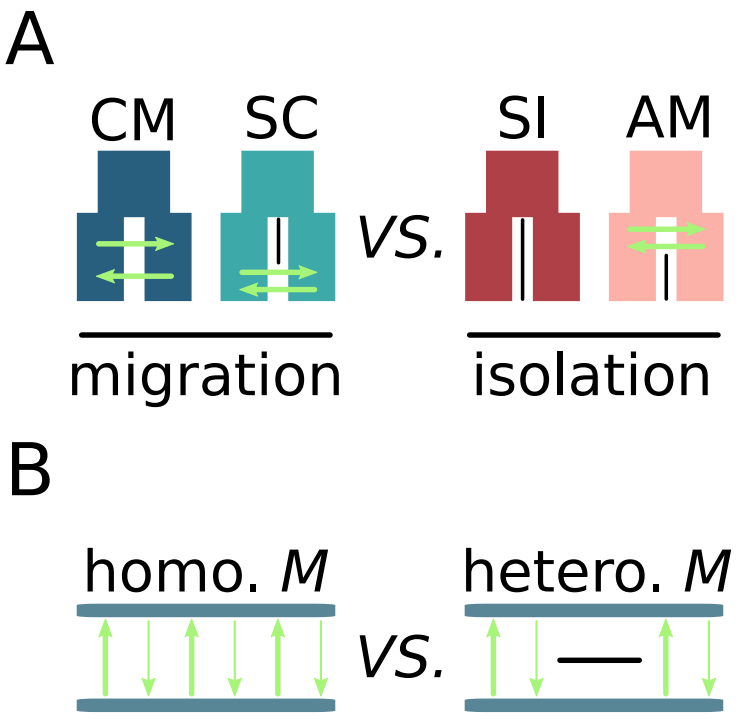

Figure S5: **Compared models using approximate Bayesian computation (ABC).** (**A**) Models with ongoing migration correspond to all CM (Continuous Migration) and SC (Secondary Contact) models. Models with current isolation correspond to all SI (Strict Isolation) and AM (Ancestral Migration) models. The first step in our ABC classification is to compare the set of CM+SC *versus* SI+AM models in order to assign a migration or isolation status to each of the 341 pairs of lineages (61 animals, 280 plants) according to the computed posterior probability.

(**B**) Pairs of plants or animals, for which our ABC framework has provided strong statistical evidence of ongoing migration, are subsequently subjected to analysis aimed at discerning the uniformity of gene flow across the genome, whether it exhibits homogeneity (characterized by the absence of local genomic barriers) or heterogeneity (signifying genetic linkage to species barriers). The comparison between homo. *M versus* hetero. *M* was carried out using the same ABC framework as in the previous step.

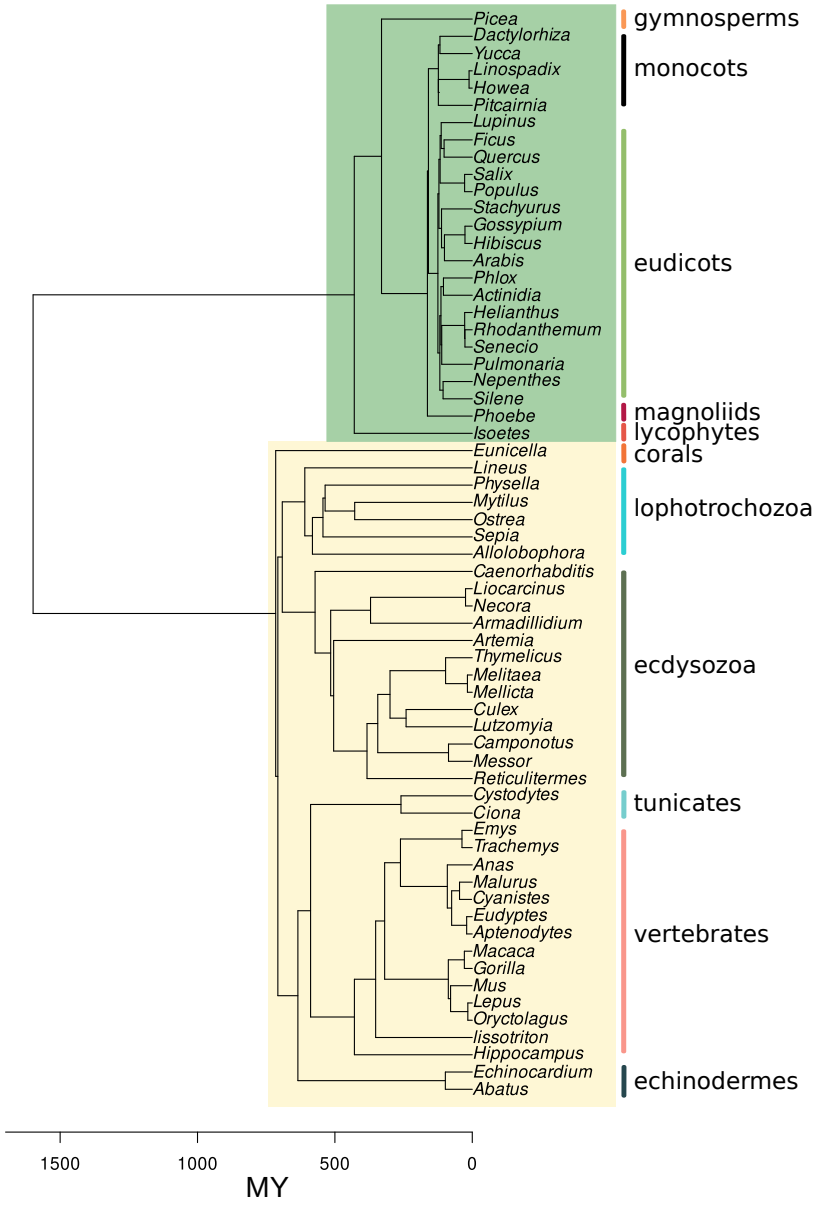

Figure S6: **Phylogenetic relationships between species included in the current study.** Plants and animals are indicated by green and yellow rectangles respectively. The scale represent the time from present expressed in million years (MY) according to `TimeTree` (*101*). Animals (yellow square) are from (*14*). Plants (green square) are included in the current study.

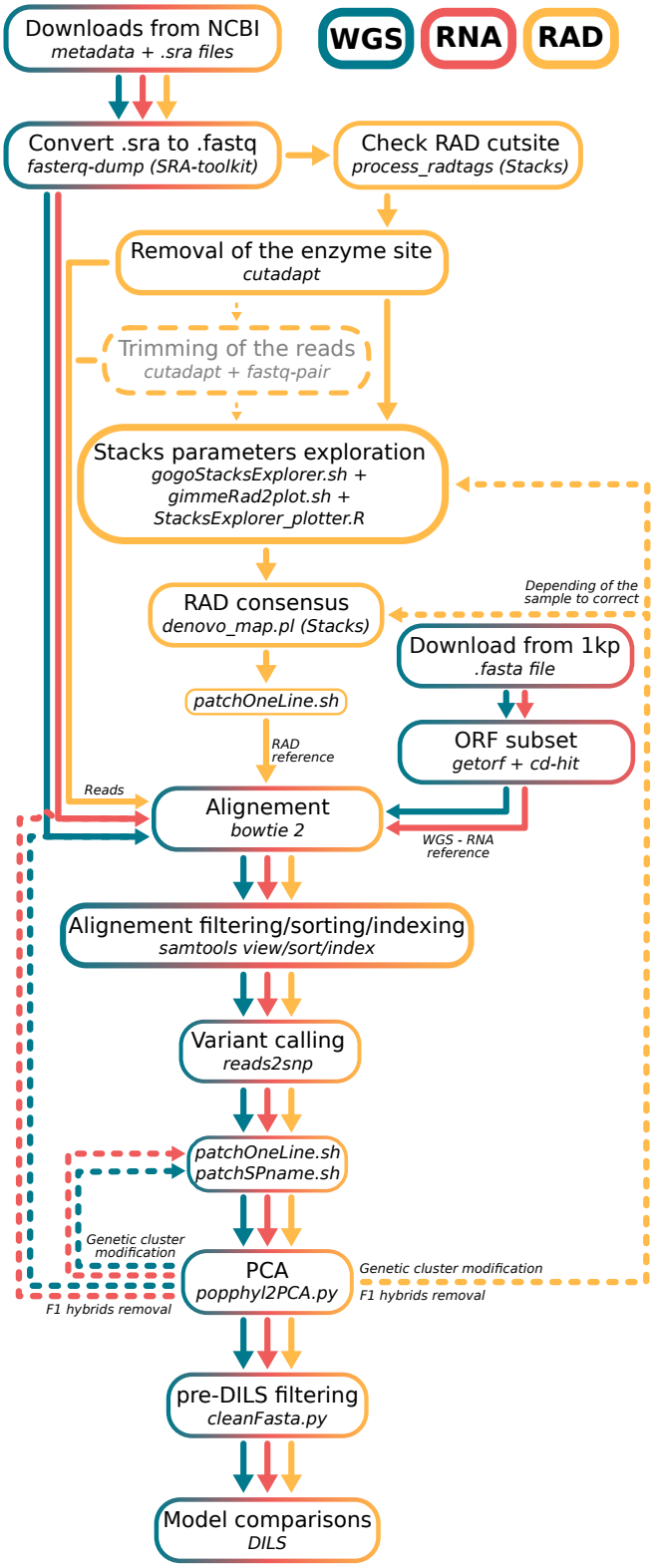

Figure S7: **Bioinformatics steps from the raw reads to demographic analysis.** Within each box, the upper line delineates an information technology procedure employed for data processing, while the lower line specifies the program or script utilized for its execution. The coloration denotes the specific sequencing technology concerned by each step.

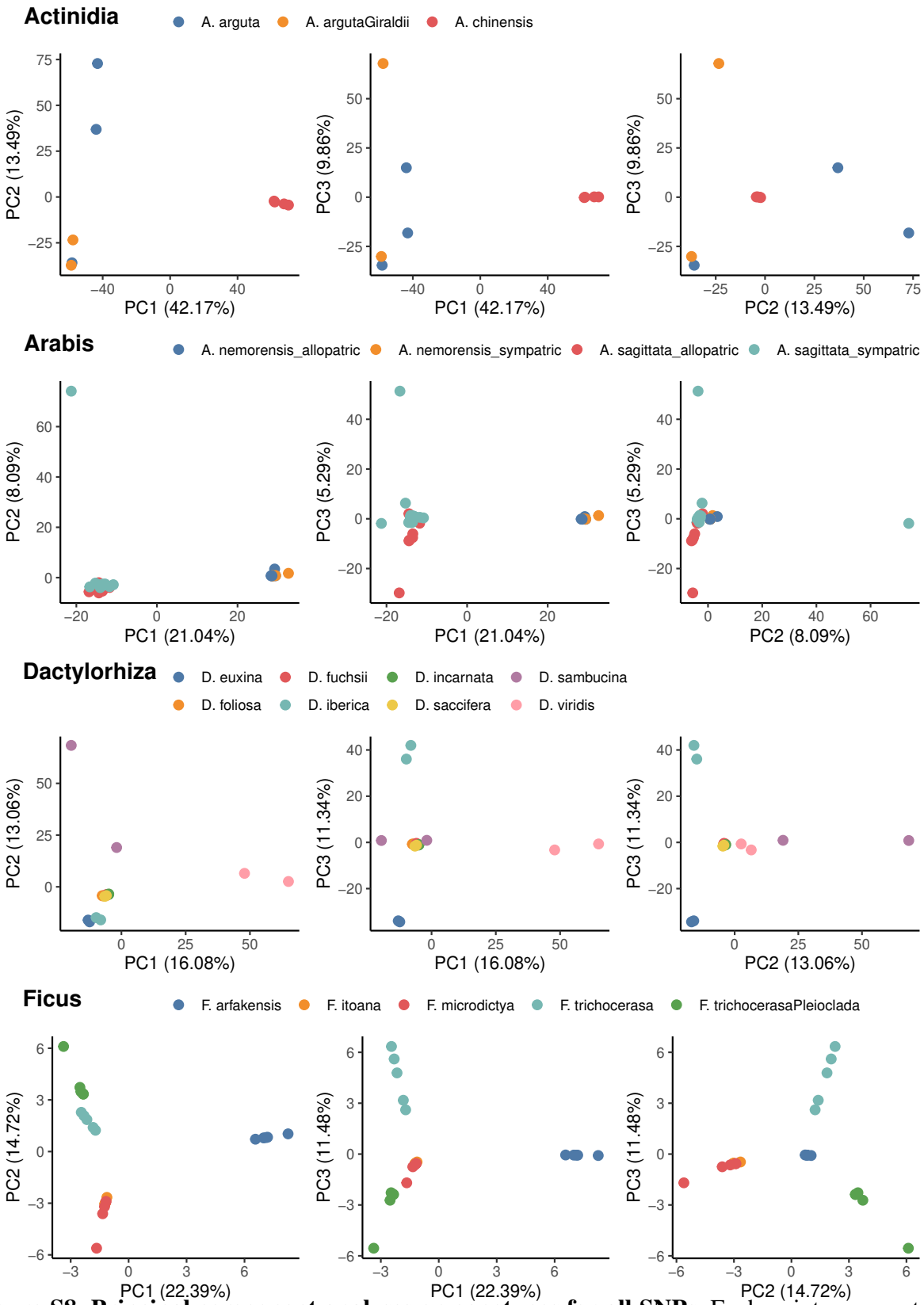

Figure S8: **Principal component analyses on genotypes for all SNPs.** Each point represents an individual. The colours represent the different populations/species named by the authors of the studies from which the data originated. (Page 1 of 6)

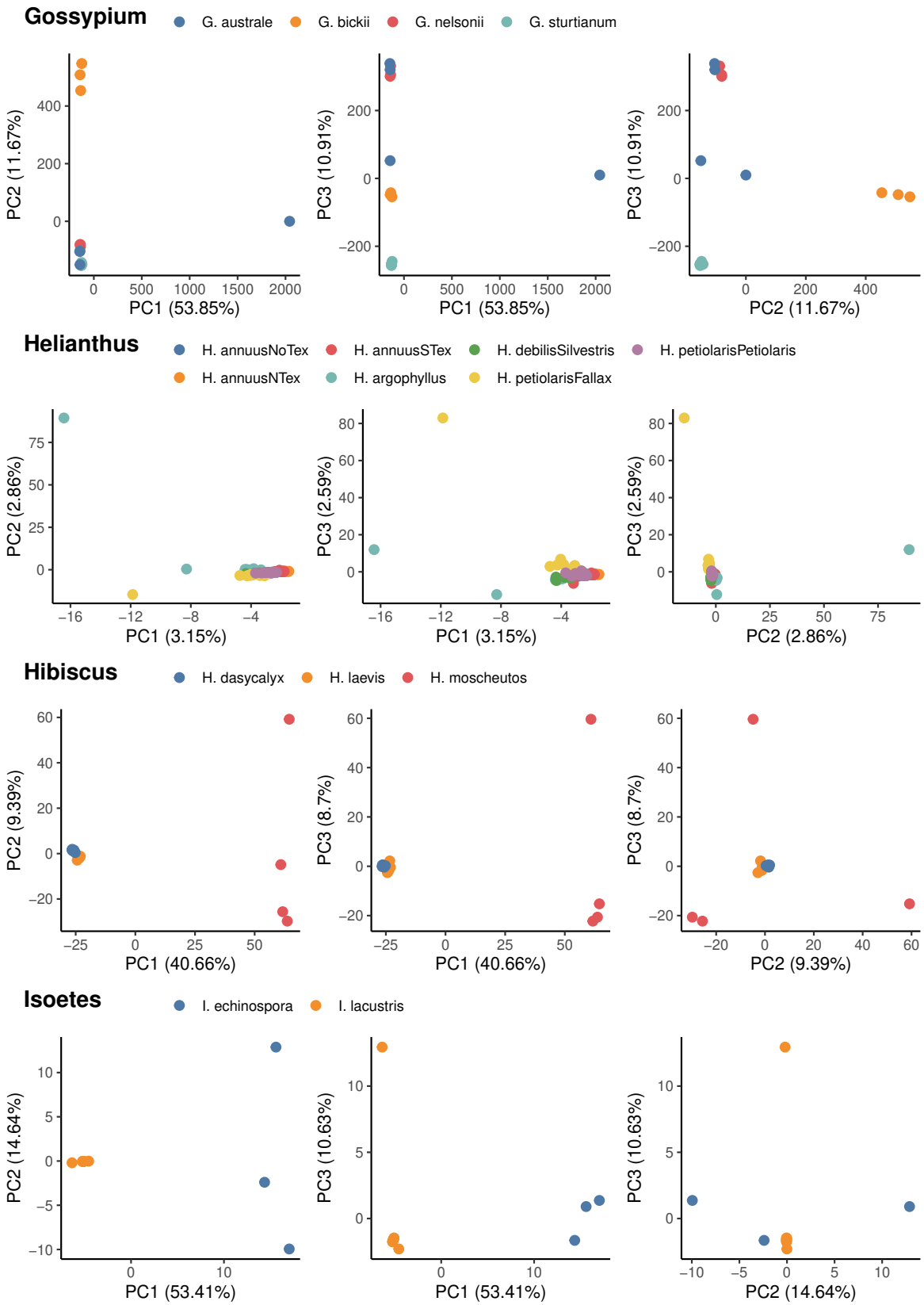

Figure S8 (continued, page 2 of 6)

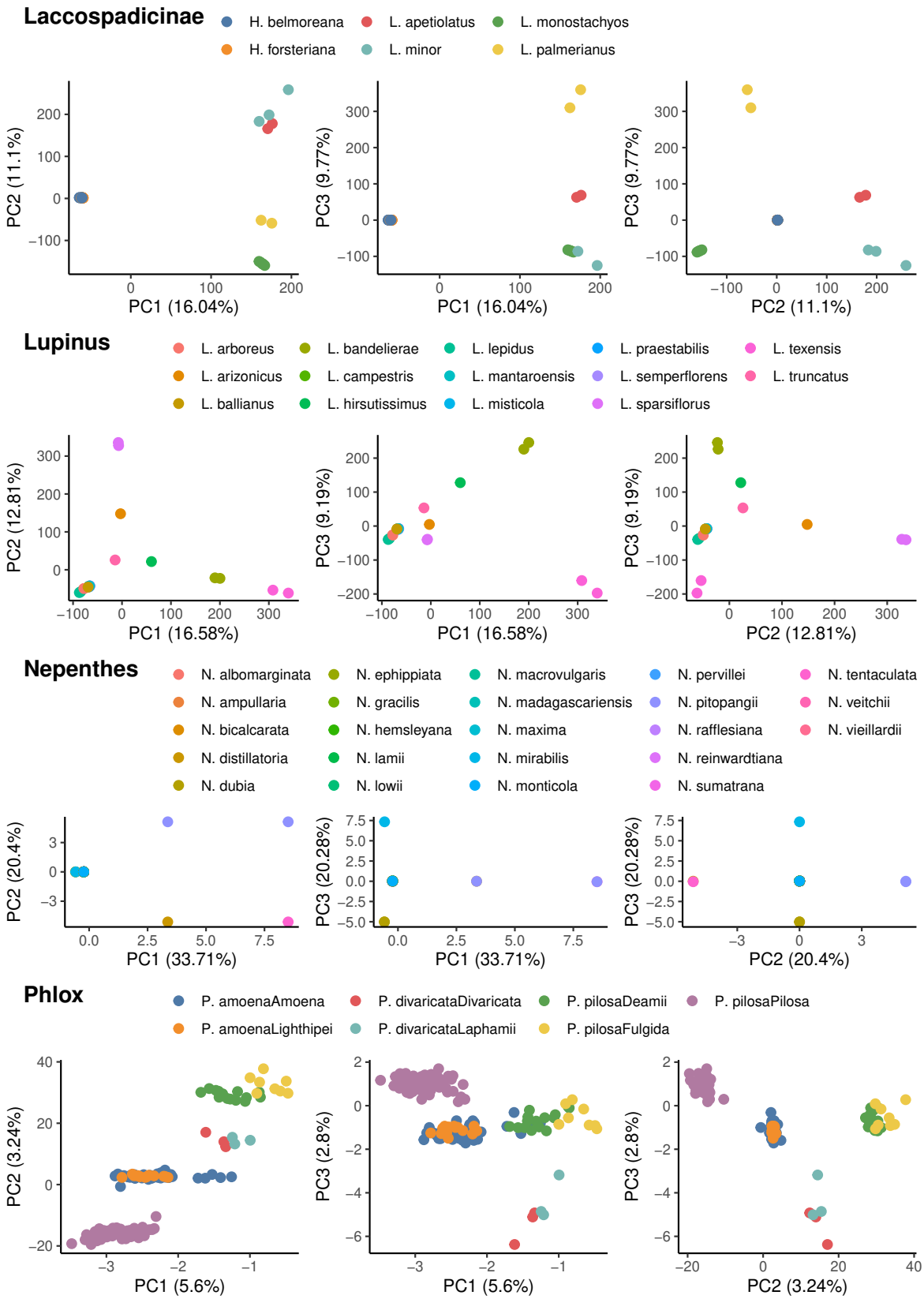

Figure S8 (continued, page 3 of 6)

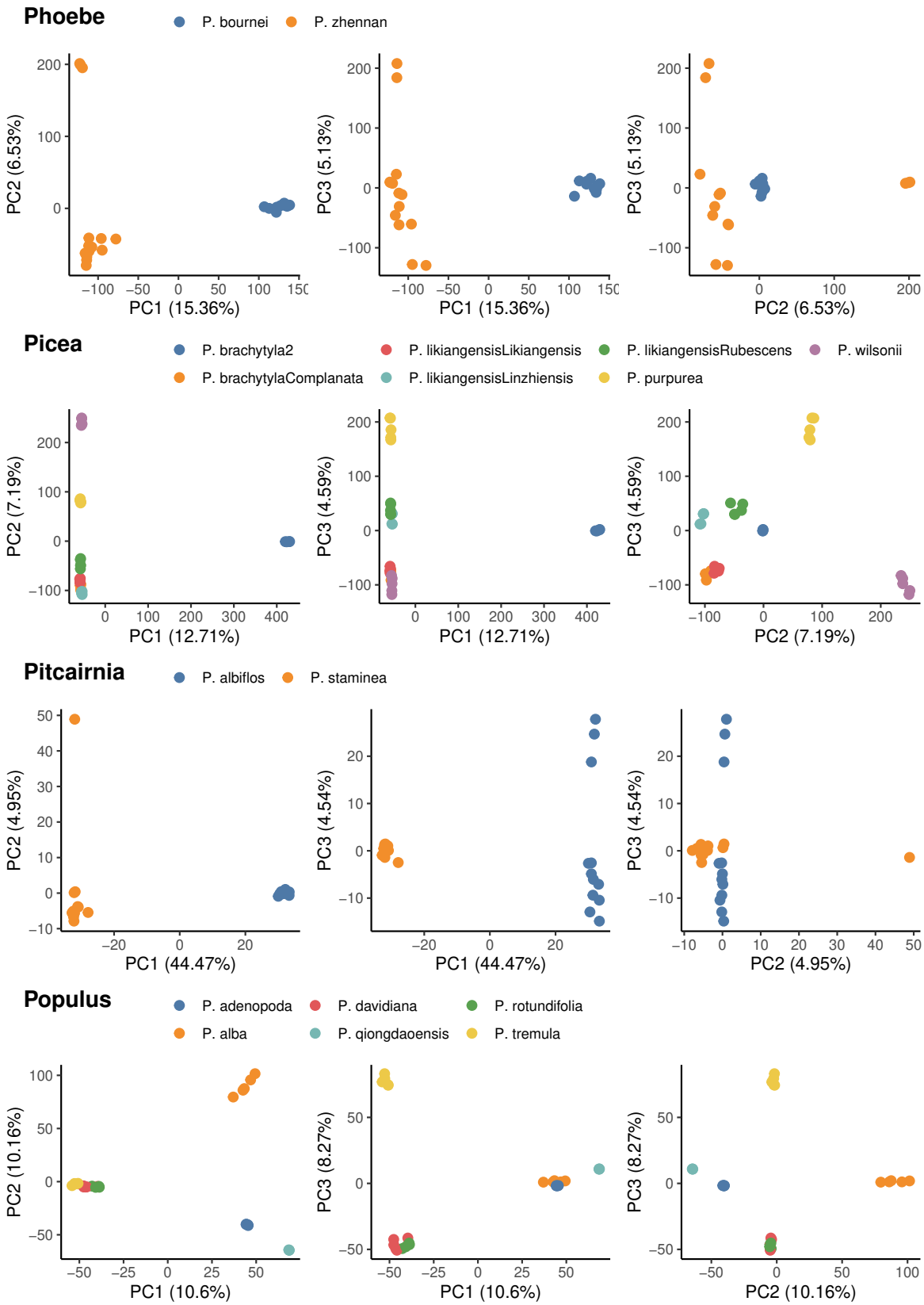

Figure S8 (continued, page 4 of 6)

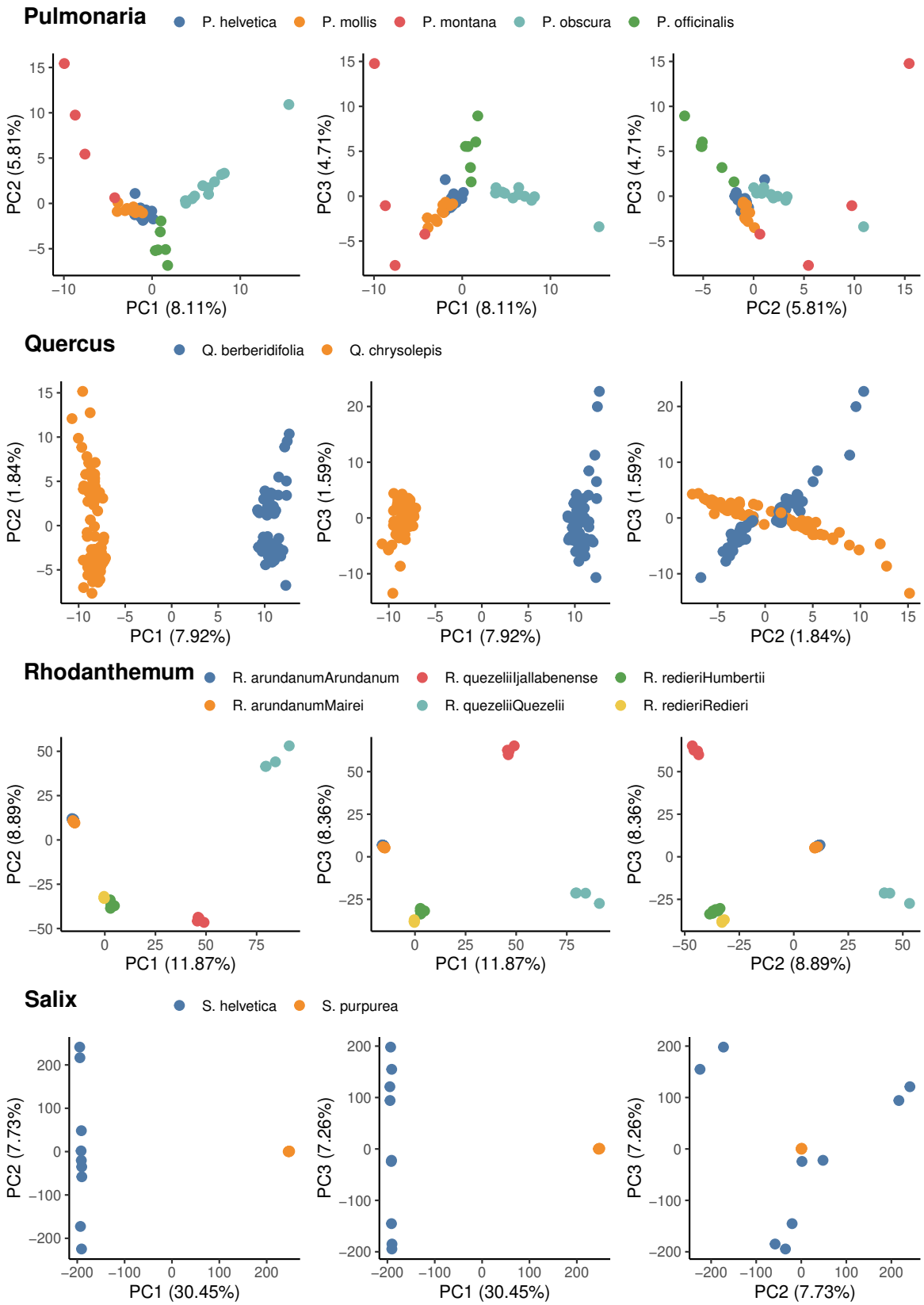

Figure S8 (continued, page 5 of 6)

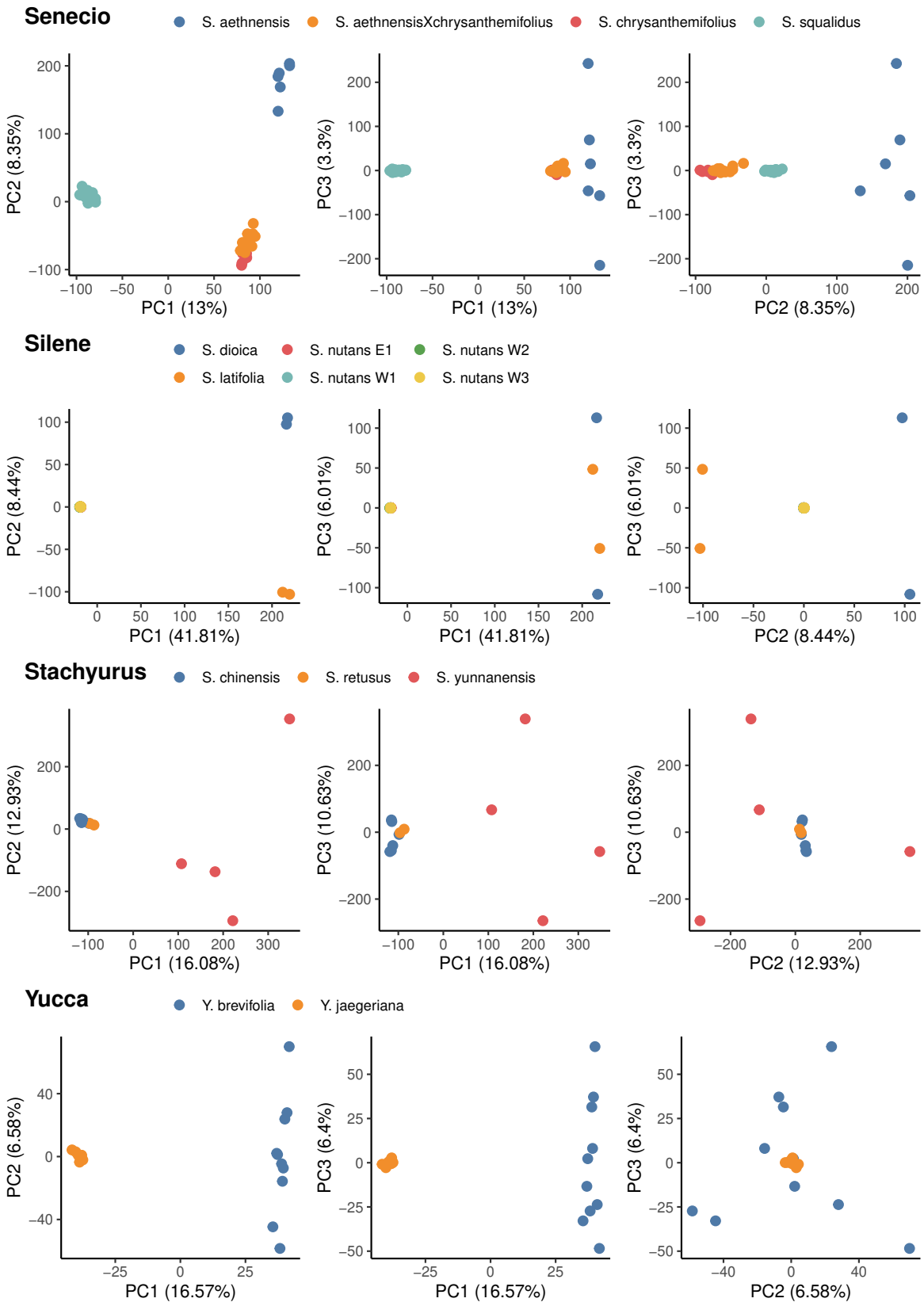

Figure S8 (continued, page 6 of 6)

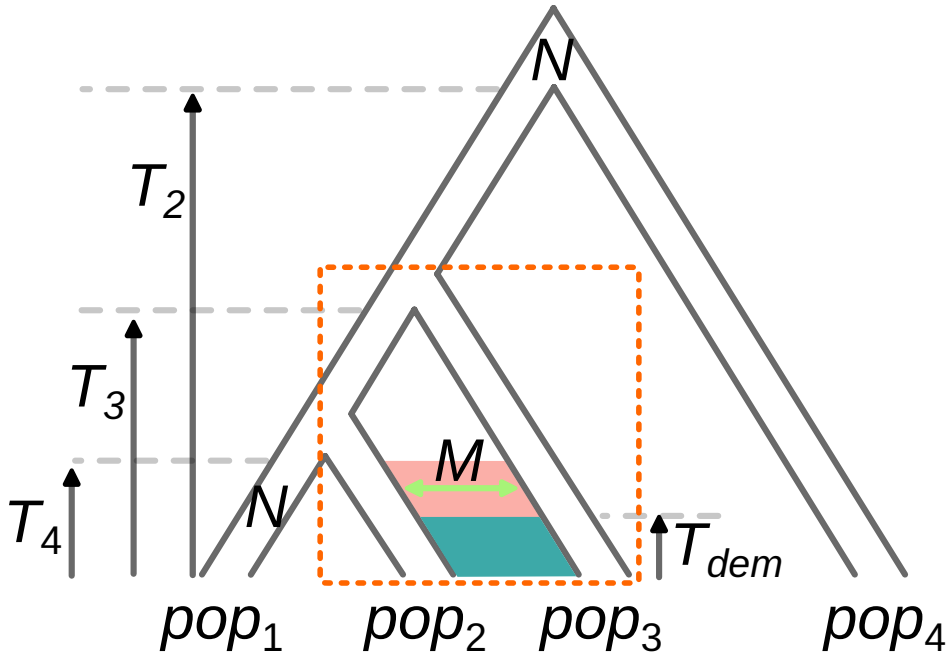

Figure S9: **Simulated demographic model for testing introgression between populations pop_2 and pop_3 (dotted rectangle) using DILS and QuIBL.** Parameters: T_4 (speciation time between 1 and 2), T_3 (speciation time between $(1, 2)$ and 3), T_2 (speciation time between $((1, 2), 3)$ and 4), T_{dem} (migration onset/end), M (number of migrants per generation), N_e (effective population size). Scenarios: $\text{SI}_{4\text{pop}}$ ($M = 0$ during red and blue periods), $\text{AM}_{4\text{pop}}$ ($M > 0$ during red, $M = 0$ during blue), $\text{IM}_{4\text{pop}}$ ($M > 0$ during red and blue), $\text{SC}_{4\text{pop}}$ ($M > 0$ during blue, $M = 0$ during red).

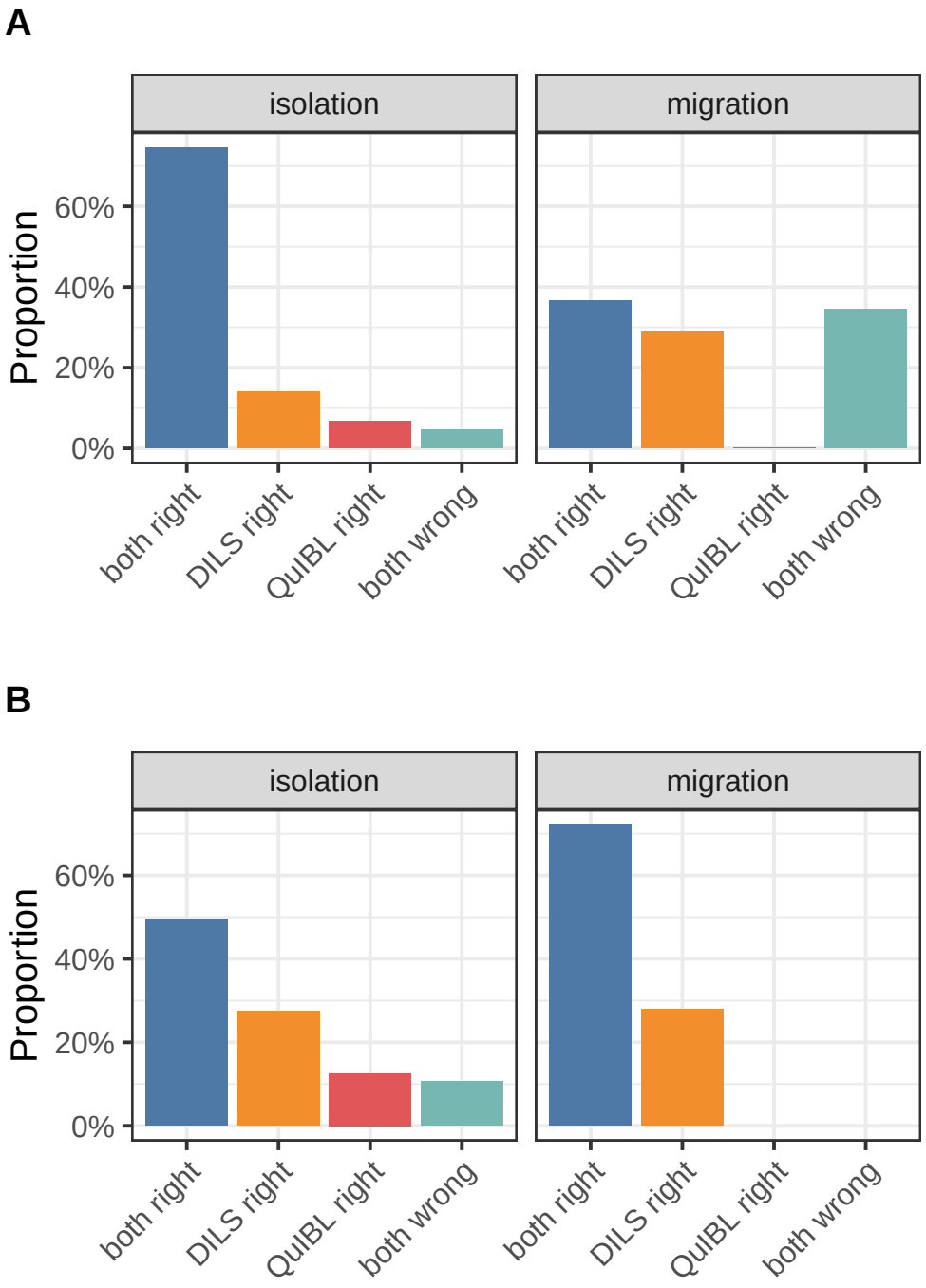

Figure S10: **Comparison of DILS and QuIBL results on simulated datasets with migration (IM_{4pop}, SC_{4pop}; Fig. S9) and without migration (SI_{4pop}, AM_{4pop}; Fig. S9))**. Colors represent the proportions of simulations where: both methods were correct (blue), only DILS was correct (orange), only QuIBL was correct (red), and both methods were incorrect (green). (**A**) Results across all explored parameters. (**B**) Results for parameters where more than 10% of loci are affected by migration and $N_e.m > 0.25$.

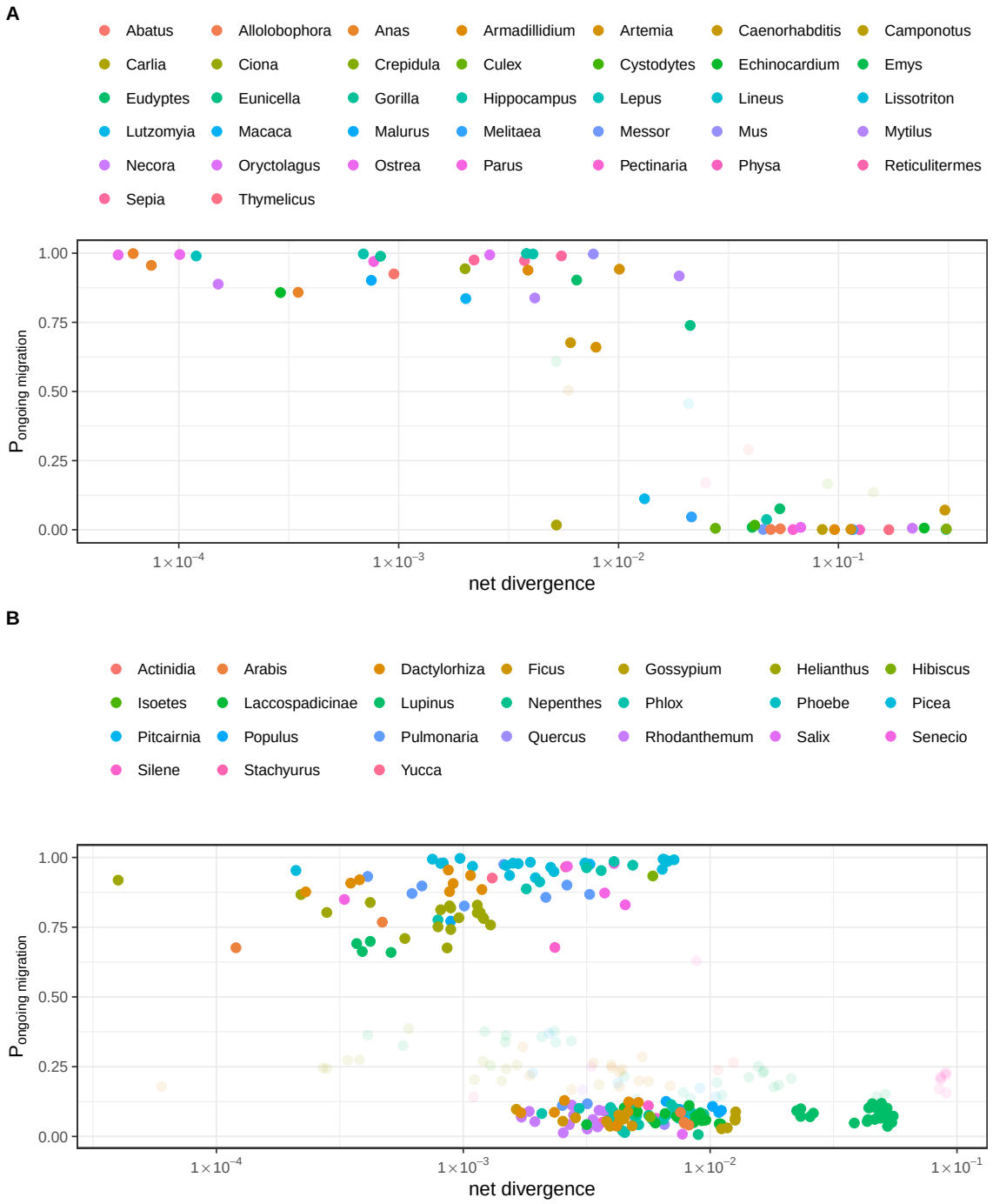

Figure S11: **Relationship between mean net divergence and posterior probability for ongoing migration.** Each point corresponds to a pair of (**A**) animals or (**B**) plants. x-axis: average net divergence. y-axis: posterior probability for ongoing migration attributed by our ABC framework. Colours correspond to surveyed genera. Solid points represent pairs for which there is strong statistical evidence either supporting or rejecting the ongoing migration model, as determined by the robustness test outlined in (*14*). In contrast, transparent points indicate pairs for which the comparison between the migration and isolation models yields an inconclusive result. Pairs for which support was inconclusive were excluded from further analysis. The remaining pairs were categorized either as exhibiting 'migration' or 'isolation', as illustrated in Figure 1-A (see section A.4.9).

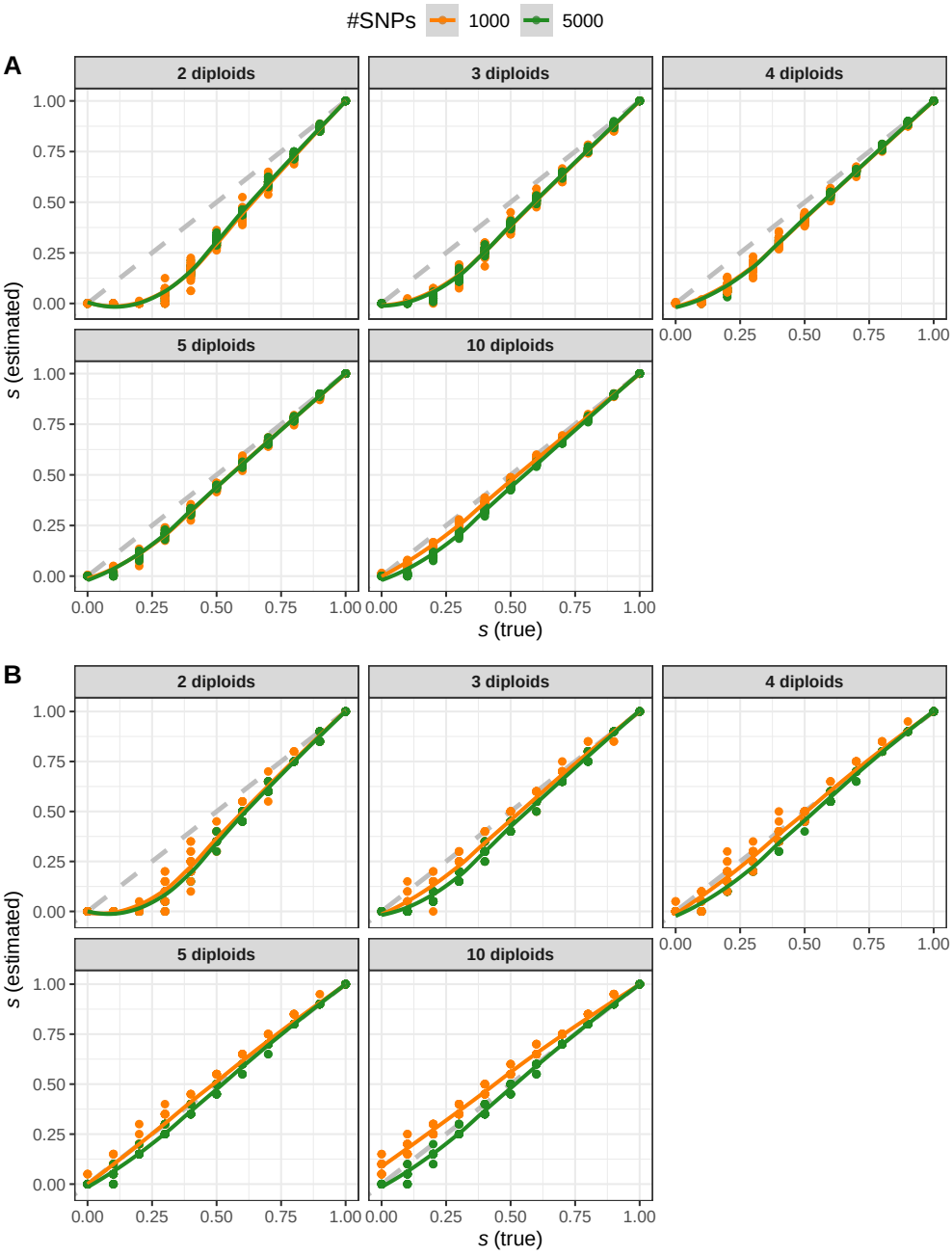

Figure S12: **Evaluation of selfing rate (*s*) estimation from simulated genomic data.**The x-axis represents the true selfing rate (*s*) used for simulations, while the y-axis shows the selfing rate estimated using our custom method (*24*). (**A**) Mean selfing rate estimated across all individuals. (**B**) Maximum selfing rate estimated among individuals. Orange and green points represent simulated datasets with 1,000 and 5,000 SNPs, respectively. The number of sampled diploid individuals is indicated in each sub-panel.

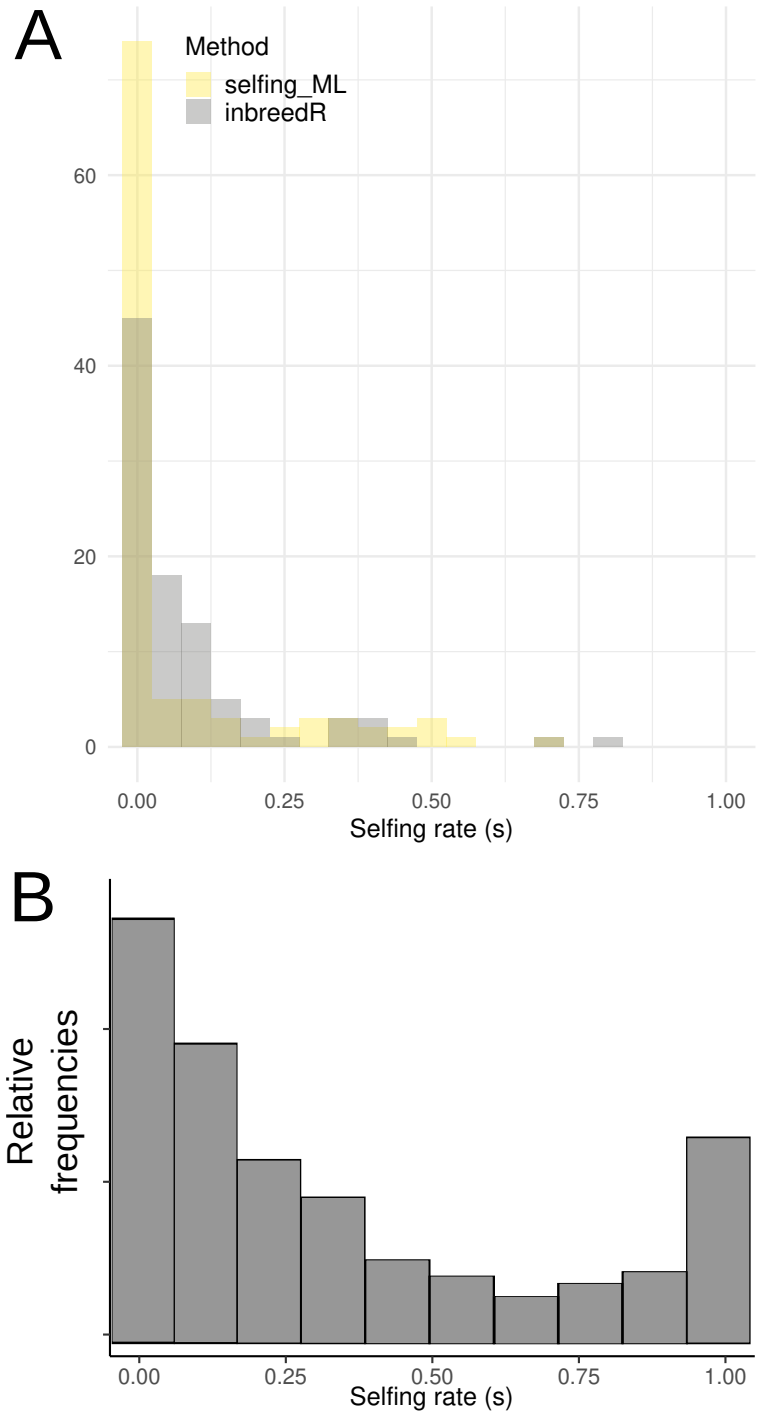

Figure S13: **Distribution of selfing rates (*s*) estimated from molecular data.** (**A**) Distribution of species-averaged selfing rates among plant species used in the plant-*versus*-animal comparison, estimated using `selfing_ML` (yellow) and `inbreedR` (*23*) (grey). (**B**) Meta-analysis of selfing rates across 329 plant species not related to our study, modified from the study by Igic and Kohn in 2006 (*97*).

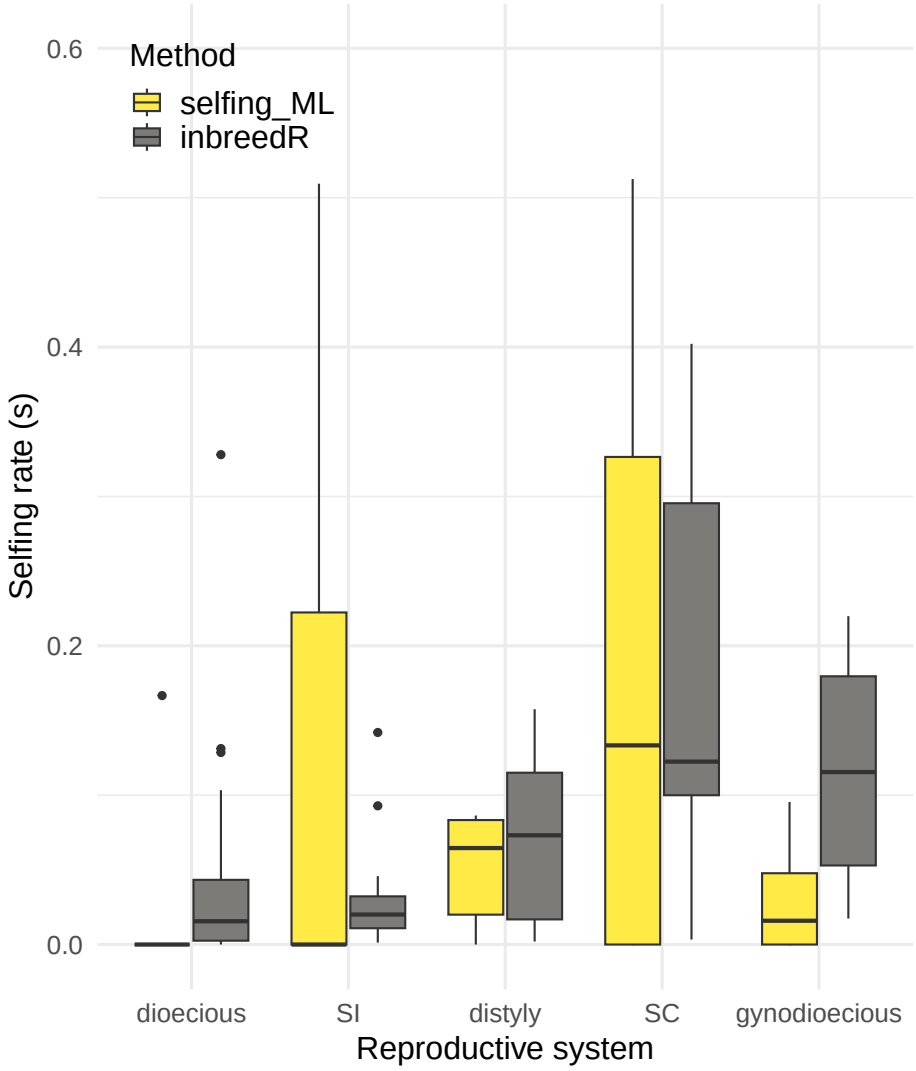

Figure S14: **Distribution of estimated selfing rates (*s*) across reproductive systems in 56 plant species.**

The reproductive systems include: 18 dioecious, 19 self-incompatible (SI), 5 distylous, 10 self-compatible (SC), and 4 gynodioecious species. Selfing rate estimates were obtained using two methods: **yellow**, the custom-developed `selfing_ML` method; and **grey**, the `inbreedR` package (*23*). Each boxplot summarizes the distribution of species-averaged selfing rates per reproductive system (*25*) and method.

Table S1: List of retained NCBI datasets.

bioproject	genus	species	n	type of data	source
PRJNA318567	*Actinidia*	*arguta*	3	WGS	(*43*)
		arguta giraldii	2		
		chinensis	4		
PRJEB33482,	*Arabis*	*nemorensis allop.*	6	RNA	(*44*)
PRJEB39992		*nemorensis symp.*	6		
		sagittata allop.	10		
		sagittata symp.	15		
PRJNA489792	*Dactylorhiza*	*euxina*	5	RAD	(*45*)
		foliosa	2		
		fuchsii	30		
		iberica	2		
		incarnata	31		
		saccifera	4		
		sambucina	3		
		viridis	3		
PRJNA445222	*Ficus*	*arfakensis*	14	RAD	(*46*)
		itoana	13		
		microdictya	15		
		trichocerasa	15		
		t. pleioclada	26		
PRJNA539957	*Gossypium*	*australe*	4	WGS	(*47*)
		bickii	3		
		nelsonii	3		
		robinsonii	2		
		sturtianum	6		
PRJNA532579	*Helianthus*	*annuus NoTex*	15	WGS	(*48*)
		annuus NTex	15		

		annuus STex	15		
		argophyllus	10		
		debilis silvestris	5		
		niveus canescens	8		
		petiolaris fallax	10		
		p. petiolaris	10		
PRJNA382435	*Hibiscus*	*dasycalyx*	6	RAD	(*49*)
		laevis	4		
		moscheutos	5		
PRJNA483403	*Isoetes*	*lacustris*	9	RAD	(*50*)
		echiospora	3		
PRJNA244607	*Howea*	*belmoreana*	40	RNA	(*51*)
		forsteriana	39		
PRJNA528594	*Linospadix*	*monostachyos*	18		(*52*)
		minor	9		
		apetiolatus	6		
		palmerianus	6		
PRJNA318864	*Lupinus*	*ballianus*	2	RNA	(*102*)
		bandelierae	2		
		misticola	2		
PRJEB37794	*Nepenthes*	*albomarginata*	3	RAD	(*53*)
		ampullaria	8		
		bicalcarata	6		
		distillatoria	2		
		dubia	2		
		ephippiata	2		
		gracilis	8		
		hemsleyana	4		
		lamii	2		

		lowii	2		
		macrovulgaris	2		
		madagascariensis	2		
		maxima	10		
		mirabilis	10		
		monticola	2		
		pervillei	16		
		pitopangii	2		
		rafflesiana	9		
		reinwardtiana	2		
		sumatrana	2		
		tentaculata	2		
		veitchii	3		
		vieillardii	2		
PRJNA701424	*Phlox*	*amoena amoena*	48	RAD	(*54*)
		a. lighthipei	14		
		divaricata divaricata	3		
		d. laphamii	3		
		pilosa deamii	15		
		p. fulgida	8		
		p. pilosa	59		
		subulata	2		
PRJNA464259	*Phoebe*	*zhennan*	9	RAD	(*55*)
		bournei	12		
PRJNA807675	*Pitcairnia*	*albiflos*	9	RAD	(*56*)
		staminea	12		
PRJNA392950,	*Picea*	*brachytyla*	4	RNA	(*57, 58*)
PRJNA401149,		*b. complanata*	5		
PRJNA378930,		*likiangensis likiangensis*	5		

PRJNA301093		*l. linzhiensis*	5		
		l. rubescens	5		
		purpurea	5		
		wilsoni	5		
PRJNA612655	*Populus*	*adenopoda*	5	WGS	(*59*)
		alba	5		
		davidiana	5		
		qiongdaoensis	3		
		rotundifolia	4		
		tremula	5		
PRJNA544114	*Pulmonaria*	*helvetica*	24	RAD	(*60*)
		mollis	10		
		montana	4		
		obscur	11		
		officinalis	6		
PRJNA639507	*Quercus*	*berberidifolia*	63	RAD	(*61*)
		chrysolepis	80		
PRJNA554975	*Rhodanthemum*	*redieri redieri*	4	RAD	(*62*)
		r. humbertii	7		
		quezelii quezelii	4		
		q. jallabenense	4		
		arundanum mairei	8		
		a. arundanum	27		
PRJNA429746	*Salix*	*helvetica*	10	RAD	(*63*)
		purpurea	10		
PRJNA549571	*Senecio*	*aethnensis*	6	RNA	(*64*)
		aethn. X chrys.	14		
		chrysanthemifolius	6		
		squalidus	28		

PRJNA295359	*Silene*	*dioica*	2	RNA	(*65, 66*)
		latifolia	2		
		nutans E1	4		
		n. W1	4		
		n. W2	4		
		n. W3	4		
PRJNA553020	*Stachyurus*	*chinensis*	6	RNA	(*67*)
		retusus	2		
		yunnanensis	4		
PRJNA329381	*Yucca*	*brevifolia*	24	RAD	(*68*)
		jaegeriana	39		

Table S2: Log-likelihood Ratio Test for logit models fitted to plant and animal datasets (Fig. 1)

Model	$\boldsymbol{\ell}$	$\boldsymbol{\beta_0}$	$\boldsymbol{\beta_1}$	$\boldsymbol{X_{p=0.5}}$	***P*-value**
M_0	−91.55841	−11.28733	−4.50495	0.00312	
M_{plants}	−61.60692	−16.15385	−6.27316	0.00266	
M_{animals}	−8.01012	−11.13402	−6.09245	0.01488	
					$< 1 \times 10^{-4}$

$\boldsymbol{\ell}$: log-likelihoods of models M_0, M_{plants}, and M_{animals}.

$\boldsymbol{\beta_0}$: estimated intercept ($\log_{10}$ scaled).

$\boldsymbol{\beta_1}$: estimated coefficient ($\log_{10}$ scaled).

$\boldsymbol{X_{p=0.5}}$: inflection point beyond which, for any level of divergence, less than 50% of pairs are expected to be connected by gene flow ($X_{p=0.5} = 10^{-\beta_0/\beta_1}$).

***P*-value**: probability that by random chance, the absolute difference between $\ell(M_{\text{plants}}) + \ell(M_{\text{animals}})$ and $\ell(M_0)$ exceeds the observed value, estimated from 10,000 random permutations.

Table S3: Log-likelihood Ratio Test for logit models fitted to plant and animal datasets obtained by RNA-sequencing only

Model	$\boldsymbol{\ell}$	$\boldsymbol{\beta_0}$	$\boldsymbol{\beta_1}$	$\boldsymbol{X_{p=0.5}}$	***P*-value**
M_0	−34.28871	−11.79074	−5.31189	0.00603	
M_{plants}	−18.28419	−19.81859	−8.47952	0.00460	
M_{animals}	−4.29462	−16.75334	−9.58339	0.01786	
					$< 1 \times 10^{-4}$

$\boldsymbol{\ell}$: log-likelihoods of models M_0, M_{plants}, and M_{animals}.

$\boldsymbol{\beta_0}$: estimated intercept ($\log_{10}$ scaled).

$\boldsymbol{\beta_1}$: estimated coefficient ($\log_{10}$ scaled).

$\boldsymbol{X_{p=0.5}}$: inflection point beyond which, for any level of divergence, less than 50% of pairs are expected to be connected by gene flow ($X_{p=0.5} = 10^{-\beta_0/\beta_1}$).

***P*-value**: probability that by random chance, the absolute difference between $\ell(M_{\text{plants}}) + \ell(M_{\text{animals}})$ and $\ell(M_0)$ exceeds the observed value, estimated from 10,000 random permutations.

Table S4: Log-likelihood Ratio Test for logit models assessing factor effects on reproductive isolation dynamics within plants

Factor	**Comparison**	**Group 1**	**Group 2**	$\ell(M_0)$	$\ell(M_1)$	$\ell(M_2)$	***P*-value**
Life form	Herbs *versus* Trees	Herb	Tree	−59.78	−36.095	−22.433	0.261
	H–L–S *versus* Trees	H–L–S	Tree	−61.607	−37.662	−22.433	0.1937
Selfing rate	s_h–s_h *vs.* s_l–s_l	s_h–s_h	s_l–s_l	−46.2	−17.2	−28.7	0.63
(`selfing_ML`)	s_h–s_h *vs.* s_h–s_l	s_h–s_h	s_h–s_l	−32.53	−17.17	−12.47	0.045
	s_l–s_l *vs.* s_h–s_l	s_l–s_l	s_h–s_l	−43.12	−28.66	−12.47	0.1515
Selfing rate	s_h–s_h *vs.* s_l–s_l	s_h–s_h	s_l–s_l	−22.21	−12.79	−8.38	0.4
(`inbreedR`)	s_h–s_h *vs.* s_h–s_l	s_h–s_h	s_h–s_l	−41.58	−12.79	−26.23	0.07
	s_l–s_l *vs.* s_h–s_l	s_l–s_l	s_h–s_l	−35.69	−8.38	−26.23	0.316

H–L–S: group of species including herbs, lianas, and shrubs.

s_h and s_l: species with selfing rates above and below the median, respectively.

Pairs are grouped as s_h–s_h (high–high), s_l–s_l (low–low), or s_h–s_l (mixed).

Selfing rates were estimated using two approaches: the custom method `selfing_ML` (*24*) and the `inbreedR` package (*23*).

$\ell(M_0)$: log-likelihood of the model fitted to all species pairs; $\ell(M_1)$ and $\ell(M_2)$: log-likelihoods of models fitted separately to each group.

P-value: probability that, by random chance, the absolute difference between $\ell(M_1) + \ell(M_2)$ and $\ell(M_0)$ exceeds the observed value, estimated from 10,000 permutations.

References and Notes

1. R. Abbott, D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, A. Brelsford, C. A. Buerkle, R. Buggs, R. K. Butlin, U. Dieckmann, F. Eroukhmanoff, A. Grill, S. H. Cahan, J. S. Hermansen, G. Hewitt, A. G. Hudson, C. Jiggins, J. Jones, B. Keller, T. Marczewski, J. Mallet, P. Martinez-Rodriguez, M. Möst, S. Mullen, R. Nichols, A. W. Nolte, C. Parisod, K. Pfennig, A. M. Rice, M. G. Ritchie, B. Seifert, C. M. Smadja, R. Stelkens, J. M. Szymura, R. Väinölä, J. B. W. Wolf, D. Zinner, Hybridization and speciation. *J. Evol. Biol.* **26**, 229–246 (2013). doi:10.1111/j.1420-9101.2012.02599.x Medline

2. G. L. Stebbins, The role of hybridization in evolution. *Proc. Am. Philos. Soc.* **103**, 231–251 (1959).

3. M. Slatkin, "The rate of spread of an advantageous allele in a subdivided population" in *Population genetics and ecology* (Elsevier, 1976), pp. 767–780.

4. M. R. Brown, P. M. Hollingsworth, L. L. Forrest, M. L. Hart, I. J. Leitch, L. Jones, C. Ford, N. de Vere, A. D. Twyford, Genetic factors predict hybrid formation in the British flora. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2220261120 (2023). doi:10.1073/pnas.2220261120 Medline

5. A. M. Westram, S. Stankowski, P. Surendranadh, N. Barton, What is reproductive isolation? *J. Evol. Biol.* **35**, 1143–1164 (2022). doi:10.1111/jeb.14005 Medline

6. C.-I. Wu, The genic view of the process of speciation. *J. Evol. Biol.* **14**, 851–865 (2001). doi:10.1046/j.1420-9101.2001.00335.x

7. M. E. Frayer, B. A. Payseur, Do genetic loci that cause reproductive isolation in the lab inhibit gene flow in nature? *Evolution* **78**, 1025–1038 (2024). doi:10.1093/evolut/qpae044 Medline

8. S. Wright, The genetical structure of populations. *Ann. Eugen.* **15**, 323–354 (1951). doi:10.1111/j.1469-1809.1949.tb02451.x Medline

9. S. H. Martin, J. W. Davey, C. D. Jiggins, Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32**, 244–257 (2015). doi:10.1093/molbev/msu269 Medline

10. A. J. Dagilis, D. Peede, J. M. Coughlan, G. I. Jofre, E. R. R. D'Agostino, H. Mavengere, A. D. Tate, D. R. Matute, A need for standardized reporting of introgression: Insights from studies across eukaryotes. *Evol. Lett.* **6**, 344–357 (2022). doi:10.1002/evl3.294 Medline

11. C. Fraïsse, I. Popovic, C. Mazoyer, B. Spataro, S. Delmotte, J. Romiguier, É. Loire, A. Simon, N. Galtier, L. Duret, N. Bierne, X. Vekemans, C. Roux, DILS: Demographic inferences with linked selection by using ABC. *Mol. Ecol. Resour.* **21**, 2629–2644 (2021). doi:10.1111/1755-0998.13323 Medline

12. D. R. Laetsch, G. Bisschop, S. H. Martin, S. Aeschbacher, D. Setter, K. Lohse, Demographically explicit scans for barriers to gene flow using gIMble. *PLOS Genet.* **19**, e1010999 (2023). doi:10.1371/journal.pgen.1010999 Medline

13. V. C. Sousa, M. Carneiro, N. Ferrand, J. Hey, Identifying loci under selection against gene flow in isolation-with-migration models. *Genetics* **194**, 211–233 (2013). doi:10.1534/genetics.113.149211 Medline

14. C. Roux, C. Fraïsse, J. Romiguier, Y. Anciaux, N. Galtier, N. Bierne, Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLOS Biol.* **14**, e2000234 (2016). doi:10.1371/journal.pbio.2000234 Medline

15. C. Fraïsse, A. Le Moan, C. Roux, G. Dubois, C. Daguin-Thiebaut, P.-A. Gagnaire, F. Viard, N. Bierne, Introgression between highly divergent sea squirt genomes: An adaptive breakthrough? *Peer Community J.* **2**, e54 (2022). doi:10.24072/pcjournal.172

16. J. Mallet, Hybridization as an invasion of the genome. *Trends Ecol. Evol.* **20**, 229–237 (2005). doi:10.1016/j.tree.2005.02.010 Medline

17. L. Gottlieb, Genetics and morphological evolution in plants. *Am. Nat.* **123**, 681–709 (1984). doi:10.1086/284231

18. E. Mayr, *Animal species and evolution* (Harvard Univ. Press, 1963).

19. L. H. Rieseberg, T. E. Wood, E. J. Baack, The nature of plant species. *Nature* **440**, 524–527 (2006). doi:10.1038/nature04402 Medline

20. S. Wang, J. E. Mank, D. Ortiz-Barrientos, L. H. Rieseberg, Genome architecture and speciation in plants and animals. *Mol. Ecol.* 10.1111/mec.70004 (2025). doi:10.1111/mec.70004 Medline

21. N. Galtier, Delineating species in the speciation continuum: A proposal. *Evol. Appl.* **12**, 657–663 (2019). doi:10.1111/eva.12748 Medline

22. N. B. Edelman, P. B. Frandsen, M. Miyagi, B. Clavijo, J. Davey, R. B. Dikow, G. García-Accinelli, S. M. Van Belleghem, N. Patterson, D. E. Neafsey, R. Challis, S. Kumar, G. R. P. Moreira, C. Salazar, M. Chouteau, B. A. Counterman, R. Papa, M. Blaxter, R. D. Reed, K. K. Dasmahapatra, M. Kronforst, M. Joron, C. D. Jiggins, W. O. McMillan, F. Di Palma, A. J. Blumberg, J. Wakeley, D. Jaffe, J. Mallet, Genomic architecture and introgression shape a butterfly radiation. *Science* **366**, 594–599 (2019). doi:10.1126/science.aaw2090 Medline

23. M. A. Stoffel, M. Esser, M. Kardos, E. Humble, H. Nichols, P. David, J. I. Hoffman, inbreedR: An R package for the analysis of inbreeding based on genetic markers. *Methods Ecol. Evol.* **7**, 1331–1339 (2016). doi:10.1111/2041-210X.12588

24. C. Roux, popgenomics/selfing_ML: selfing_ML, Zenodo (2025); https://doi.org/10.5281/zenodo.15296403.

25. A. J. Helmstetter, M. Méndez, J. Schönenberger, C. Burgarella, B. Anderson, M. von Balthazar, S. Billiard, H. de Boer, J. Cros, P.-A. Delecroix, M. Dufay, J. R. Pannell, D. S. Bianchi, D. J. Schoen, M. Vallejo-Marin, R. Zenil-Ferguson, H. Sauquet, S. Glémin, J. Käfer, Pollination and mating traits underlie diverse reproductive strategies in flowering plants. bioRxiv 2024.02.26.582019 [Preprint] (2024); https://doi.org/10.1101/2024.02.26.582019.

26. C. Roux, G. Tsagkogeorga, N. Bierne, N. Galtier, Crossing the species barrier: Genomic hotspots of introgression between two highly divergent Ciona intestinalis species. *Mol. Biol. Evol.* **30**, 1574–1587 (2013). doi:10.1093/molbev/mst066 Medline

27. L. C. Moyle, E. B. Graham, Genetics of hybrid incompatibility between Lycopersicon esculentum and L. hirsutum. *Genetics* **169**, 355–373 (2005). doi:10.1534/genetics.104.029546 Medline

28. L. C. Moyle, T. Nakazato, Comparative genetics of hybrid incompatibility: Sterility in two Solanum species crosses. *Genetics* **179**, 1437–1453 (2008). doi:10.1534/genetics.107.083618 Medline

29. A. L. Sweigart, L. Fishman, J. H. Willis, A simple genetic incompatibility causes hybrid male sterility in mimulus. *Genetics* **172**, 2465–2479 (2006). doi:10.1534/genetics.105.053686 Medline

30. R. J. Abbott, Plant speciation across environmental gradients and the occurrence and nature of hybrid zones. *J. Syst. Evol.* **55**, 238–258 (2017). doi:10.1111/jse.12267

31. R. G. Harrison, Hybrid zones: Windows on evolutionary process. *Oxf. Surv. Evol. Biol.* **7**, 69–128 (1990).

32. C. L. Morjan, L. H. Rieseberg, How species evolve collectively: Implications of gene flow and selection for the spread of advantageous alleles. *Mol. Ecol.* **13**, 1341–1356 (2004). doi:10.1111/j.1365-294X.2004.02164.x Medline

33. R. Frankham, C. J. Bradshaw, B. W. Brook, Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biol. Conserv.* **170**, 56–63 (2014). doi:10.1016/j.biocon.2013.12.036

34. M. P. Zuellig, A. L. Sweigart, A two-locus hybrid incompatibility is widespread, polymorphic, and active in natural populations of Mimulus. *Evolution* **72**, 2394–2405 (2018). doi:10.1111/evo.13596 Medline

35. L. H. Rieseberg, B. K. Blackman, Speciation genes in plants. *Ann. Bot. (Lond.)* **106**, 439–455 (2010). doi:10.1093/aob/mcq126 Medline

36. M. E. Frayer, N. V. Robles, M. J. R. Barrera, J. M. Coughlan, M. Schumer, The molecular evolutionary basis of species formation revisited. EcoEvoRxiv [Preprint] (2025); https://doi.org/10.32942/X2CP9B.

37. J. P. Scholl, J. J. Wiens, Diversification rates and species richness across the Tree of Life. *Proc. Biol. Sci.* **283**, 20161334 (2016).

38. H. Morlon, J. Andréoletti, J. Barido-Sottani, S. Lambert, B. Perez-Lamarque, I. Quintero, V. Senderov, P. Veron, Phylogenetic Insights into Diversification. *Annu. Rev. Ecol. Evol. Syst.* **55**, 1–21 (2024). doi:10.1146/annurev-ecolsys-102722-020508

39. D. I. Bolnick, A. K. Hund, P. Nosil, F. Peng, M. Ravinet, S. Stankowski, S. Subramanian, J. B. W. Wolf, R. Yukilevich, A multivariate view of the speciation continuum. *Evolution* **77**, 318–328 (2023). doi:10.1093/evolut/qpac004 Medline

40. S. Stankowski, A. D. Cutter, I. Satokangas, B. A. Lerch, J. Rolland, C. M. Smadja, J. C. Segami Marzal, C. R. Cooney, P. G. D. Feulner, F. M. C. B. Domingos, H. L. North, R.

Yamaguchi, R. K. Butlin, J. B. W. Wolf, J. Coughlan, P. Heidbreder, R. Hernández-Gutiérrez, K. B. Barnard-Kubow, D. Peede, L. Rancilhac, R. B. Salvador, K. A. Thompson, E. A. Stacy, L. C. Moyle, M. D. Garlovsky, A. Maulana, A. Kantelinen, N. I. Cacho, H. Schneemann, M. Domínguez, E. B. Dopman, K. Lohse, S. J. Rometsch, A. A. Comeault, R. M. Merrill, E. S. C. Scordato, S. Singhal, V. Pärssinen, A. C. R. Lackey, S. Kumar, J. I. Meier, N. Barton, C. Fraïsse, M. Ravinet, J. Kulmuni, Toward the integration of speciation research. *Evol. J. Linn. Soc.* **3**, kzae001 (2024). doi:10.1093/evolinnean/kzae001

41. F. Monnet, P. Zoe, T. Pascal, F. Christelle, V. d. P. Yves, V. Xavier, R. Camille, Rapid establishment of species barriers in plants compared to animals, Zenodo (2025); https://zenodo.org/records/15288584.

42. F. Monnet, Ladarwall/Greenworld: Main release, Zenodo (2025); https://doi.org/10.5281/zenodo.15881247.

43. Y. Liu, D. Li, Q. Zhang, C. Song, C. Zhong, X. Zhang, Y. Wang, X. Yao, Z. Wang, S. Zeng, Y. Wang, Y. Guo, S. Wang, X. Li, L. Li, C. Liu, H. C. McCann, W. He, Y. Niu, M. Chen, L. Du, J. Gong, P. M. Datson, E. Hilario, H. Huang, Rapid radiations of both kiwifruit hybrid lineages and their parents shed light on a two-layer mode of species diversification. *New Phytol.* **215**, 877–890 (2017). doi:10.1111/nph.14607 Medline

44. H. Dittberner, A. Tellier, J. de Meaux, Approximate Bayesian computation untangles signatures of contemporary and historical hybridization between two endangered species. *Mol. Biol. Evol.* **39**, msac015 (2022). doi:10.1093/molbev/msac015 Medline

45. M. K. Brandrud, J. Baar, M. T. Lorenzo, A. Athanasiadis, R. M. Bateman, M. W. Chase, M. Hedrén, O. Paun, Phylogenomic relationships of diploids and the origins of allotetraploids in Dactylorhiza (Orchidaceae). *Syst. Biol.* **69**, 91–109 (2020). doi:10.1093/sysbio/syz035 Medline

46. D. Souto-Vilaros, M. Proffit, B. Buatois, M. Rindos, M. Sisol, T. Kuyaiva, B. Isua, J. Michalek, C. T. Darwell, M. Hossaert-McKey, G. D. Weiblen, V. Novotny, S. T. Segar, Pollination along an elevational gradient mediated both by floral scent and pollinator compatibility in the fig and fig-wasp mutualism. *J. Ecol.* **106**, 2256–2273 (2018). doi:10.1111/1365-2745.12995

47. C. E. Grover, M. A. Arick II, A. Thrash, J. Sharbrough, G. Hu, D. Yuan, S. Snodgrass, E. R. Miller, T. Ramaraj, D. G. Peterson, J. A. Udall, J. F. Wendel, Dual domestication, diversity, and differential introgression in Old World cotton diploids. *Genome Biol. Evol.* **14**, evac170 (2022). doi:10.1093/gbe/evac170 Medline

48. G. L. Owens, M. Todesco, N. Bercovich, J.-S. Légaré, N. Mitchell, K. D. Whitney, L. H. Rieseberg, Standing variation rather than recent adaptive introgression probably underlies differentiation of the texanus subspecies of Helianthus annuus. *Mol. Ecol.* **30**, 6229–6245 (2021). doi:10.1111/mec.16008 Medline

49. J. Norrell, Differentiating the Neches River Rose Mallow (hibiscus Dasycalyx) from Its Congeners by Means of Phylogenetics and Population Genetics, thesis, University of Texas at Tyler (2017).

50. D. P. Wood, J. K. Olofsson, S. W. McKenzie, L. T. Dunning, Contrasting phylogeographic structures between freshwater lycopods and angiosperms in the British Isles. *Bot. Lett.* **165**, 476–486 (2018). doi:10.1080/23818107.2018.1505545

51. L. T. Dunning, H. Hipperson, W. J. Baker, R. K. Butlin, C. Devaux, I. Hutton, J. Igea, A. S. T. Papadopulos, X. Quan, C. M. Smadja, C. G. N. Turnbull, V. Savolainen, Ecological speciation in sympatric palms: 1. Gene expression, selection and pleiotropy. *J. Evol. Biol.* **29**, 1472–1487 (2016). doi:10.1111/jeb.12895 Medline

52. O. G. Osborne, A. Ciezarek, T. Wilson, D. Crayn, I. Hutton, W. J. Baker, C. G. N. Turnbull, V. Savolainen, Speciation in Howea palms occurred in sympatry, was preceded by ancestral admixture, and was associated with edaphic and phenological adaptation. *Mol. Biol. Evol.* **36**, 2682–2697 (2019). doi:10.1093/molbev/msz166 Medline

53. M. Scharmann, A. Wistuba, A. Widmer, Introgression is widespread in the radiation of carnivorous Nepenthes pitcher plants. *Mol. Phylogenet. Evol.* **163**, 107214 (2021). doi:10.1016/j.ympev.2021.107214 Medline

54. B. E. Goulet-Scott, A. G. Garner, R. Hopkins, Genomic analyses overturn two long-standing homoploid hybrid speciation hypotheses. *Evolution* **75**, 1699–1710 (2021). doi:10.1111/evo.14279 Medline

55. X. Ding, J. H. Xiao, L. Li, J. G. Conran, J. Li, Congruent species delimitation of two controversial gold-thread nanmu tree species based on morphological and restriction site associated DNA sequencing data. *J. Syst. Evol.* **57**, 234–246 (2019). doi:10.1111/jse.12433

56. M. M. Tavares, M. Ferro, B. S. S. Leal, C. Palma-Silva, Speciation with gene flow between two Neotropical sympatric species (*Pitcairnia* spp.: Bromeliaceae). *Ecol. Evol.* **12**, e8834 (2022). doi:10.1002/ece3.8834 Medline

57. Y. Sun, R. J. Abbott, Z. Lu, K. Mao, L. Zhang, X. Wang, D. Ru, J. Liu, Reticulate evolution within a spruce (Picea) species complex revealed by population genomic analysis. *Evolution* **72**, 2669–2681 (2018). doi:10.1111/evo.13624 Medline

58. D. Ru, Y. Sun, D. Wang, Y. Chen, T. Wang, Q. Hu, R. J. Abbott, J. Liu, Population genomic analysis reveals that homoploid hybrid speciation can be a lengthy process. *Mol. Ecol.* **27**, 4875–4887 (2018). doi:10.1111/mec.14909 Medline

59. H. Shang, J. Hess, M. Pickup, D. L. Field, P. K. Ingvarsson, J. Liu, C. Lexer, Evolution of strong reproductive isolation in plants: broad-scale patterns and lessons from a perennial model group. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **375**, 20190544 (2020).

60. S. Grünig, M. Fischer, C. Parisod, Recent hybrid speciation at the origin of the narrow endemic Pulmonaria helvetica. *Ann. Bot.* **127**, 21–31 (2021). doi:10.1093/aob/mcaa145 Medline

61. J. Ortego, L. L. Knowles, Incorporating interspecific interactions into phylogeographic models: A case study with Californian oaks. *Mol. Ecol.* **29**, 4510–4524 (2020). doi:10.1111/mec.15548 Medline

62. F. Wagner, T. Ott, M. Schall, U. Lautenschlager, R. Vogt, C. Oberprieler, Taming the Red Bastards: Hybridisation and species delimitation in the Rhodanthemum arundanum-group

(Compositae, Anthemideae). *Mol. Phylogenet. Evol.* **144**, 106702 (2020). doi:10.1016/j.ympev.2019.106702 Medline

63. S. Gramlich, N. D. Wagner, E. Hörandl, RAD-seq reveals genetic structure of the F_2-generation of natural willow hybrids (Salix L.) and a great potential for interspecific introgression. *BMC Plant Biol.* **18**, 317 (2018). doi:10.1186/s12870-018-1552-6 Medline

64. B. Nevado, S. A. Harris, M. A. Beaumont, S. J. Hiscock, Rapid homoploid hybrid speciation in British gardens: The origin of Oxford ragwort (Senecio squalidus). *Mol. Ecol.* **29**, 4221–4233 (2020). doi:10.1111/mec.15630 Medline

65. X.-S. Hu, D. A. Filatov, The large-X effect in plants: Increased species divergence and reduced gene flow on the Silene X-chromosome. *Mol. Ecol.* **25**, 2609–2619 (2016). doi:10.1111/mec.13427 Medline

66. A. Muyle, H. Martin, N. Zemp, M. Mollion, S. Gallina, R. Tavares, A. Silva, T. Bataillon, A. Widmer, S. Glémin, P. Touzet, G. A. B. Marais, Dioecy is associated with high genetic diversity and adaptation rates in the plant genus Silene. *Mol. Biol. Evol.* **38**, 805–818 (2021). doi:10.1093/molbev/msaa229 Medline

67. Y. Feng, H. P. Comes, Y.-X. Qiu, Phylogenomic insights into the temporal-spatial divergence history, evolution of leaf habit and hybridization in Stachyurus (Stachyuraceae). *Mol. Phylogenet. Evol.* **150**, 106878 (2020). doi:10.1016/j.ympev.2020.106878 Medline

68. A. M. Royer, M. A. Streisfeld, C. I. Smith, Population genomics of divergence within an obligate pollination mutualism: Selection maintains differences between Joshua tree species. *Am. J. Bot.* **103**, 1730–1741 (2016). doi:10.3732/ajb.1600069 Medline

69. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). doi:10.1038/nmeth.1923 Medline

70. N. Matasci, L.-H. Hung, Z. Yan, E. J. Carpenter, N. J. Wickett, S. Mirarab, N. Nguyen, T. Warnow, S. Ayyampalayam, M. Barker, J. G. Burleigh, M. A. Gitzendanner, E. Wafula, J. P. Der, C. W. dePamphilis, B. Roure, H. Philippe, B. R. Ruhfel, N. W. Miles, S. W. Graham, S. Mathews, B. Surek, M. Melkonian, D. E. Soltis, P. S. Soltis, C. Rothfels, L. Pokorny, J. A. Shaw, L. DeGironimo, D. W. Stevenson, J. C. Villarreal, T. Chen, T. M. Kutchan, M. Rolf, R. S. Baucom, M. K. Deyholos, R. Samudrala, Z. Tian, X. Wu, X. Sun, Y. Zhang, J. Wang, J. Leebens-Mack, G. K.-S. Wong, Data access for the 1,000 Plants (1KP) project. *Gigascience* **3**, 17 (2014). doi:10.1186/2047-217X-3-17 Medline

71. J. M. Catchen, A. Amores, P. Hohenlohe, W. Cresko, J. H. Postlethwait, Stacks: Building and genotyping Loci de novo from short-read sequences. *G3* **1**, 171–182 (2011). doi:10.1534/g3.111.000240 Medline

72. J. Catchen, P. A. Hohenlohe, S. Bassham, A. Amores, W. A. Cresko, Stacks: An analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140 (2013). doi:10.1111/mec.12354 Medline

73. J. R. Paris, J. R. Stevens, J. M. Catchen, Lost in parameter space: A road map for stacks. *Methods Ecol. Evol.* **8**, 1360–1373 (2017). doi:10.1111/2041-210X.12775

74. L. Excoffier, N. Marchi, D. A. Marques, R. Matthey-Doret, A. Gouy, V. C. Sousa, fastsimcoal2: Demographic inference under complex evolutionary scenarios. *Bioinformatics* **37**, 4882–4885 (2021). doi:10.1093/bioinformatics/btab468 Medline

75. R. Gutenkunst, R. Hernandez, S. Williamson, C. Bustamante, Diffusion approximations for demographic inference: DaDi. *Nat. Prec.* 10.1038/npre.2010.4594.1 (2010).

76. J. Jouganous, W. Long, A. P. Ragsdale, S. Gravel, Inferring the joint demographic history of multiple populations: Beyond the diffusion approximation. *Genetics* **206**, 1549–1567 (2017). doi:10.1534/genetics.117.200493 Medline

77. M. Malinsky, M. Matschiner, H. Svardal, Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Mol. Ecol. Resour.* **21**, 584–595 (2021). doi:10.1111/1755-0998.13265 Medline

78. T. E. Cruickshank, M. W. Hahn, Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* **23**, 3133–3157 (2014). doi:10.1111/mec.12796 Medline

79. C. Roux, C. Fraïsse, V. Castric, X. Vekemans, G. H. Pogson, N. Bierne, Can we continue to neglect genomic variation in introgression rates when inferring the history of speciation? A case study in a Mytilus hybrid zone. *J. Evol. Biol.* **27**, 1662–1675 (2014). doi:10.1111/jeb.12425 Medline

80. T. Leroy, C. Roux, L. Villate, C. Bodénès, J. Romiguier, J. A. P. Paiva, C. Dossat, J.-M. Aury, C. Plomion, A. Kremer, Extensive recent secondary contacts between four European white oak species. *New Phytol.* **214**, 865–878 (2017). doi:10.1111/nph.14413 Medline

81. T. Capblancq, C. Roux, F. Boyer, F. Legeai, M. Joron, L. Després, Untangling the contribution of adaptive versus non-adaptive processes in the evolution of reproductive isolation between Coenonympha butterflies. bioRxiv 2024.11.29.625973 [Preprint] (2024); https://doi.org/10.1101/2024.11.29.625973.

82. T. Koppetsch, M. Malinsky, M. Matschiner, Towards reliable detection of introgression in the presence of among-species rate variation. *Syst. Biol.* **73**, 769–788 (2024). doi:10.1093/sysbio/syae028 Medline

83. N. Galtier, An approximate likelihood method reveals ancient gene flow between human, chimpanzee and gorilla. *Peer Community J.* **4**, e3 (2024). doi:10.24072/pcjournal.359

84. J. Ross-Ibarra, S. I. Wright, J. P. Foxe, A. Kawabe, L. DeRose-Wilson, G. Gos, D. Charlesworth, B. S. Gaut, Patterns of polymorphism and demographic history in natural populations of Arabidopsis lyrata. *PLOS ONE* **3**, e2411 (2008). doi:10.1371/journal.pone.0002411 Medline

85. R. R. Hudson, Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337–338 (2002). doi:10.1093/bioinformatics/18.2.337 Medline

86. B. Charlesworth, M. T. Morgan, D. Charlesworth, The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303 (1993). doi:10.1093/genetics/134.4.1289 Medline

87. N. Barton, B. O. Bengtsson, The barrier to genetic exchange between hybridising populations. *Heredity* **57**, 357–376 (1986). doi:10.1038/hdy.1986.135 Medline

88. M. A. Beaumont, Approximate Bayesian computation in evolution and ecology. *Annu. Rev. Ecol. Evol. Syst.* **41**, 379–406 (2010). doi:10.1146/annurev-ecolsys-102209-144621

89. F. Tajima, Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437–460 (1983). doi:10.1093/genetics/105.2.437 Medline

90. G. A. Watterson, On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**, 256–276 (1975). doi:10.1016/0040-5809(75)90020-9 Medline

91. F. Tajima, Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989). doi:10.1093/genetics/123.3.585 Medline

92. M. Nei, W.-H. Li, Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5269–5273 (1979). doi:10.1073/pnas.76.10.5269 Medline

93. S. Wright, Isolation by Distance. *Genetics* **28**, 114–138 (1943). doi:10.1093/genetics/28.2.114 Medline

94. J. Wakeley, J. Hey, Estimating ancestral population parameters. *Genetics* **145**, 847–855 (1997). doi:10.1093/genetics/145.3.847 Medline

95. J. L. Hamrick, J. D. Nason, "Gene flow in forest trees" in *Forest conservation genetics: Principles and practice*, A. Young, D. Boshier, T. Boyle, Eds. (CSIRO, 2000), pp. 81–90.

96. F. Austerlitz, S. Mariette, N. Machon, P.-H. Gouyon, B. Godelle, Effects of colonization processes on genetic diversity: Differences between annual plants and tree species. *Genetics* **154**, 1309–1321 (2000). doi:10.1093/genetics/154.3.1309 Medline

97. B. Igic, J. R. Kohn, The distribution of plant mating systems: Study bias against obligately outcrossing species. *Evolution* **60**, 1098–1103 (2006). Medline

98. B. Anderson, J. Pannell, S. Billiard, C. Burgarella, H. de Boer, M. Dufay, A. J. Helmstetter, M. Méndez, S. P. Otto, D. Roze, H. Sauquet, D. Schoen, J. Schönenberger, M. Vallejo-Marin, R. Zenil-Ferguson, J. Käfer, S. Glémin, Opposing effects of plant traits on diversification. *iScience* **26**, 106362 (2023). doi:10.1016/j.isci.2023.106362 Medline

99. J. A. P. Hamlin, M. S. Hibbins, L. C. Moyle, Assessing biological factors affecting postspeciation introgression. *Evol. Lett.* **4**, 137–154 (2020). doi:10.1002/evl3.159 Medline

100. J. Goudet, Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* **5**, 184–186 (2005). doi:10.1111/j.1471-8286.2004.00828.x

101. S. Kumar, G. Stecher, M. Suleski, S. B. Hedges, TimeTree: A resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017). doi:10.1093/molbev/msx116 Medline

102. B. Nevado, G. W. Atchison, C. E. Hughes, D. A. Filatov, Widespread adaptive evolution during repeated evolutionary radiations in New World lupins. *Nat. Commun.* **7**, 12384 (2016). doi:10.1038/ncomms12384 Medline