

SUPPORTING INFORMATION

SUPPLEMENTARY RESULTS

The impact of incomplete lineage sorting

We visualized the distributions of the four-taxon training dataset in a three-dimensional space. The data displayed full coverage of the entire space restricted by the non-negative and sum-to one constraints (**Fig. S29a**). Note that the proportion of the corresponding topology was not always 100%, even for scenarios without migration (**Fig. S29b**). Therefore, the phylogenies could be influenced by either gene flow or incomplete lineage sorting (ILS).

To evaluate the impact of ILS, we examined a set of data having topology (((P1, P2), P3), O), and the proportions of two alternative topologies were found to decrease as the existence time of the ancestral species increased (**Fig. S30a**). As the intensity of ILS depends on the population size and divergence time, there are two alternative topologies with the same probability for a four-taxon case without gene flow:

$$P = \frac{1}{3} \times e^{-\tau} \quad (1) \quad (\text{Hudson 1983})$$

where τ is the time difference between two speciation events in units of $2N$ generations.

Given that the estimated topology proportions were influenced by both gene flow and ILS, we wanted to determine a threshold that was less affected by demographic histories to further distinguish patterns of introgression from ILS. As the equation shows that probabilities of two discordant topologies are the greatest when the difference of two divergence times is zero, we evaluated the distributions of topology proportions on the simulated datasets with two speciation events at the same time (**Fig. S30b**). The three

topologies had the same mean value of 0.33, and there was no significant difference between the three absolute divergence times (Kruskal-Wallis test, P value = 0.3264). Next, we merged the estimated values of these datasets and calculated the 95% quantile of probability density function based on the kernel density estimation, which was 0.5 for 5-kb and 10-kb windows, and 0.4 for 50-kb and 100-kb windows (**Fig. S30c**). The threshold indicated that the chance of a topology proportion caused by ILS greater than this value was smaller than 5%. Thus, the false positive rate (FPR) was less than 5% for the four-taxon model under all evolutionary scenarios. This threshold is conservative because the true intensity of ILS is likely less than the above assumption.

ILS in the five-taxon model is more complex, and the fourteen alternative topologies do not have the same probabilities (Pamilo and Nei 1988; Rosenberg 2002). For example, for the symmetric phylogeny (with topo M as an example), the probability of each alternative topology is:

$$\begin{aligned} P(A) = P(J) &= \frac{1}{3} \times (1 - e^{-\tau_1}) \times e^{-\tau_2} + \frac{1}{6} \times \frac{1}{3} \times e^{-\tau_1} \times e^{-\tau_2} \\ &= \frac{1}{3} \times e^{-\tau_2} - \frac{5}{18} \times e^{-(\tau_1+\tau_2)} \quad (2) \quad (\text{Rosenberg 2002}) \end{aligned}$$

$$\begin{aligned} P(F) = P(I) &= \frac{1}{3} \times e^{-\tau_1} \times (1 - e^{-\tau_2}) + \frac{1}{6} \times \frac{1}{3} \times e^{-\tau_1} \times e^{-\tau_2} \\ &= \frac{1}{3} \times e^{-\tau_1} - \frac{5}{18} \times e^{-(\tau_1+\tau_2)} \quad (3) \quad (\text{Rosenberg 2002}) \end{aligned}$$

$$P(N) = P(O) = \frac{2}{6} \times \frac{1}{3} \times e^{-\tau_1} \times e^{-\tau_2} = \frac{1}{9} \times e^{-(\tau_1+\tau_2)} \quad (4) \quad (\text{Rosenberg 2002})$$

$$\begin{aligned} P(B) = P(C) = P(D) = P(E) = P(G) = P(H) = P(K) = P(L) \\ = \frac{1}{6} \times \frac{1}{3} \times e^{-\tau_1} \times e^{-\tau_2} = \frac{1}{18} \times e^{-(\tau_1+\tau_2)} \quad (5) \quad (\text{Rosenberg 2002}) \end{aligned}$$

which can be divided into four groups, and are determined by the time difference, τ_1 and τ_2 , determined by speciation events. We also noticed that the τ values to reach the

maximum proportion were not consistent for different topology groups (**Fig. S30d–e**).

The probabilities of alternative topologies for the asymmetric model (with topo A as the species tree) are:

$$\begin{aligned}
P(M) &= \frac{1}{3} \times (1 - e^{-\tau_1}) \times e^{-\tau_2} \\
&\quad + \left[\frac{1}{3} \times \frac{1}{3} \times \left(\frac{3}{2} \times e^{-\tau_2} - \frac{3}{2} \times e^{-3 \times \tau_2} \right) + \frac{2}{6} \times \frac{1}{3} \times e^{-3 \times \tau_2} \right] \times e^{-\tau_1} \\
&= \frac{1}{3} \times e^{-\tau_2} - \frac{1}{6} \times e^{-(\tau_1 + \tau_2)} - \frac{1}{18} \\
&\quad \times e^{-(\tau_1 + 3 \times \tau_2)} \quad (6) \quad (\text{Rosenberg 2002})
\end{aligned}$$

$$\begin{aligned}
P(J) &= \frac{1}{3} \times (1 - e^{-\tau_1}) \times e^{-\tau_2} \\
&\quad + \left[\frac{1}{3} \times \frac{1}{3} \times \left(\frac{3}{2} \times e^{-\tau_2} - \frac{3}{2} \times e^{-3 \times \tau_2} \right) + \frac{1}{6} \times \frac{1}{3} \times e^{-3 \times \tau_2} \right] \times e^{-\tau_1} \\
&= \frac{1}{3} \times e^{-\tau_2} - \frac{1}{6} \times e^{-(\tau_1 + \tau_2)} - \frac{1}{9} \\
&\quad \times e^{-(\tau_1 + 3 \times \tau_2)} \quad (7) \quad (\text{Rosenberg 2002})
\end{aligned}$$

$$\begin{aligned}
P(B) = P(C) &= \left[\frac{1}{3} \times \left(1 - \frac{3}{2} \times e^{-\tau_2} + \frac{1}{2} \times e^{-3 \times \tau_2} \right) + \frac{1}{3} \times \frac{1}{3} \right. \\
&\quad \left. \times \left(\frac{3}{2} \times e^{-\tau_2} - \frac{3}{2} \times e^{-3 \times \tau_2} \right) + \frac{1}{6} \times \frac{1}{3} \times e^{-3 \times \tau_2} \right] \times e^{-\tau_1} \\
&= \frac{1}{3} \times e^{-\tau_1} - \frac{1}{3} \times e^{-(\tau_1 + \tau_2)} + \frac{1}{18} \\
&\quad \times e^{-(\tau_1 + 3 \times \tau_2)} \quad (8) \quad (\text{Rosenberg 2002})
\end{aligned}$$

$$\begin{aligned}
P(N) = P(O) &= \left[\frac{1}{3} \times \frac{1}{3} \times \left(\frac{3}{2} \times e^{-\tau_2} - \frac{3}{2} \times e^{-3 \times \tau_2} \right) + \frac{2}{6} \times \frac{1}{3} \times e^{-3 \times \tau_2} \right] \times e^{-\tau_1} \\
&= \frac{1}{6} \times e^{-(\tau_1 + \tau_2)} - \frac{1}{18} \times e^{-(\tau_1 + 3 \times \tau_2)} \quad (9) \quad (\text{Rosenberg 2002})
\end{aligned}$$

$$\begin{aligned}
P(D) = P(G) &= \left[\frac{1}{3} \times \frac{1}{3} \times \left(\frac{3}{2} \times e^{-\tau_2} - \frac{3}{2} \times e^{-3 \times \tau_2} \right) + \frac{1}{6} \times \frac{1}{3} \times e^{-3 \times \tau_2} \right] \times e^{-\tau_1} \\
&= \frac{1}{6} \times e^{-(\tau_1 + \tau_2)} - \frac{1}{9} \times e^{-(\tau_1 + 3 \times \tau_2)} \quad (10) \quad (\text{Rosenberg 2002})
\end{aligned}$$

$$\begin{aligned}
P(E) = P(F) = P(H) = P(I) = P(K) = P(L) &= \frac{1}{6} \times \frac{1}{3} \times e^{-3 \times \tau_2} \times e^{-\tau_1} \\
&= \frac{1}{18} \times e^{-(\tau_1 + 3 \times \tau_2)} \quad (11) \quad (\text{Rosenberg 2002})
\end{aligned}$$

which showed a similar landscape, but contained more topology groups (**Fig. S30f–g**).

Thus, it is not straightforward to distinguish between introgression and ILS for a

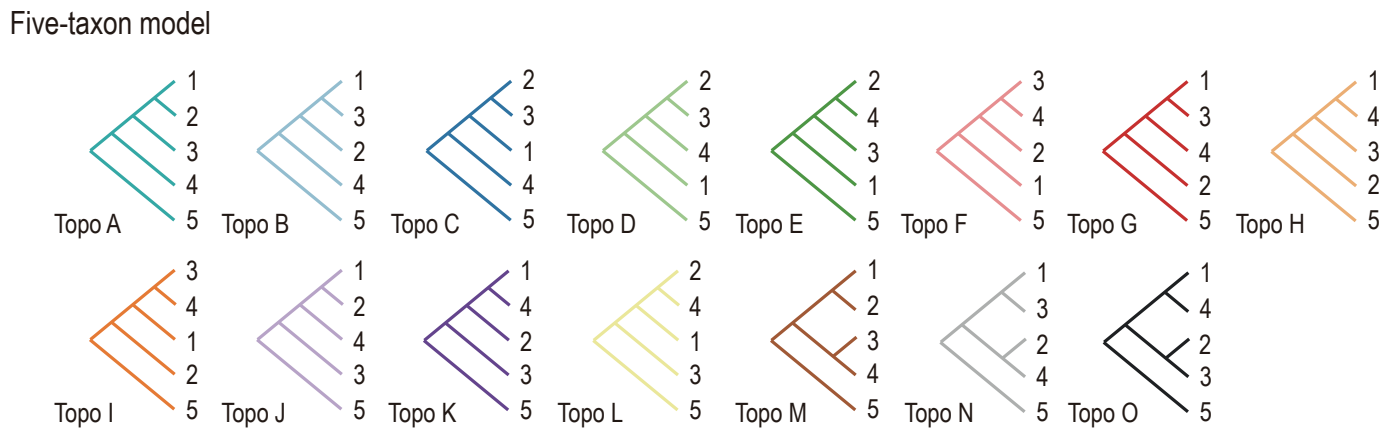
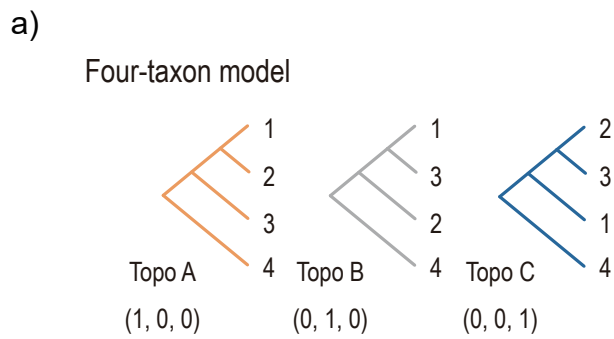
five-taxon system. Instead, it is a better way to evaluate the intensity of ILS by performing demographic modeling with specific parameters. If the demographic history is not available, the previously discussed threshold of the four-taxon model can be applied as a strict standard to detect introgression and ensure that the FPR is less than 5%, because the probability of a given alternative topology in the five-taxon model is also smaller than or equal to one-third, like the four-taxon model.

SUPPLEMENTARY REFERENCE

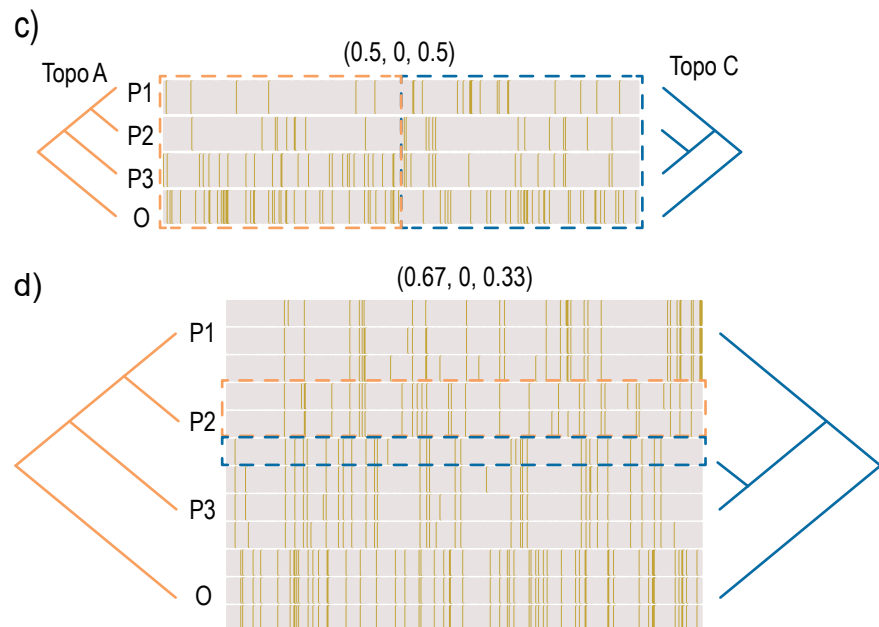
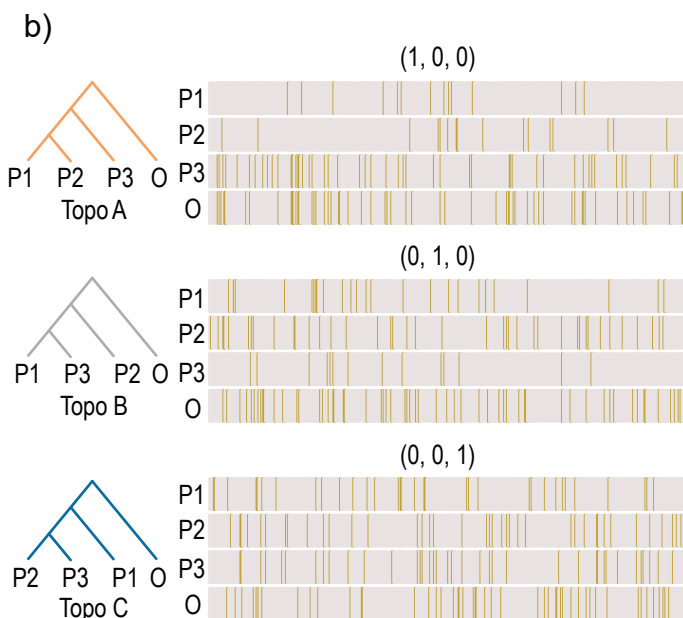
Hudson R.R. 1983. Testing the constant-rate neutral allele model with protein sequence data. *Evolution* 37:203–217.

Pamilo P., Nei M. 1988. Relationships between gene trees and species trees. *Mol. Biol. Evol.* 5:568–583.

Rosenberg N.A. 2002. The probability of topological concordance of gene trees and species trees. *Theor. Popul. Biol.* 61:225–247.

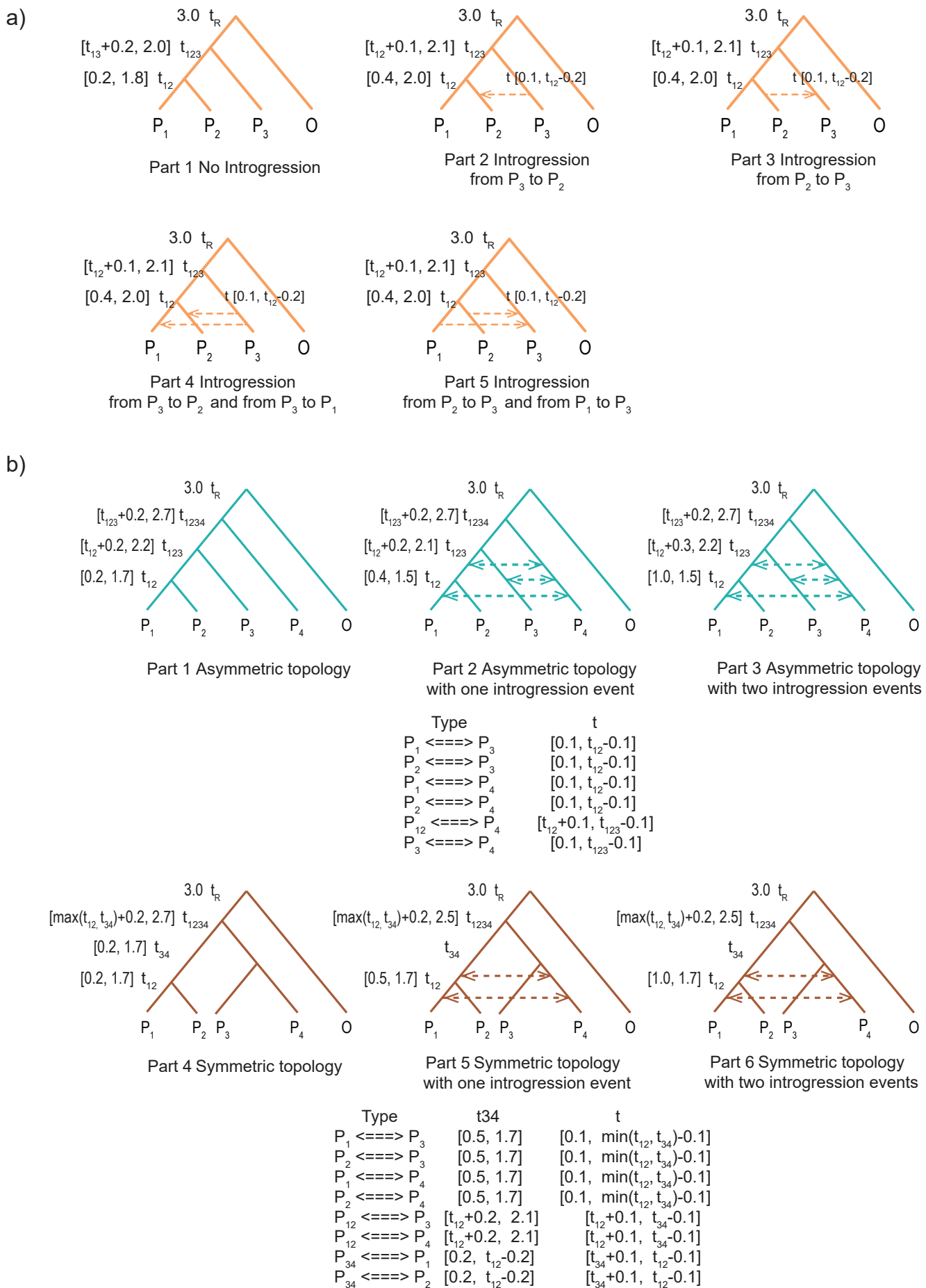


Major allele Minor allele



Supplementary Figure S1. Possible relationships of three ingroup taxa and an outgroup taxon.

(a) There are three rooted topology trees for three ingroup taxa and one outgroup taxon, and the three categories are encoded in the one-hot format. Similarly, fifteen topologies exist when there are four ingroup taxa. (b) An example of sequence alignment with one haplotype for each taxon is shown for each possible topology, in which major alleles and minor alleles are labeled in light grey and dark yellow, respectively. (c) An example containing two segments with different evolutionary histories, and the label is specified as (0.5, 0, 0.5). (d) An example with multiple haplotypes per taxon shows individual heterogeneity in taxon P2, with two haplotypes similar to P1 and one similar to P3, likely due to ancestral polymorphism or gene flow. A vector (0.67, 0, 0.33) is therefore used to represent the relationships among taxa.



Supplementary Figure S2. Scenarios used in model training.

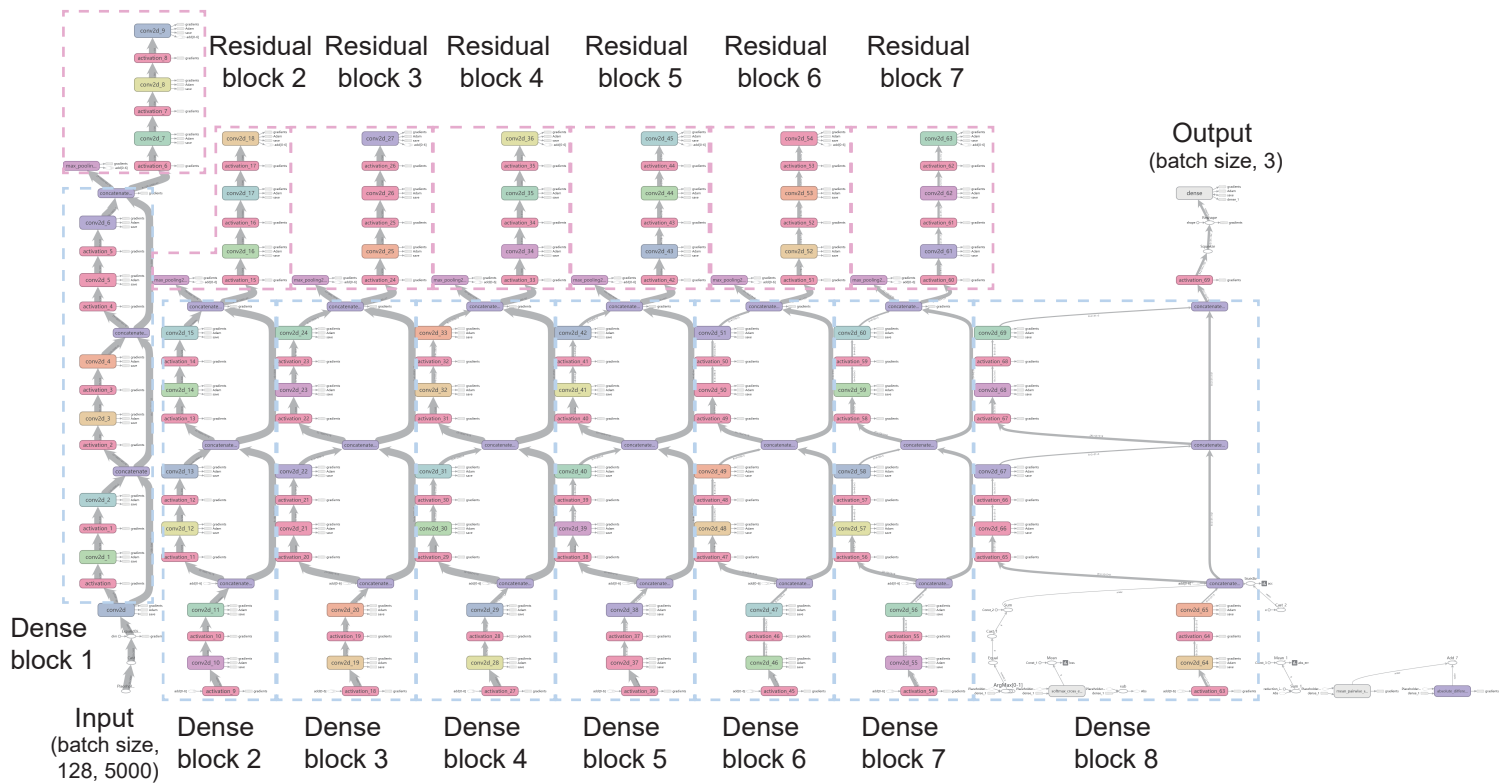
(a) A training dataset including four taxa with different numbers of introgression events was simulated for the four-taxon model according to multiple evolutionary scenarios. The split times and introgression times were restricted in a range given in units of $4N$ generations.

(b) Similarly, a training dataset including asymmetric and symmetric topologies and different numbers of introgression events was simulated for the five-taxon model.

a)

Residual block 1

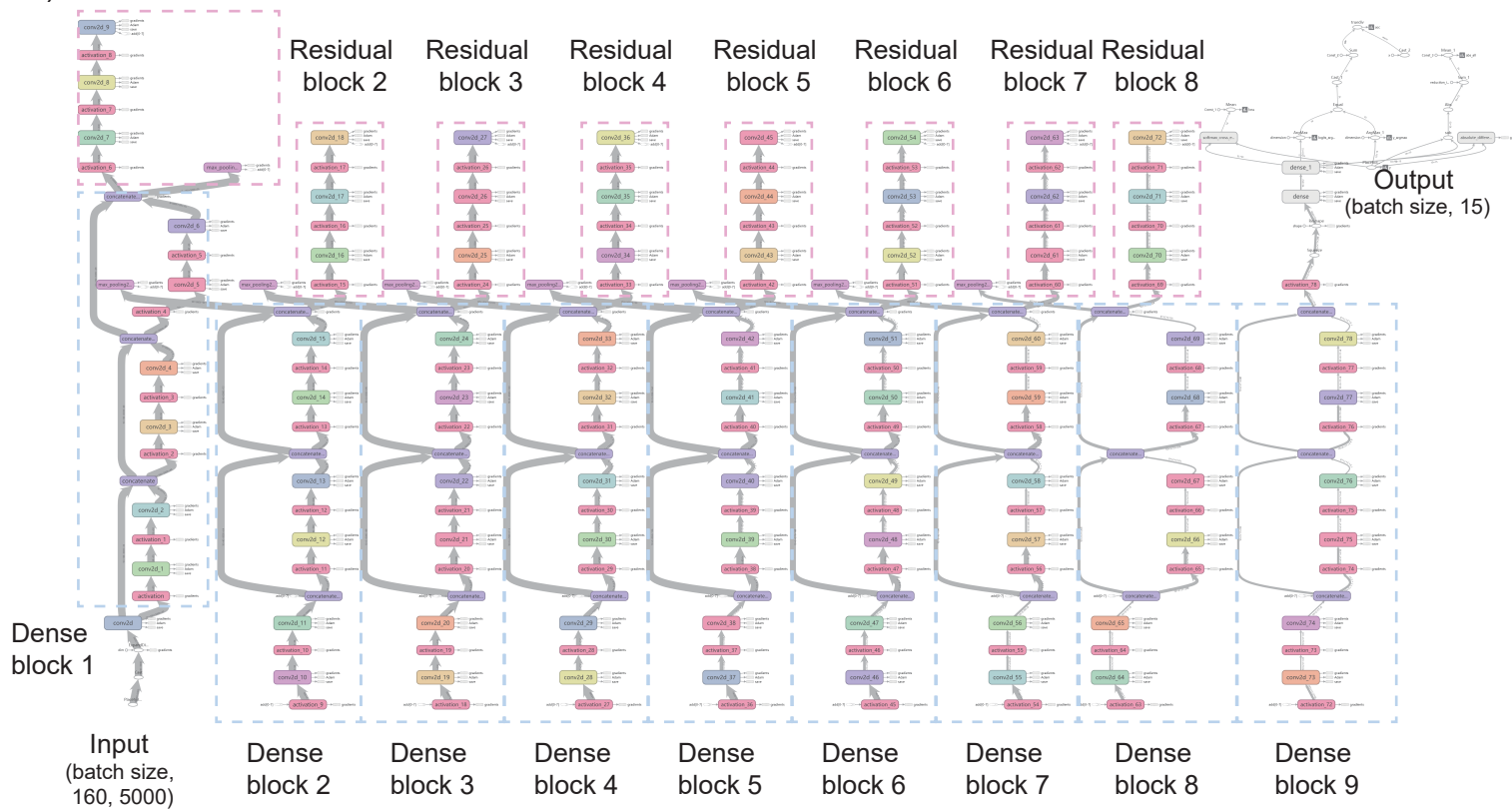
Four-taxon model

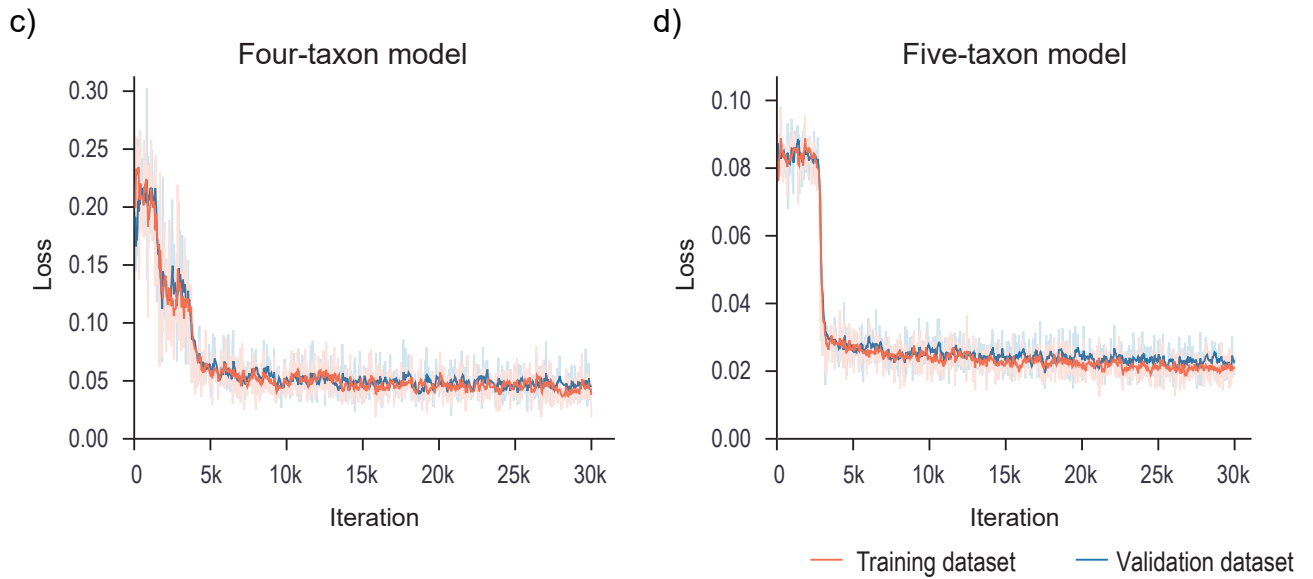


b)

Residual block 1

Five-taxon model

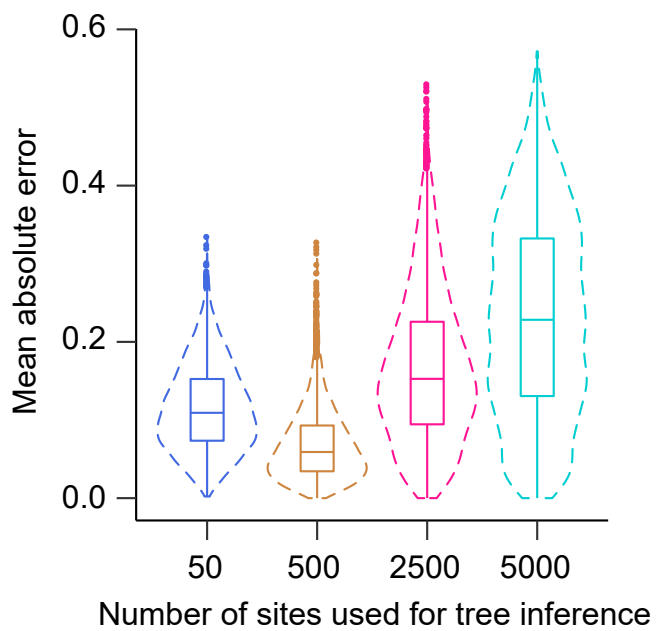




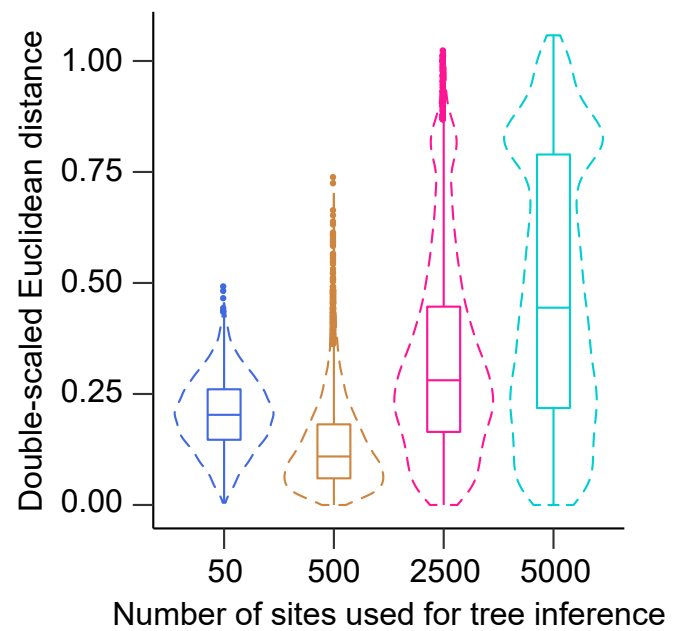
Supplementary Figure S3. The architectures and training process of CNNs.

The two diagrams show detailed architectures of the two convolution neural networks used for four taxa (a) and five taxa (b). The boxed indicate the boundary of Dense blocks and Residual blocks. The mean absolute error (MAE) was used as the loss function to optimize the neural networks. The MAE losses are plotted against the training step to measure the performance of the neural networks using four-taxon data (c) and five-taxon data (d). The orange line and blue line indicate the training dataset and the validation dataset, respectively.

a)

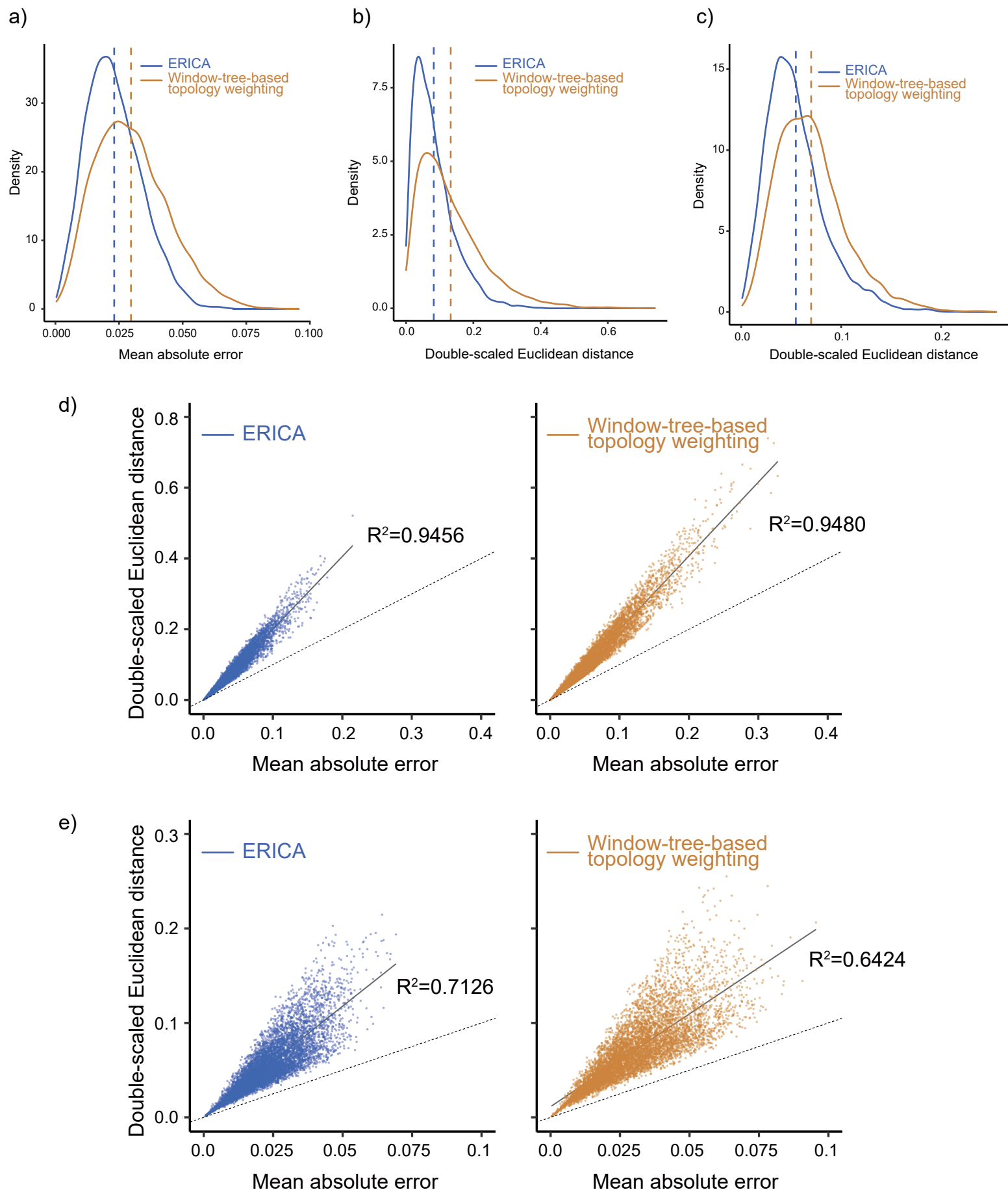


b)



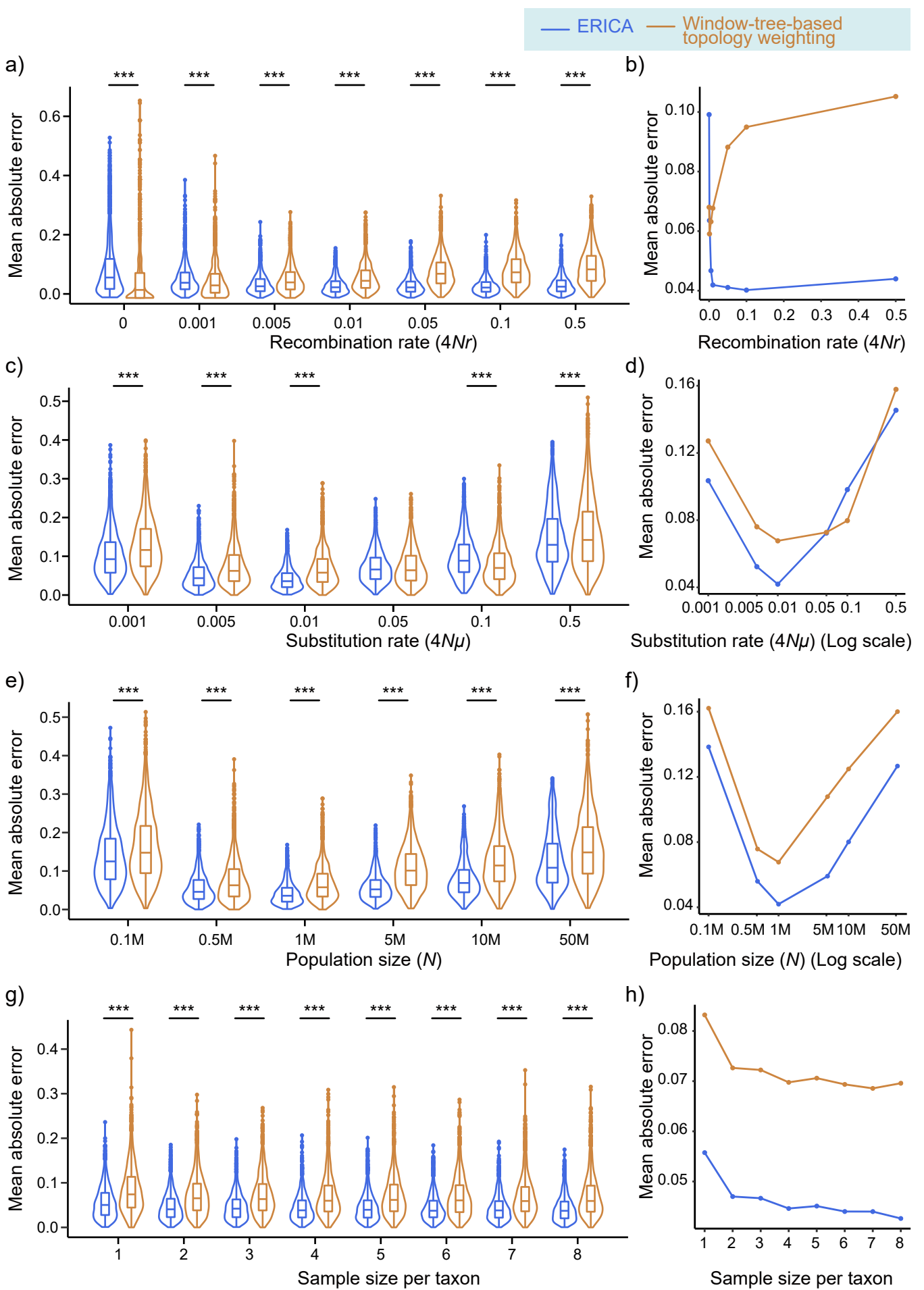
Supplementary Figure S4. The error rates of the window-tree-based topology weighting methods with different window sizes.

The distributions of MAEs (a) and scaled Euclidean distances (b) of the maximum-likelihood trees (evaluated using test dataset D1, $n = 6,030$) constructed using different window sizes.

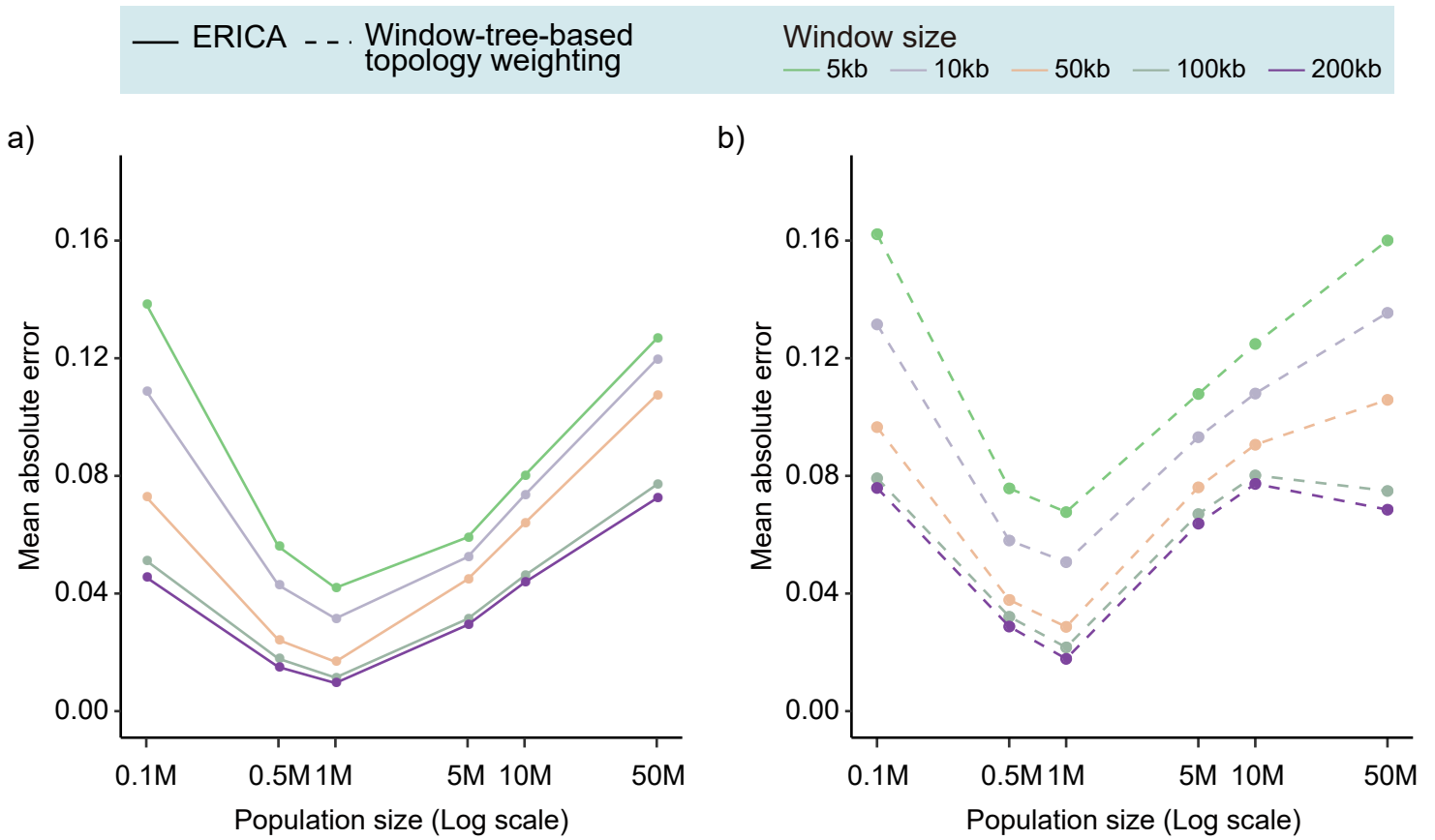


Supplementary Figure S5. The error rates of ERICA models and tree-based methods.

The distributions of MAEs for the five-taxon model (a) (evaluated using dataset D1, $n = 7,410$) and scaled Euclidean distances for the four-taxon (b) (evaluated using dataset D1, $n = 6,030$) and five-taxon models (c) (evaluated using dataset D1, $n = 7,410$). Dot plots show the strong correlations between mean absolute errors and scaled Euclidean distances for both four-taxon (d) and five-taxon cases (e).

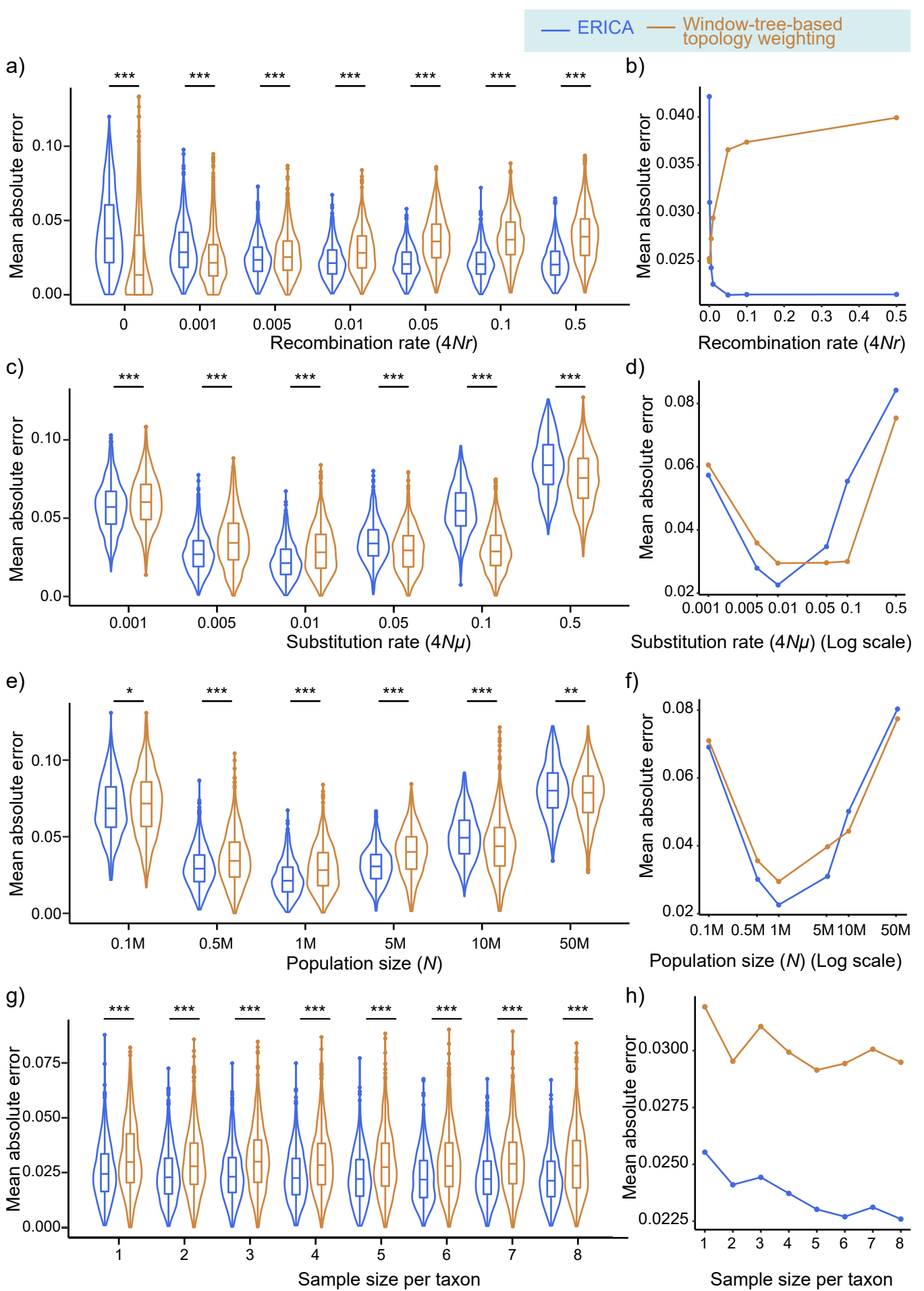


Supplementary Figure S6. Robustness of the four-taxon ERICA model for phylogenetic inference. Violin plots show the distributions of MAEs for the four-taxon ERICA model and the window-tree-based topology weighting method for different recombination rates ($4Nr$) (test dataset D2) (a), substitution rates ($4N\mu$) (test dataset D3) (c), effective population sizes (N) (test dataset D4) (e), and numbers of samples per taxon (test dataset D5) (g). Line plots show the average value for each dataset (b, d, f, h). *** indicates Mann-Whitney U test P value < 0.001 . 2,010 windows were used for each dataset.

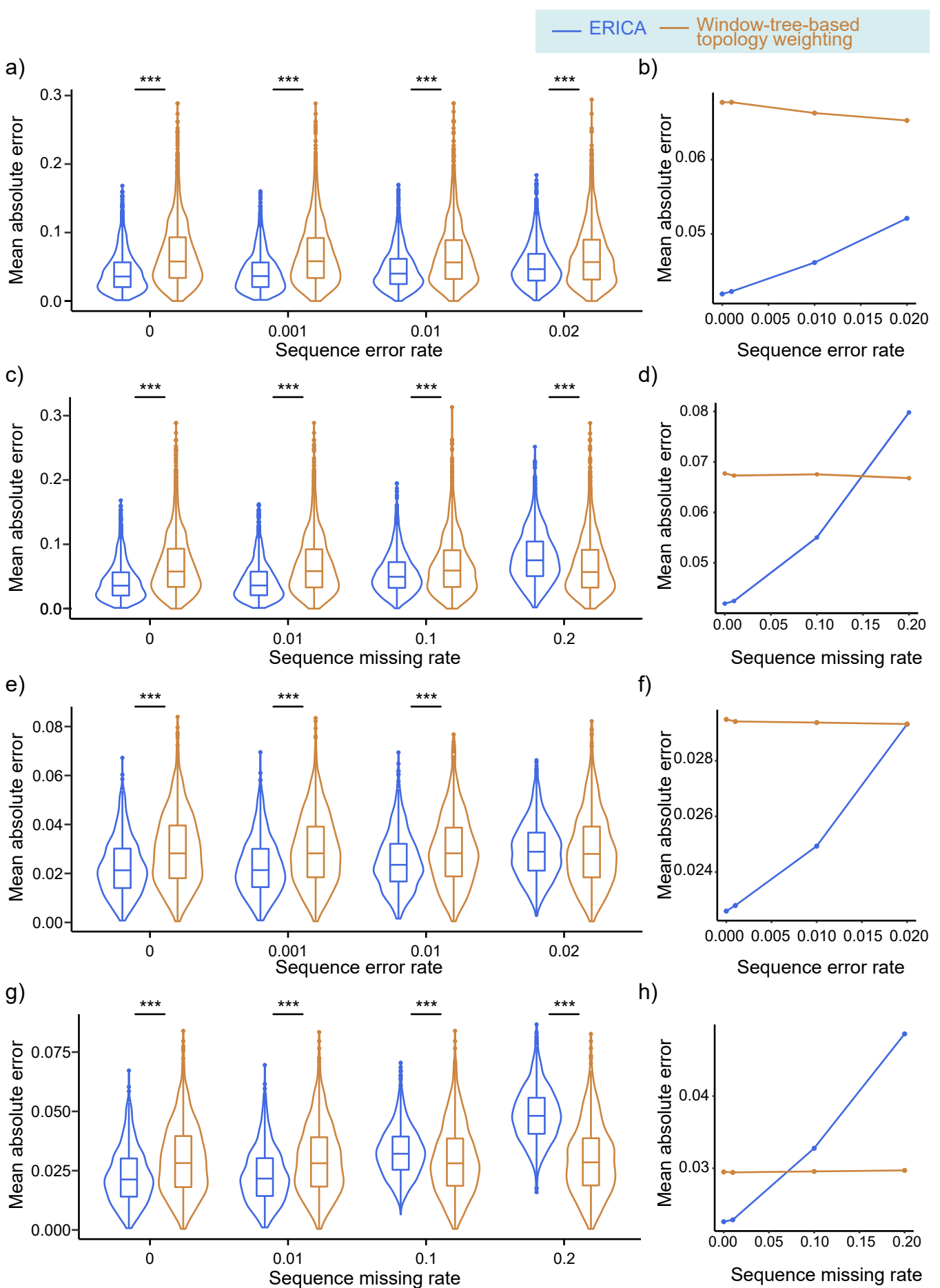


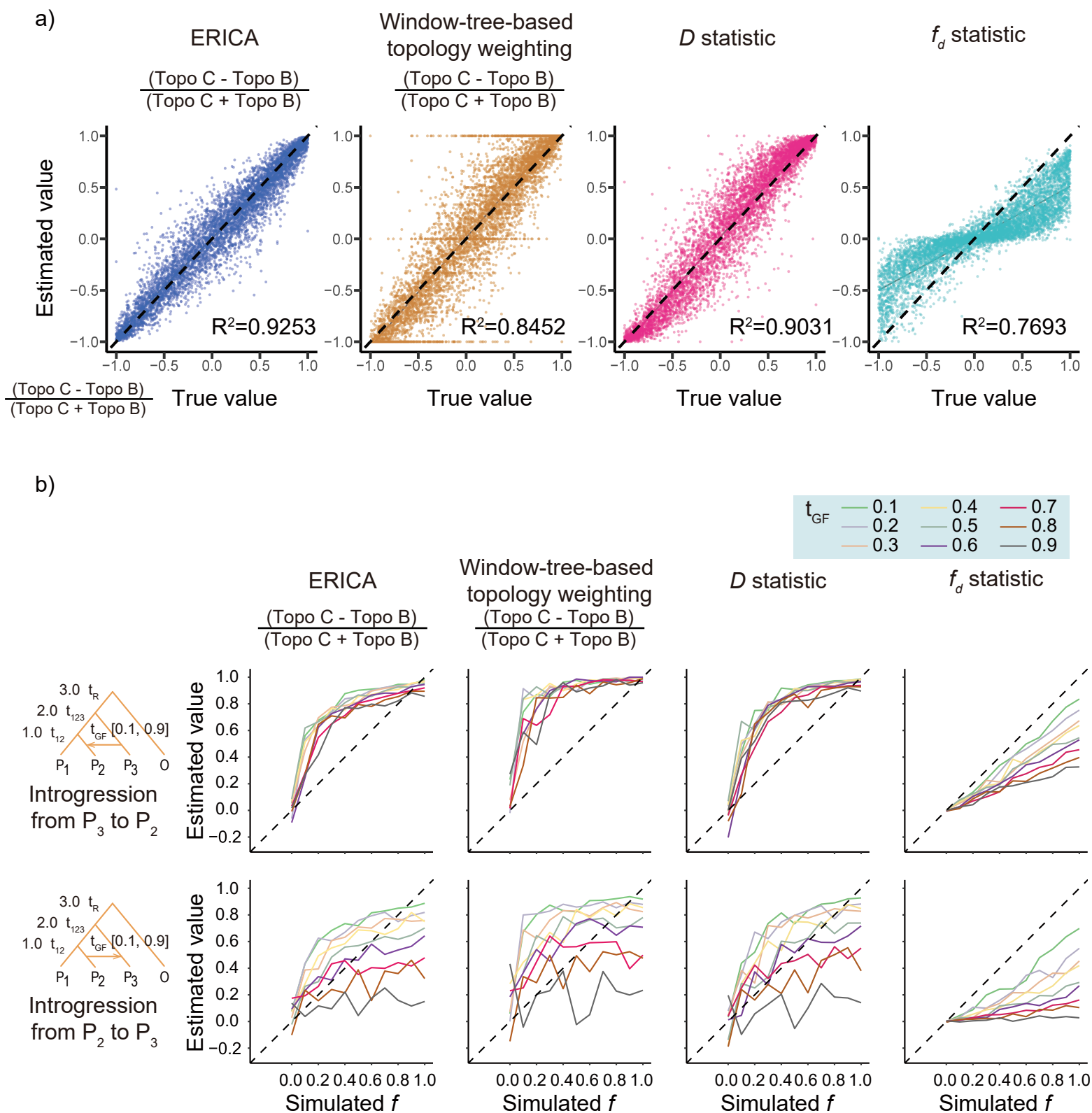
Supplementary Figure S7. Effect of window size on phylogenetic inference.

Line plots show the average MAEs for different population sizes for the four-taxon ERICA model (a) and the tree-based approach (b). The line colors indicate different window sizes. (evaluated using dataset D4, 5-kb window: $n = 2,010$; 10-kb window: $n = 1,005$; 50-kb window: $n = 201$; 100-kb window: $n = 100$; 200-kb window: $n = 50$).

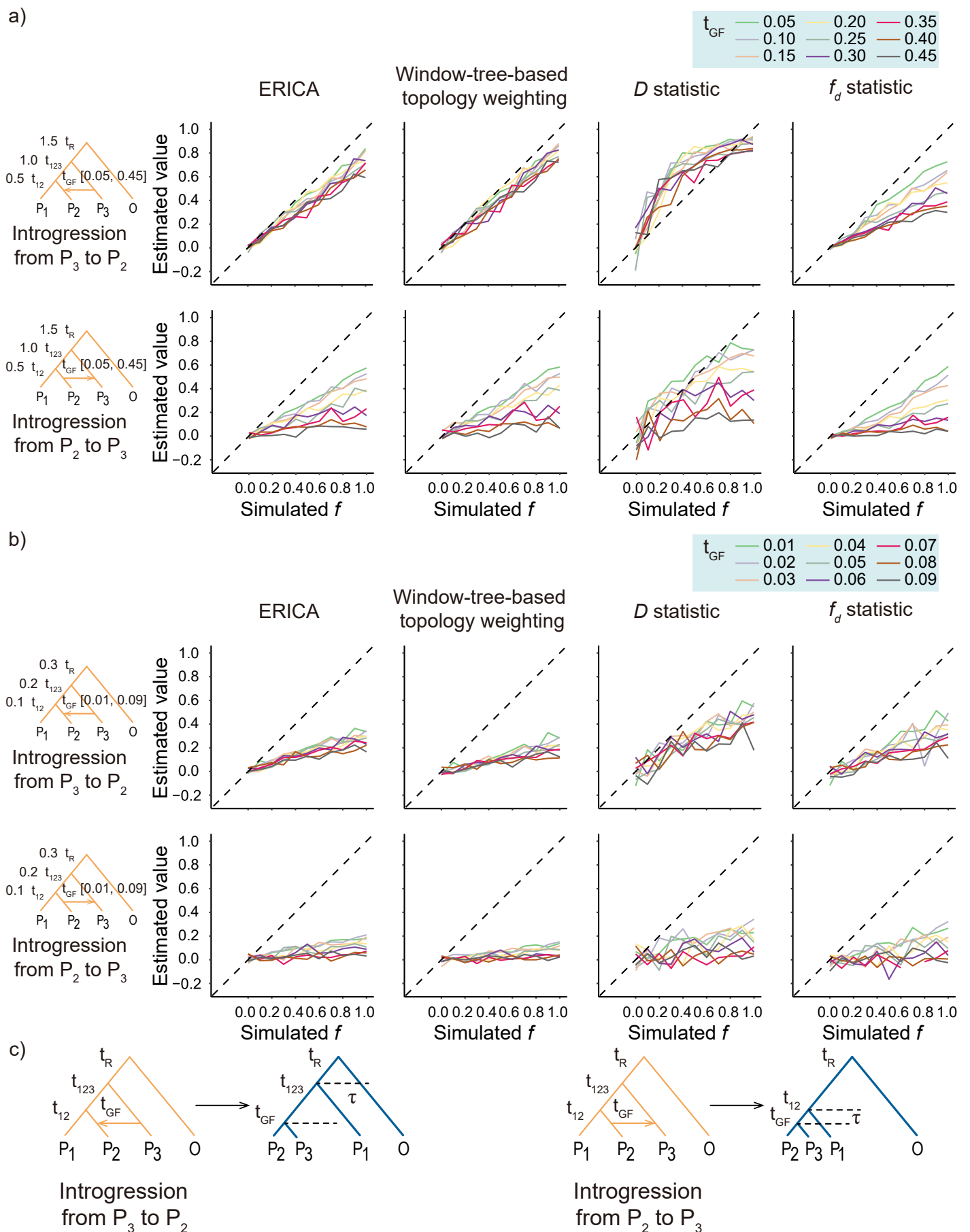


Supplementary Figure S8. Robustness of the five-taxon ERICA model for phylogenetic inference. Violin plots show the distributions of MAEs for the five-taxon ERICA model and the window-tree-based topology weighting method for different recombination rates ($4Nr$) (test dataset D2) (a), substitution rates ($4N\mu$) (test dataset D3) (c), effective population sizes (N) (test dataset D4) (e), and numbers of samples per taxon (test dataset D5) (g). Line plots show the average value for each dataset (b, d, f, h). *** indicates Mann-Whitney U test P value < 0.001 . ** indicates P value < 0.01 , * indicates P value < 0.05 . 916 windows were used for each dataset.



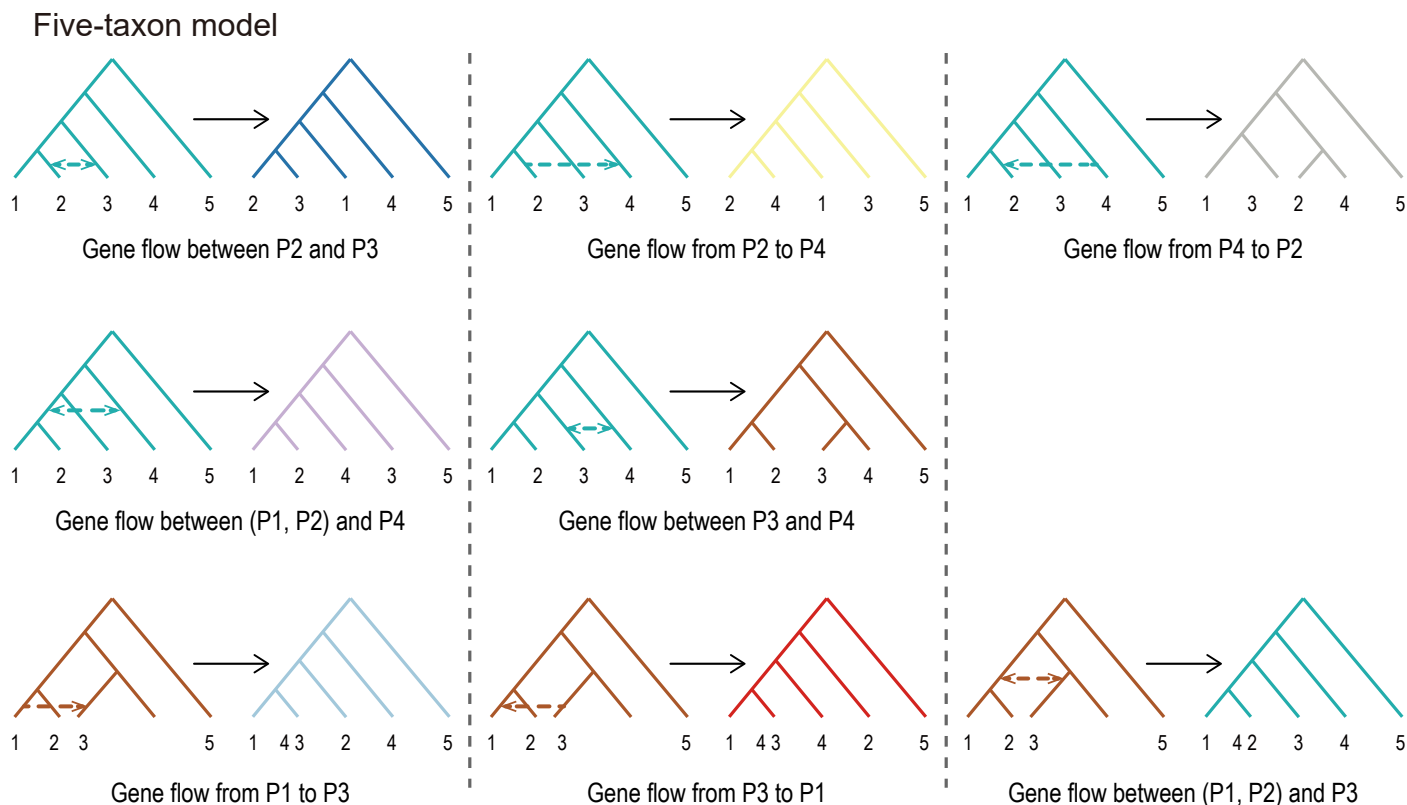
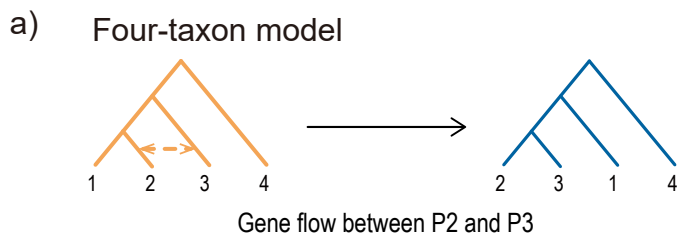


Supplementary Figure S10. Introgression proportion inference using relative topology differences.(a) The relationships between true values and estimated values of the relative differences between two alternative topologies (see Materials and Methods) (evaluated using test dataset D1, $n = 6,030$). (b) The estimated values of introgression fractions (f) for different introgression times (t_{GF}) (evaluated using test dataset D8). 20 replicates were performed for each scenario.



Supplementary Figure S11. Introgression proportion inference using absolute topology differences.

The estimated values of introgression fractions (f) with different species split times. 11 different values were simulated. The split times t_{12} , t_{123} , and t_R were set to 0.5, 1.0, and 1.5 (in units of $4N$ generations, test dataset D9) (a), and 0.1, 0.2, and 0.3 (test dataset D10) (b). The time of gene flow ranged from 10% to 90% of the split time (t_{12}). Gene flows from P_3 to P_2 (the top panel) and in the opposite direction (the bottom panel) were tested. The values of ERICA and the tree-based method were calculated as the absolute difference between two alternative topologies (see Materials and Methods). 20 replicates were performed for each scenario. (c) Expected genealogies arose from gene flow between non-sister species. The topologies were identical but the internal branch lengths were different in the two introgression scenarios. Compared with the introgression from P_3 into P_2 , when the direction was P_2 to P_3 , the probability of ILS increased as a result of a smaller divergence time (τ).

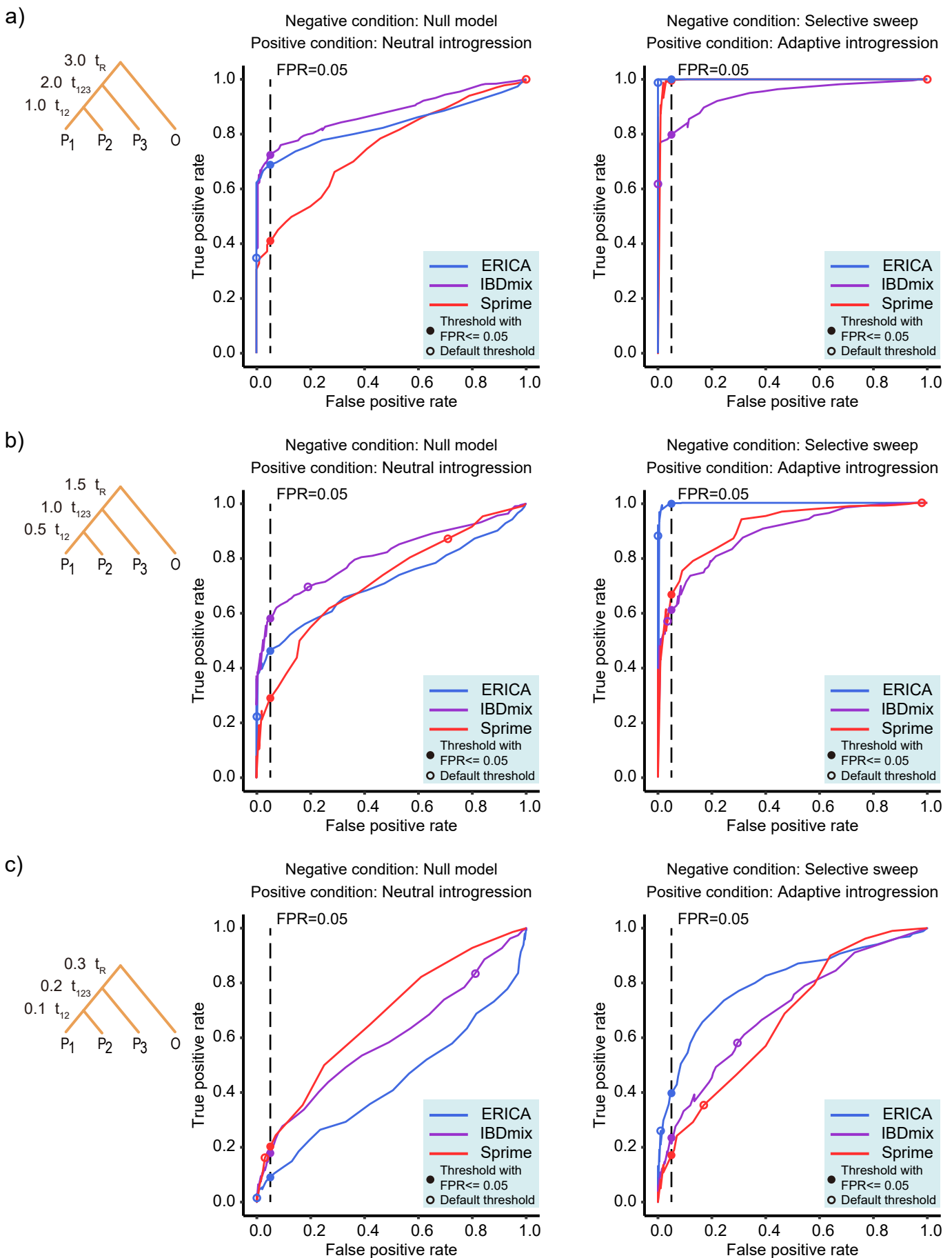


b)

Model	Example species tree	Direction of gene flow	Corresponding topology
Four-taxon model	A (((P1, P2), P3), O)	P1<->P3	B (((P1, P3), P2), O)
		P2<->P3	C (((P2, P3), P1), O)
Asymmetric five-taxon model	A ((((P1, P2), P3), P4), O)	P1<->P3	B ((((P1, P3), P2), P4), O)
		P1 ->P4	K ((((P1, P4), P2), P3), O)
		P4 ->P1	O (((P1, P4), (P2, P3)), O)
		P2<->P3	C (((P2, P3), P1), P4), O)
		P2 ->P4	L (((P2, P4), P1), P3), O)
		P4 ->P2	N (((P1, P3), (P2, P4)), O)
		P12 <-> P4	J (((P1, P2), P4), P3), O)
		P3 <->P4	M (((P1, P2), (P3, P4)), O)
Symmetric five-taxon model	M (((P1, P2), (P3, P4)), O)	P1 ->P3	B ((((P1, P3), P2), P4), O)
		P3 ->P1	G ((((P1, P3), P4), P2), O)
		P1 ->P4	K ((((P1, P4), P2), P3), O)
		P4 ->P1	H (((P1, P4), P3), P2), O)
		P2 ->P3	C (((P2, P3), P1), P4), O)
		P3 ->P2	D (((P2, P3), P4), P1), O)
		P2 ->P4	L (((P2, P4), P1), P3), O)
		P4 ->P2	E (((P2, P4), P3), P1), O)
P12 <-> P3	A (((P1, P2), P3), P4), O)		
P12 <-> P4	J (((P1, P2), P4), P3), O)		

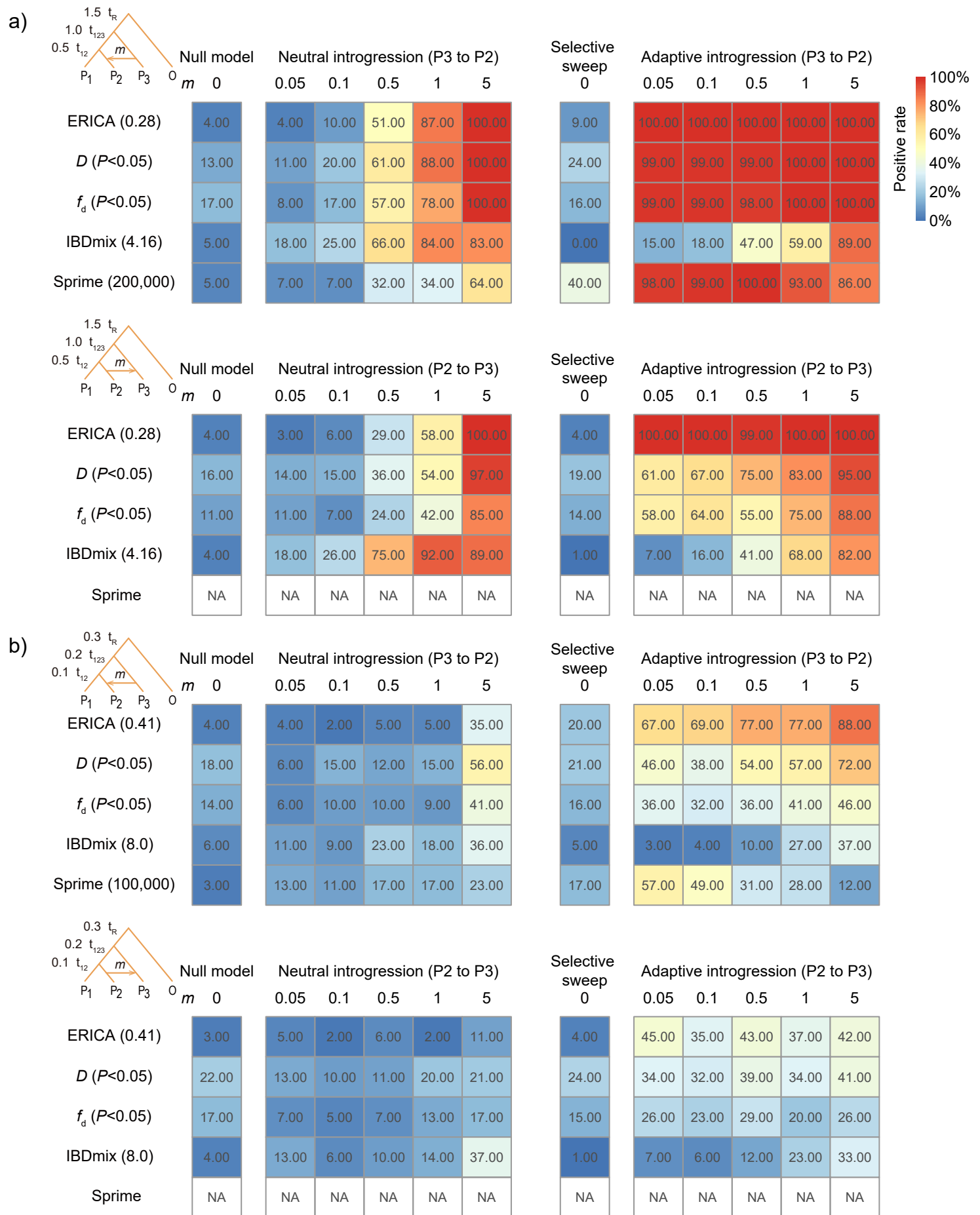
Supplementary Figure S12. The impact of gene flow on the topological structures.

The possible gene flow between non-sister species and their resulting topologies for a specific four-taxon phylogeny (Topo A) and five-taxon phylogenies (asymmetric Topo A and symmetric Topo M) are illustrated (a) and listed below (b). The corresponding relationships between gene flow directions and topology changes under other species trees could be inferred by simply exchanging orders of taxa.



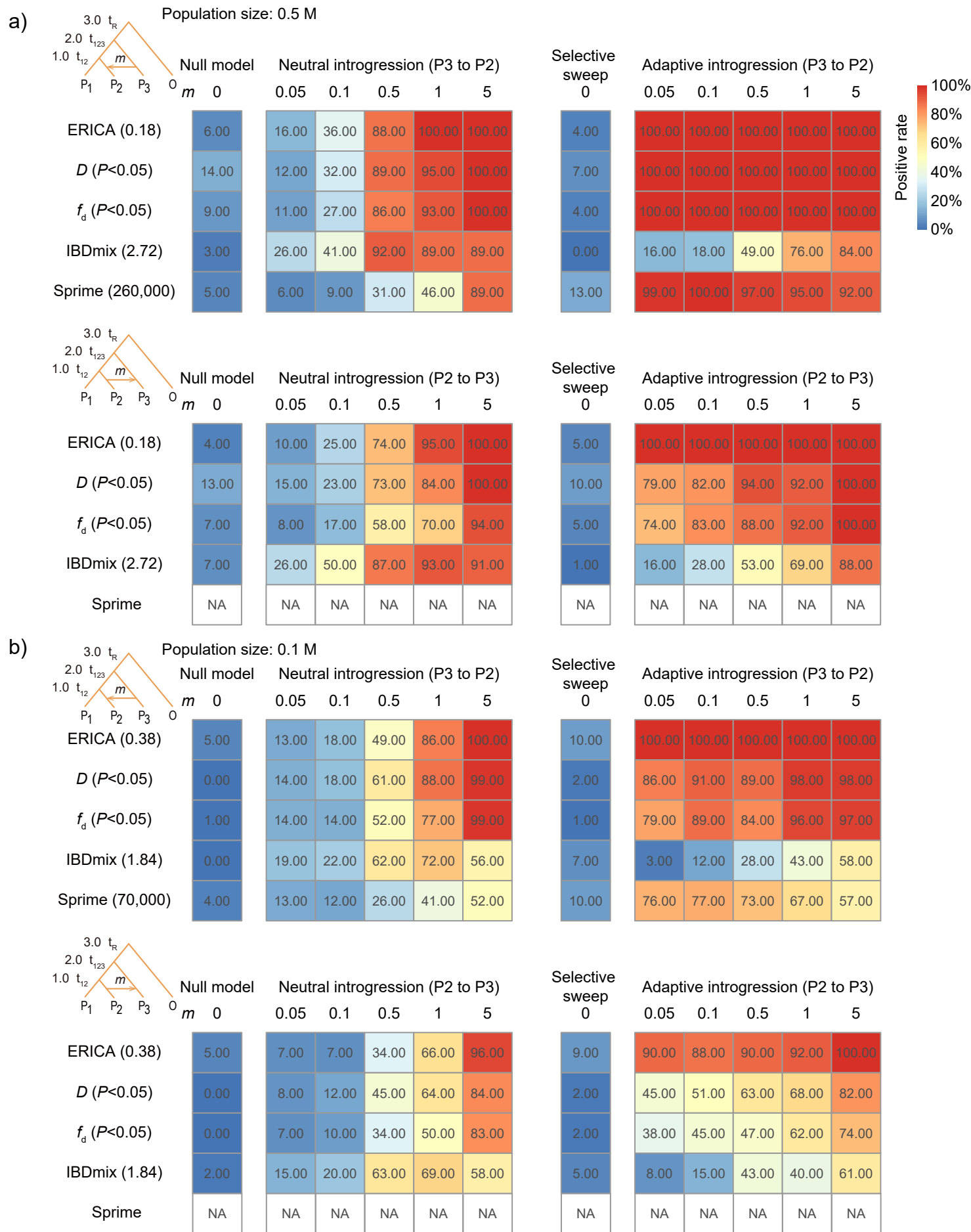
Supplementary Figure S13. The receiver operating characteristic (ROC) curves of the four-taxon ERICA model and other algorithms for introgression detection.

The ROC curves of the ERICA, IBDmix, and Sprime methods for neutral and adaptive introgressions with different species split times. The split times t_{12} , t_{123} , and t_R were set to 1, 2, and 3 (in units of $4N$ generations, test dataset D11) (a), and 0.5, 1.0, and 1.5 (test dataset D12) (b), and 0.1, 0.2, and 0.3 (test dataset D13) (c). The solid circles indicate the TPRs with FPRs equal to 5%, and the hollow circles indicate the TPRs and FPRs obtained using the software default thresholds for classification. ($n = 1,200$ for the ERICA and IBDmix methods, and $n = 600$ for Sprime, as it was only applicable to the detection of gene flow from P3 to P2.)



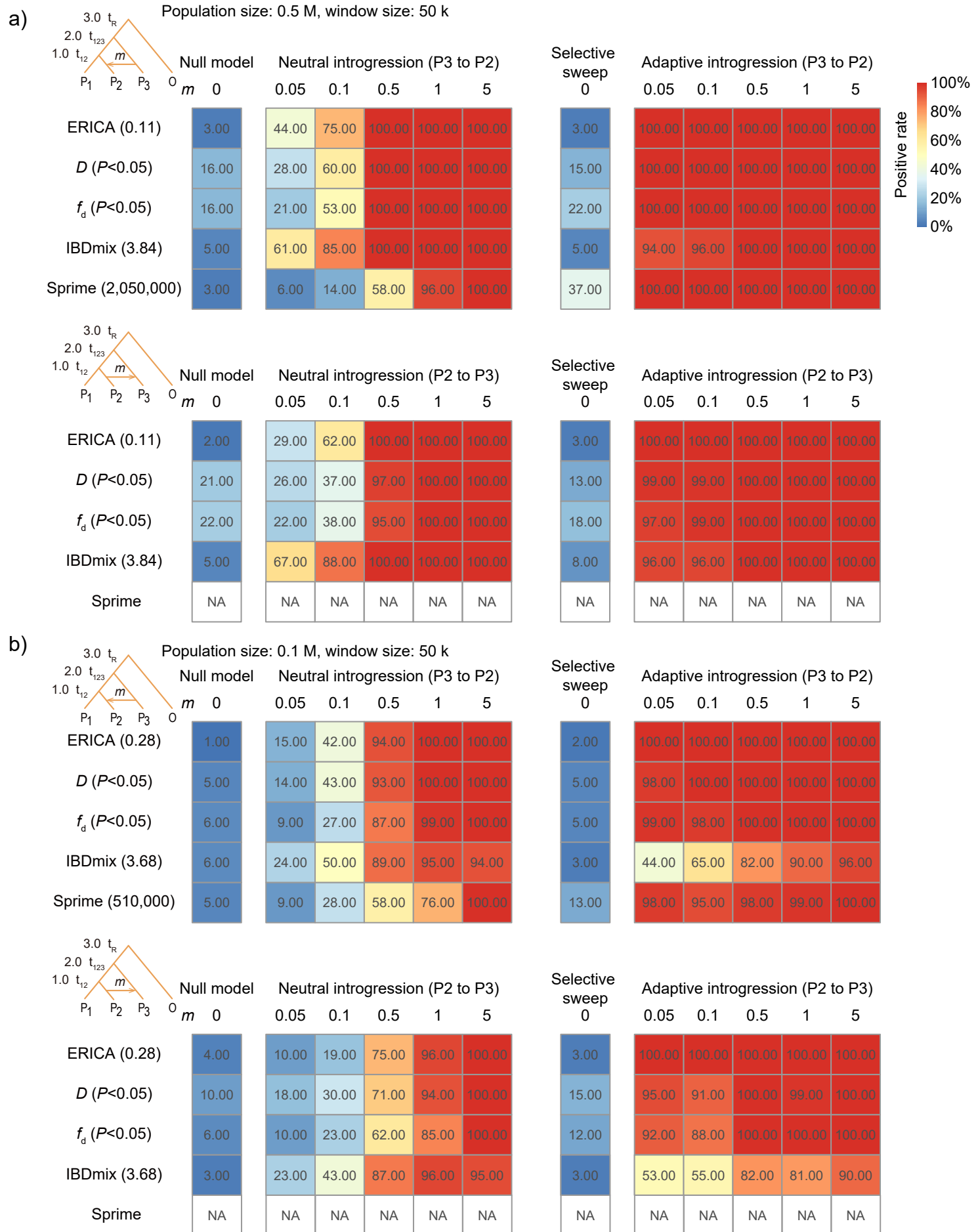
Supplementary Figure S14. Performance of the four-taxon ERICA model and other algorithms for introgression detection with different split times.

The heatmaps show the FPRs for scenarios without gene flow and the TPRs for neutral and adaptive introgression at different migration rates (0.05–5 migrants per generation). The split times t_{12} , t_{123} , and t_R were set to 0.5, 1.0, and 1.5 (in units of $4N$ generations, test dataset D12) (a), and 0.1, 0.2, and 0.3 (test dataset D13) (b). The numbers in the brackets are the thresholds used for each method. NA indicates that Sprime was not applicable for the detection of gene flow from P2 to P3. 100 replicates were tested for each case.

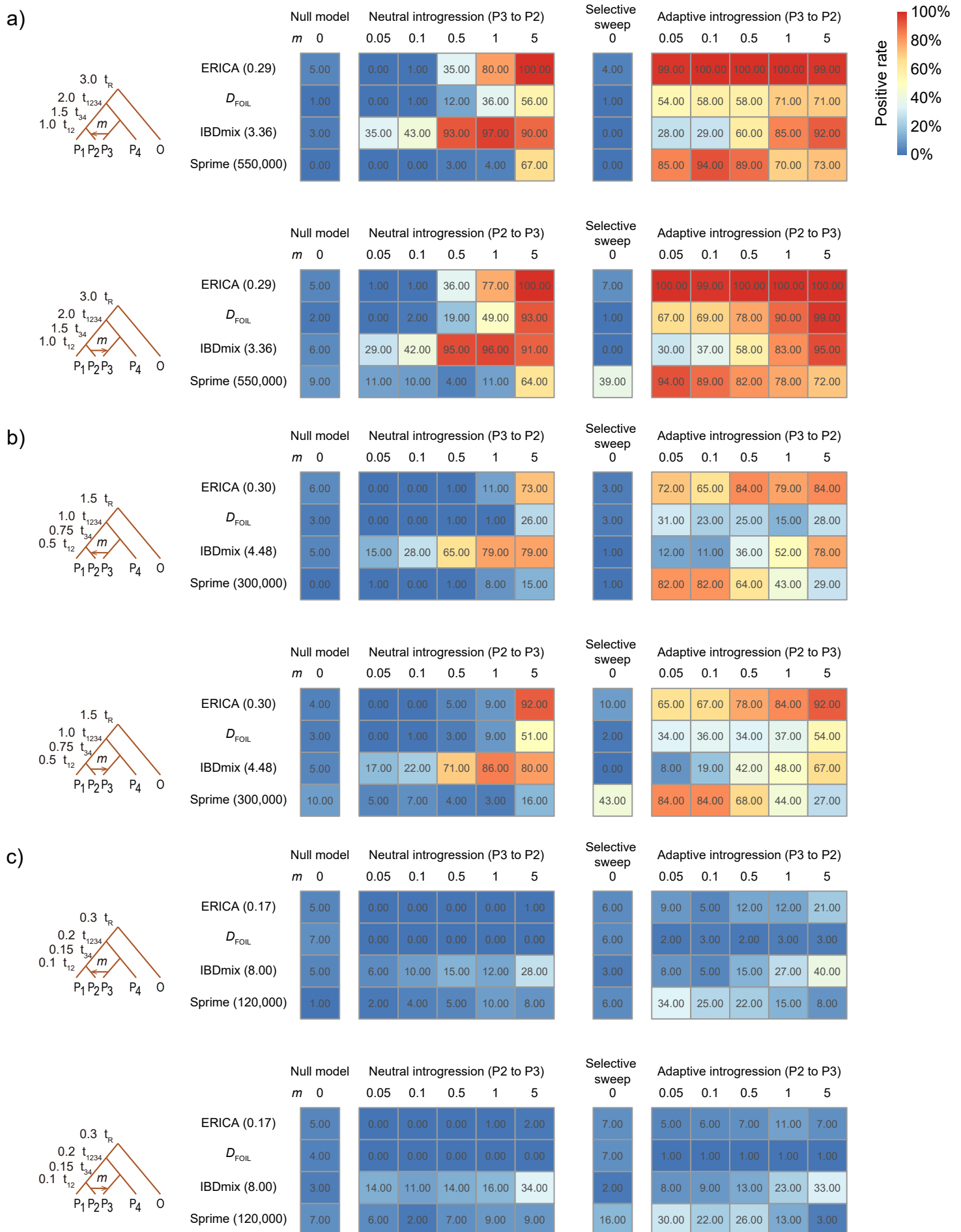


Supplementary Figure S15. Performance of the four-taxon ERICA model and other algorithms for introgression detection for different population sizes.

The heatmaps show the FPRs for scenarios without gene flow and the TPRs for neutral and adaptive introgression at different migration rates (0.05–5 migrants per generation). The population sizes were set to 0.5 M (test dataset D14) (a) and 0.1 M (test dataset D15) (b). The numbers in the brackets are the thresholds used for each method. NA indicates that Sprime was not applicable for the detection of gene flow from P2 to P3. 100 replicates were tested for each case.

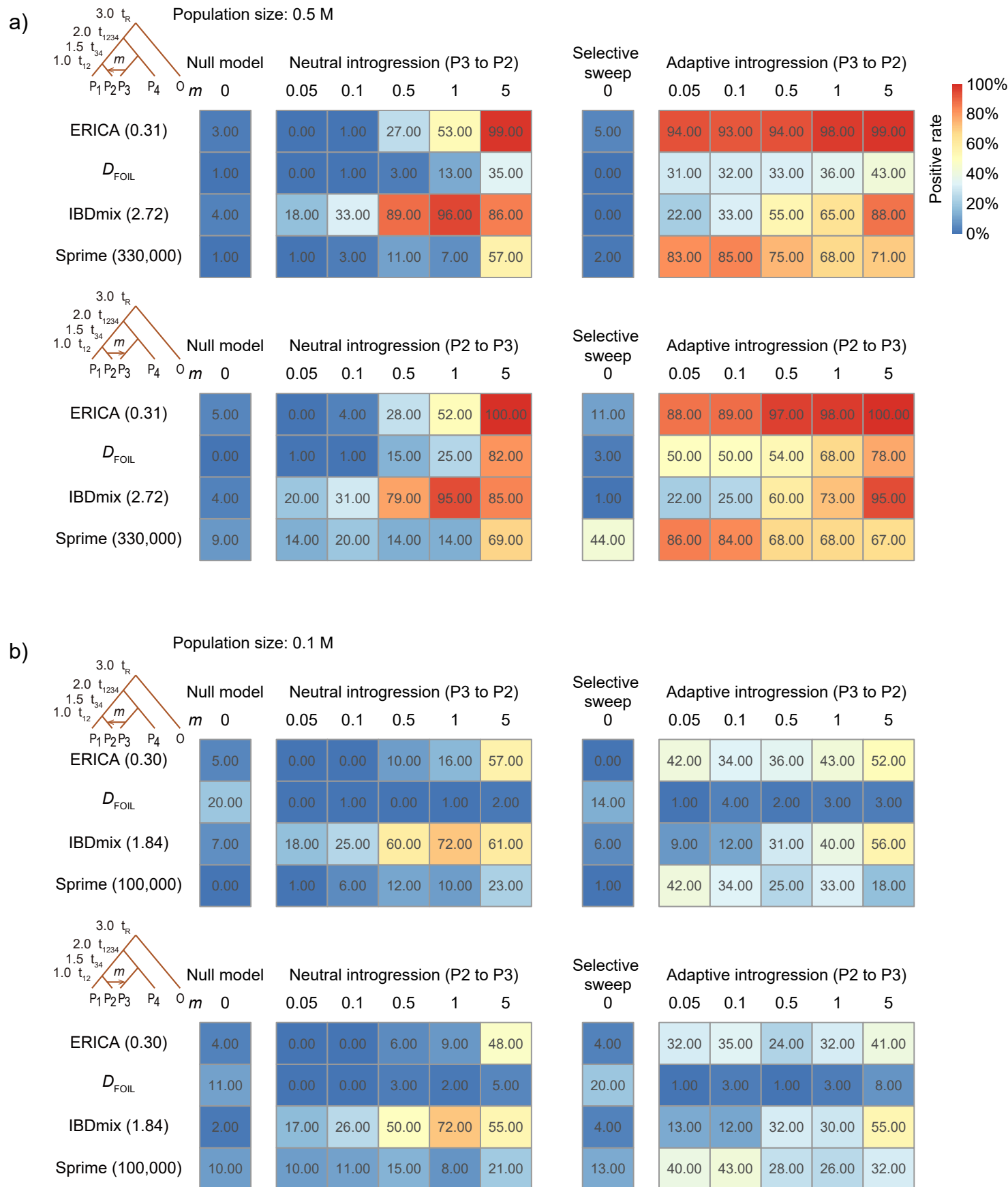


Supplementary Figure S16. Effect of window size on introgression detection for the four-taxon data. The heatmaps show the FPRs for scenarios without gene flow and the TPRs for neutral and adaptive introgression at different migration rates (0.05–5 migrants per generation). The population sizes were set to 0.5 M (test dataset D16) (a) and 0.1 M (test dataset D17) (b). Each window was 50 kb in length. The numbers in the brackets are the thresholds used for each method. NA indicates that Sprime was not applicable for the detection of gene flow from P2 to P3. 100 replicates were tested for each case.



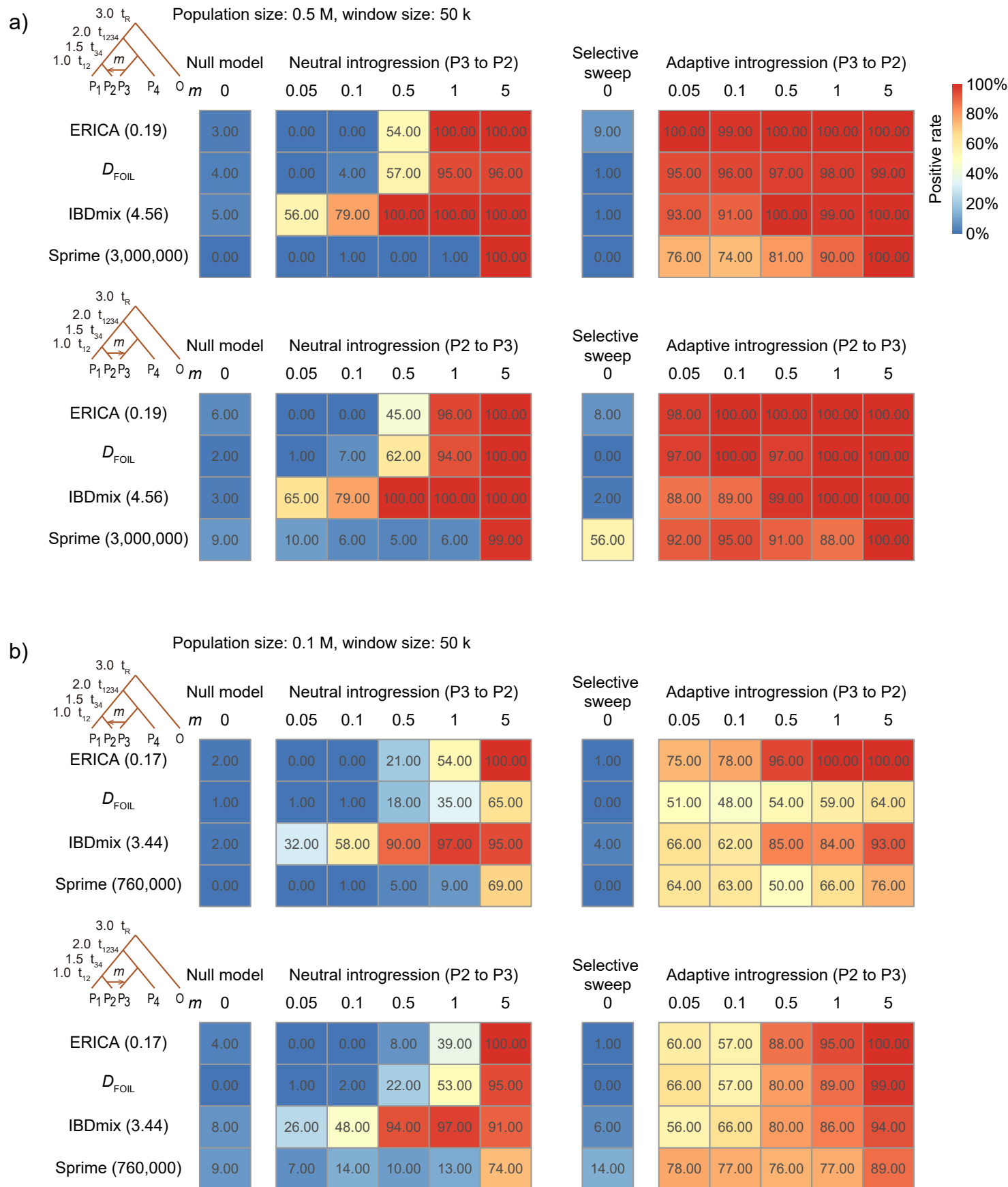
Supplementary Figure S17. Performance of the five-taxon ERICA model and other algorithms for introgression detection with different split times.

The heatmaps show the FPRs for scenarios without gene flow and the TPRs for neutral and adaptive introgression at different migration rates (0.05–5 migrants per generation). The split times t_{12} , t_{123} , and t_R were set to 1, 1.5, 2, and 3 (test dataset D11) (a), 0.5, 0.75, 1, and 1.5 (test dataset D12) (b), and 0.1, 0.15, 0.2, and 0.3 (test dataset D13) (c). The numbers in the brackets are the thresholds used for each method. 100 replicates were tested for each case.



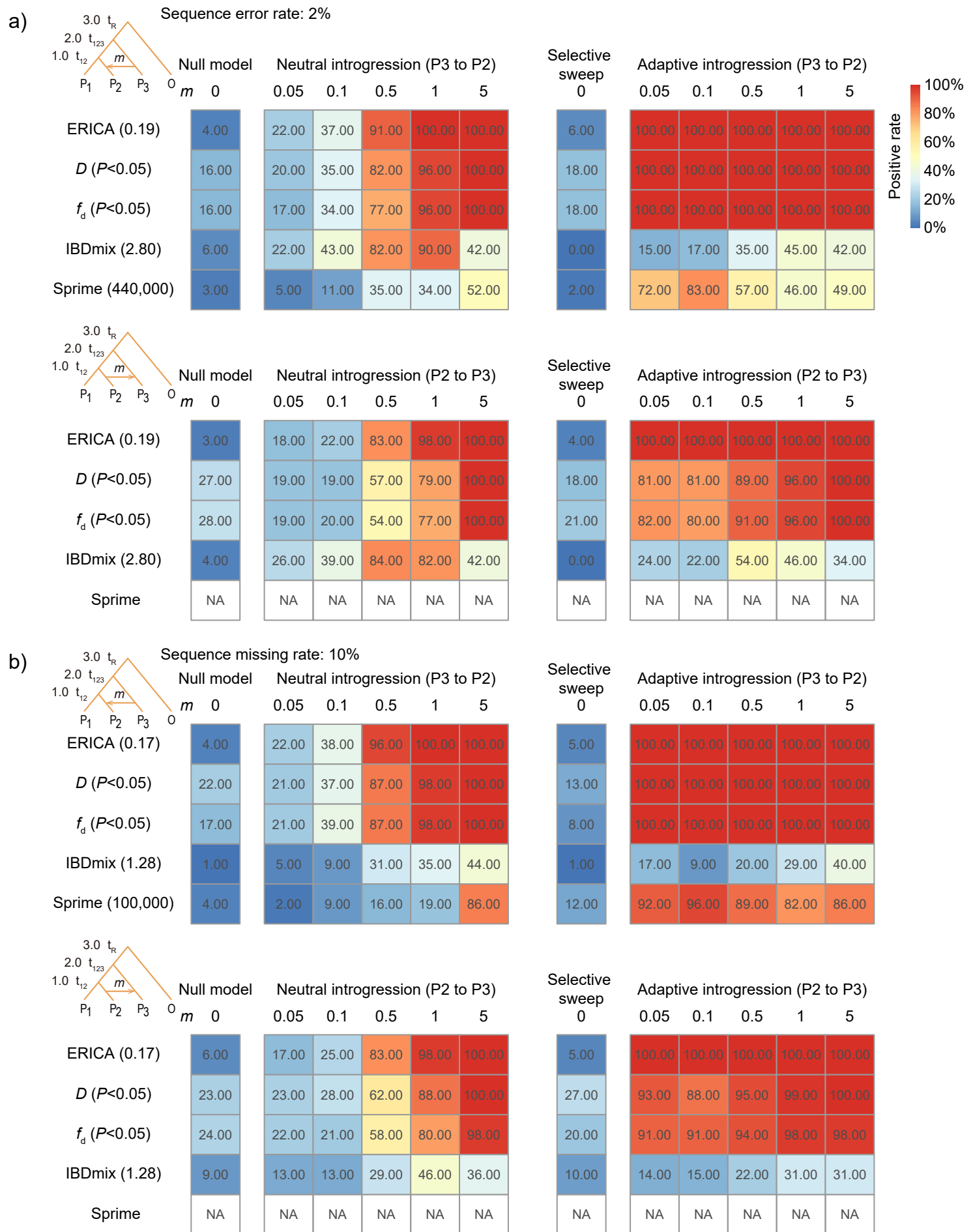
Supplementary Figure S18. Performance of the five-taxon ERICA model and other algorithms for introgression detection for different population sizes.

The heatmaps show the FPRs for scenarios without gene flow and the TPRs for neutral and adaptive introgression at different migration rates (0.05–5 migrants per generation). The population sizes were set to 0.5 M (test dataset D14) (a) and 0.1 M (test dataset D15) (b). The numbers in the brackets are the thresholds used for each method. 100 replicates were tested for each case.



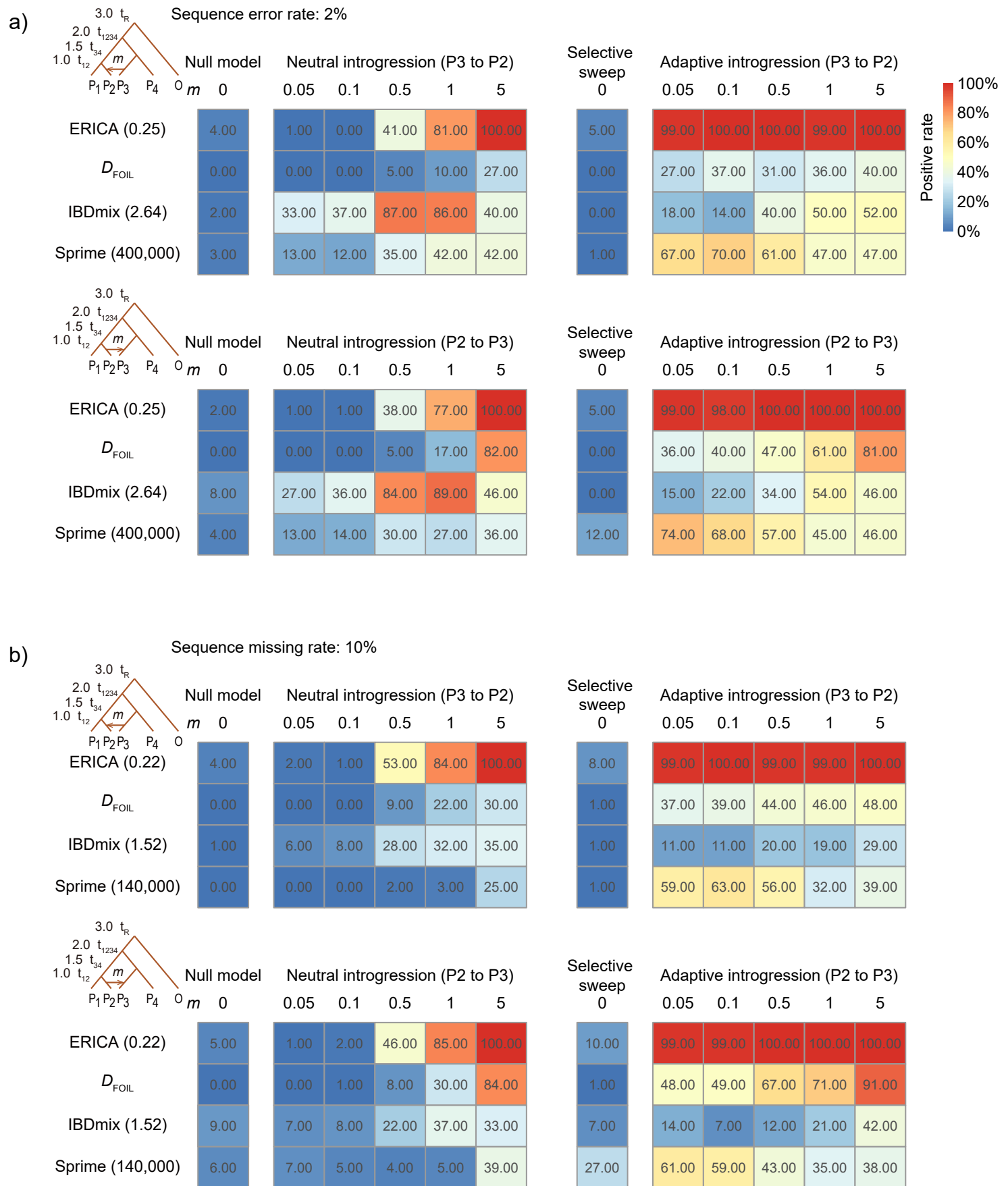
Supplementary Figure S19. Effect of window size on introgression detection for five-taxon data.

The heatmaps show the FPRs for scenarios without gene flow and the TPRs for neutral and adaptive introgression at different migration rates (0.05–5 migrants per generation). The population sizes were set to 0.5 M (test dataset D16) (a) and 0.1 M (test dataset D17) (b). Each window was 50 kb in length. The numbers in the brackets are the thresholds used for each method. 100 replicates were tested for each case.



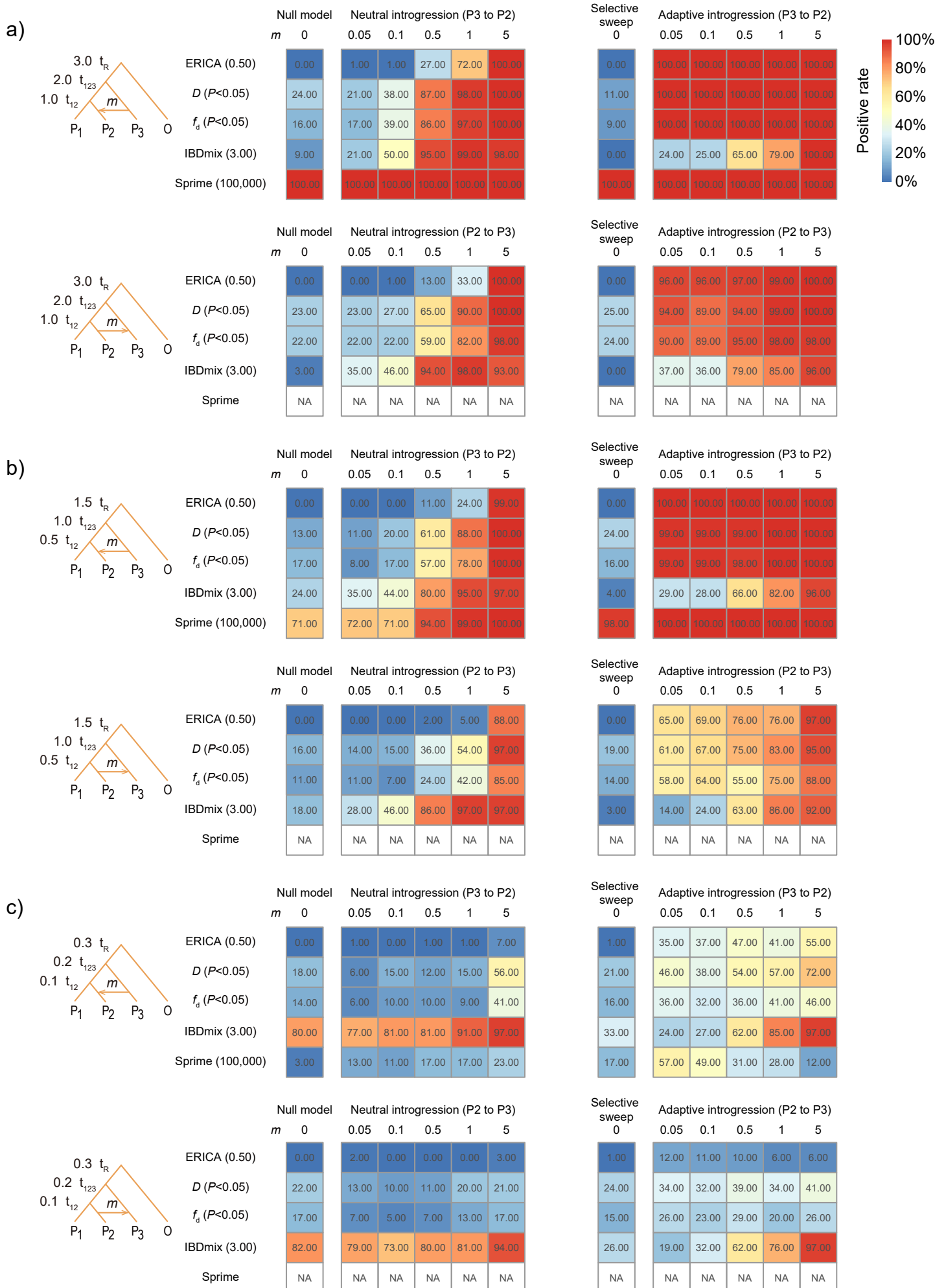
Supplementary Figure S20. Effect of sequence errors on introgression detection for four-taxon data.

The heatmaps show the FPRs for scenarios without gene flow and the TPRs for neutral and adaptive introgression at different migration rates (0.05–5 migrants per generation). The population sizes were set to 1 M and 2% random errors (test dataset D18) (a) and 10% missing data (test dataset D19) (b) were introduced in the sequences. The numbers in the brackets are the thresholds used for each method. NA indicates that Sprime was not applicable for the detection of gene flow from P2 to P3. 100 replicates were tested for each case.



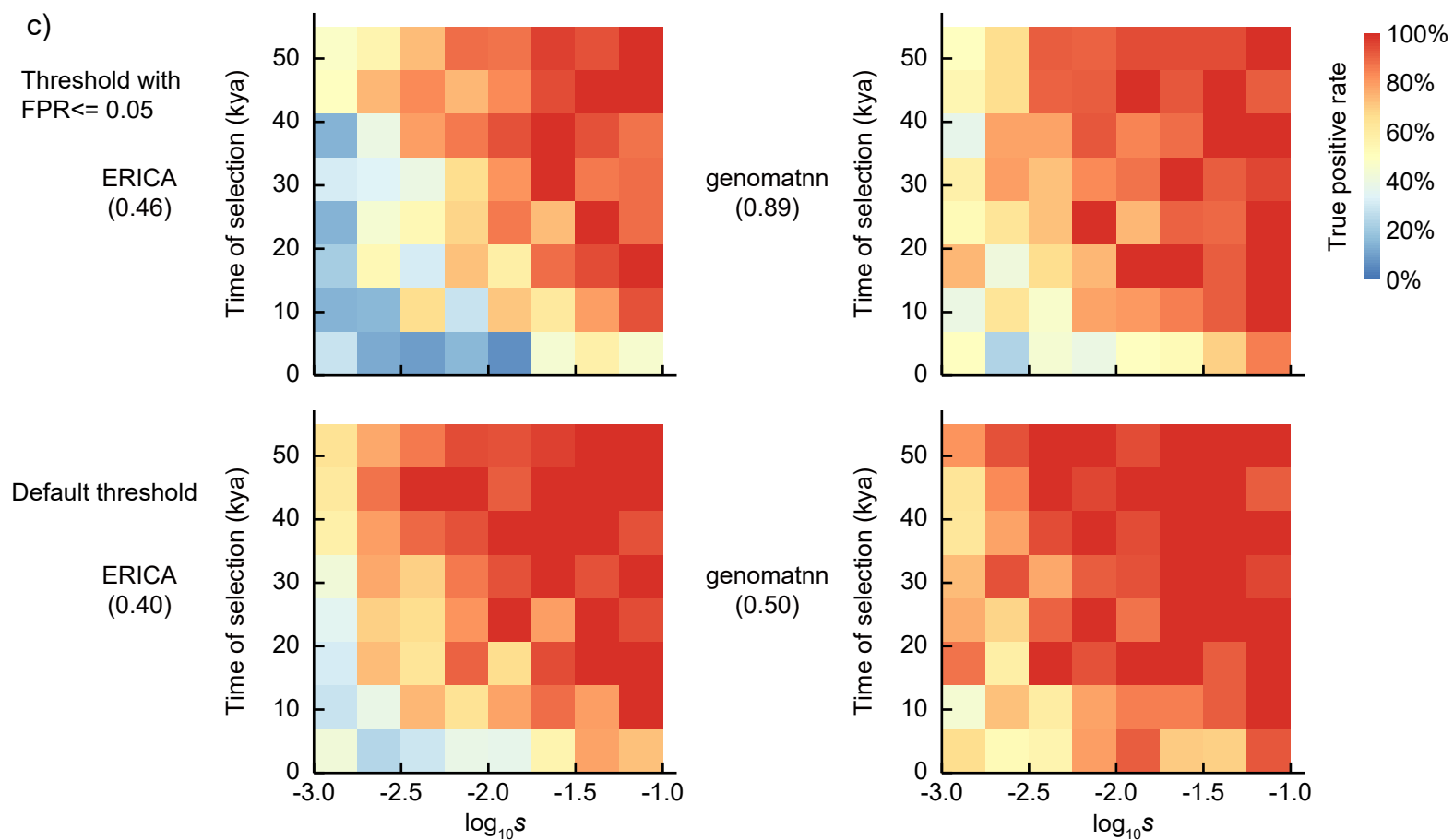
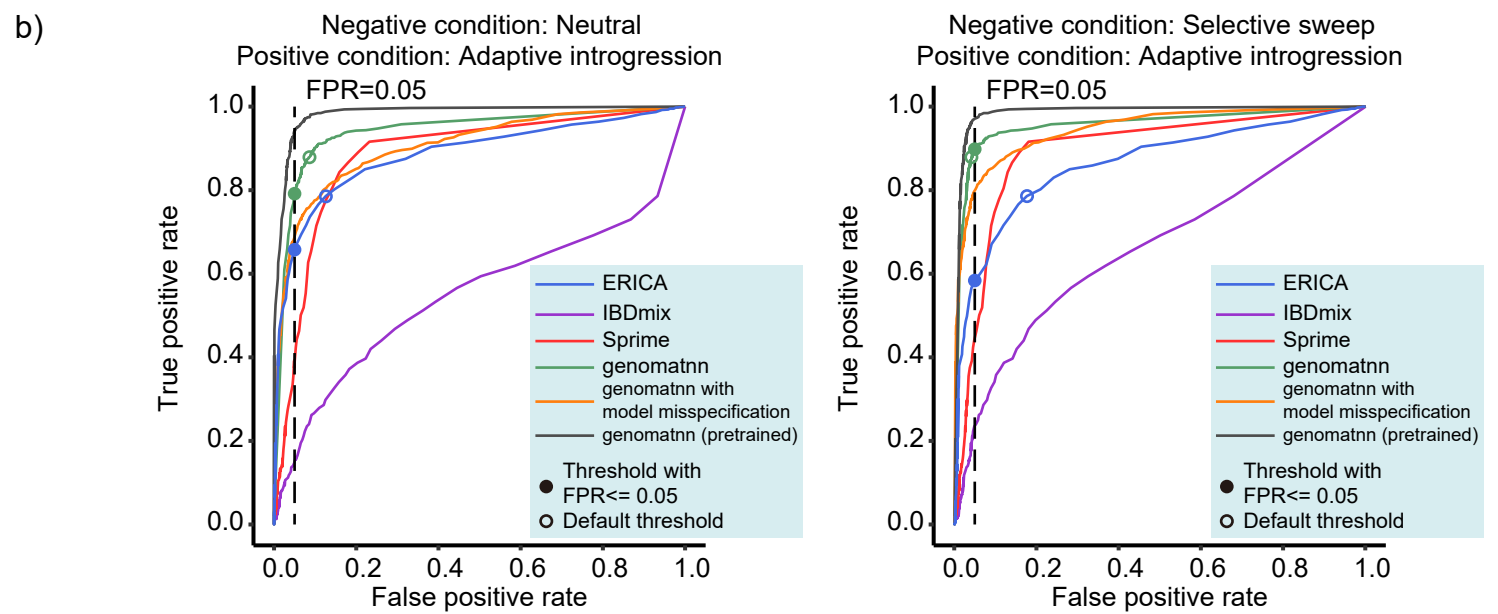
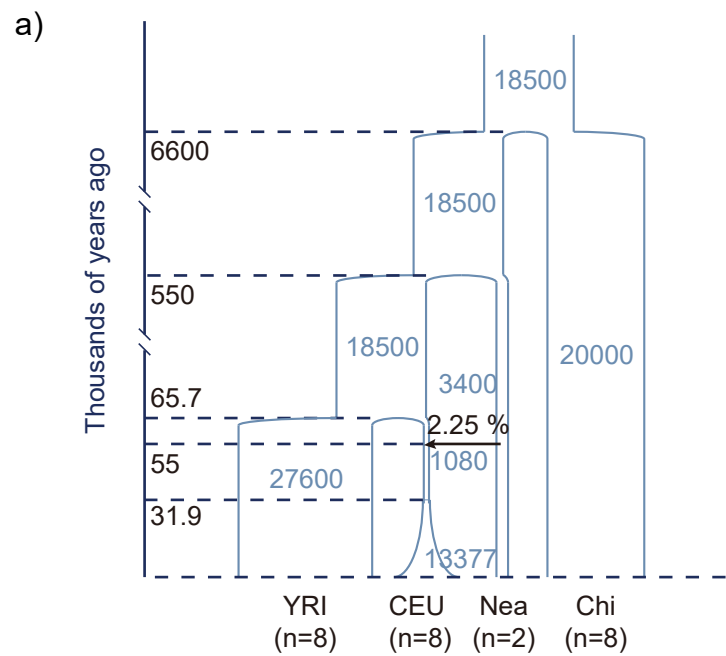
Supplementary Figure S21. Effect of sequence errors on introgression detection for five-taxon data.

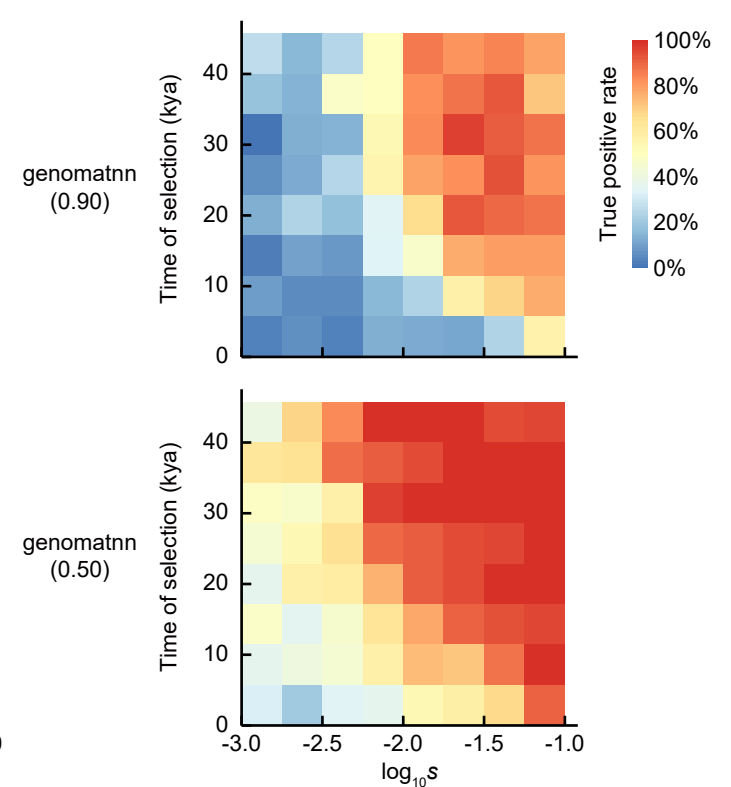
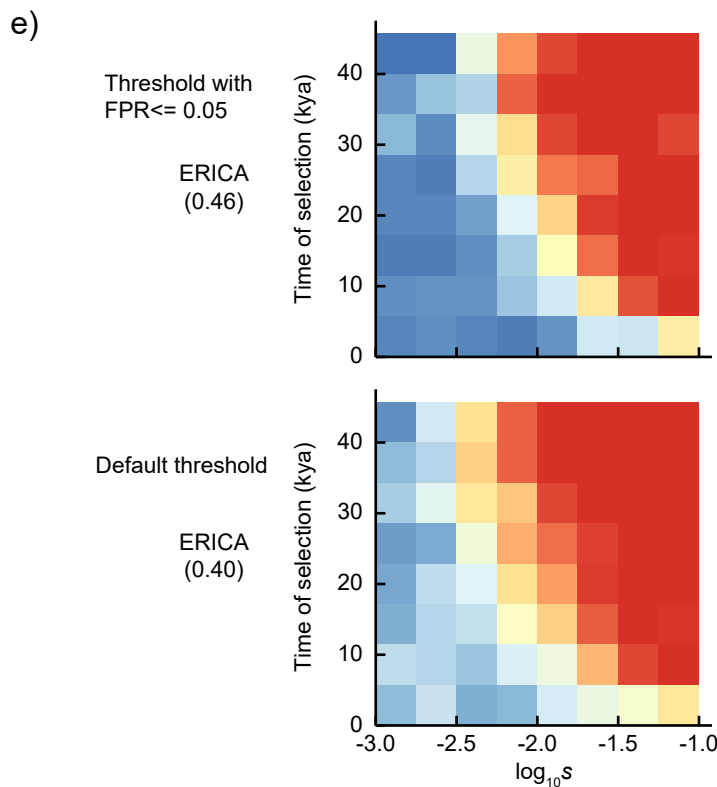
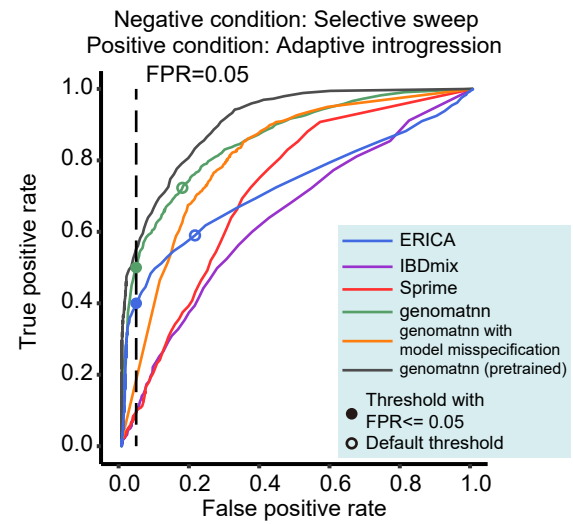
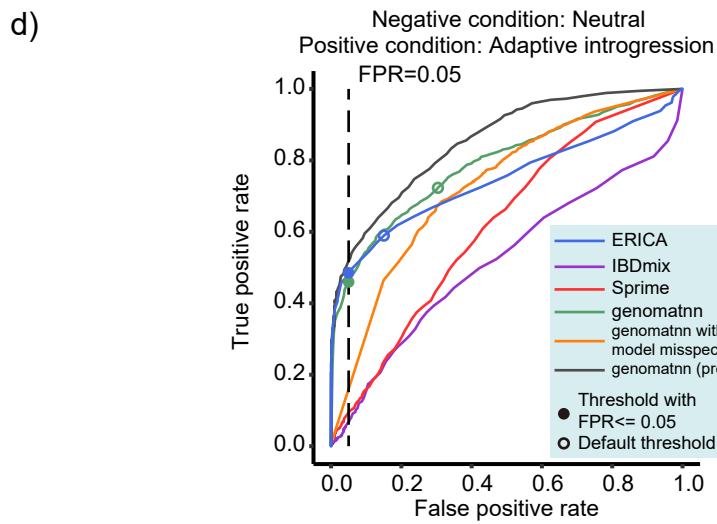
The heatmaps show the FPRs for scenarios without gene flow and the TPRs for neutral and adaptive introgression at different migration rates (0.05–5 migrants per generation). The population sizes were set to 1 M and 2% random errors (test dataset D18) (a) and 10% missing data (test dataset D19) (b) were introduced in the sequences. The numbers in the brackets are the thresholds used for each method. 100 replicates were tested for each case.



Supplementary Figure S22. Effect of thresholds on introgression detection.

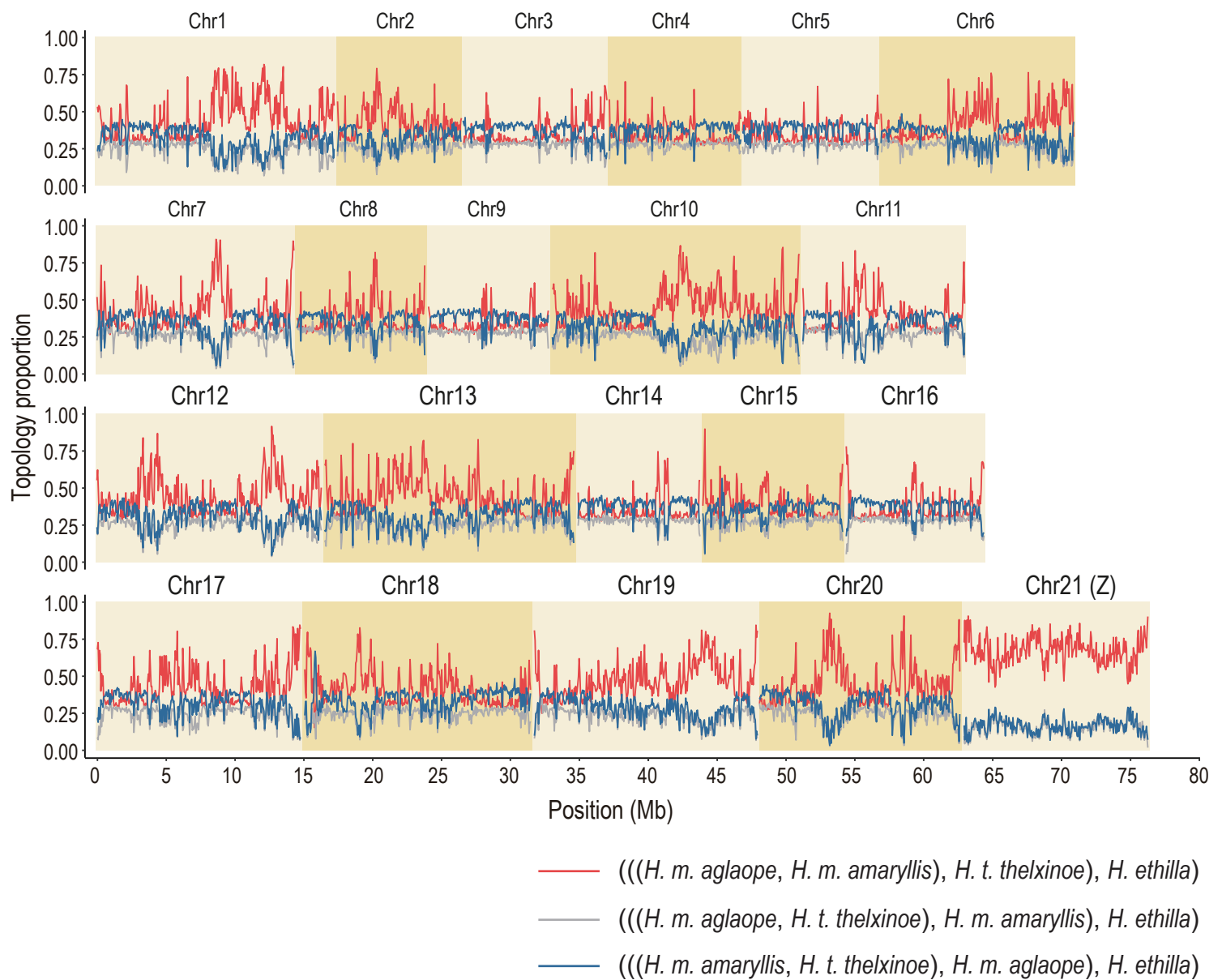
The heatmaps show the FPRs for scenarios without gene flow and the TPRs for neutral and adaptive introgression at different migration rates (0.05–5 migrants per generation) of test datasets D11 (a), D12 (b), and D13 (c) obtained using the software default thresholds. NA indicates that Sprime was not applicable for the detection of gene flow from P2 to P3. 100 replicates were tested for each case.





Supplementary Figure S23. Performance of ERICA and other algorithms for introgression detection in human demographic scenarios.

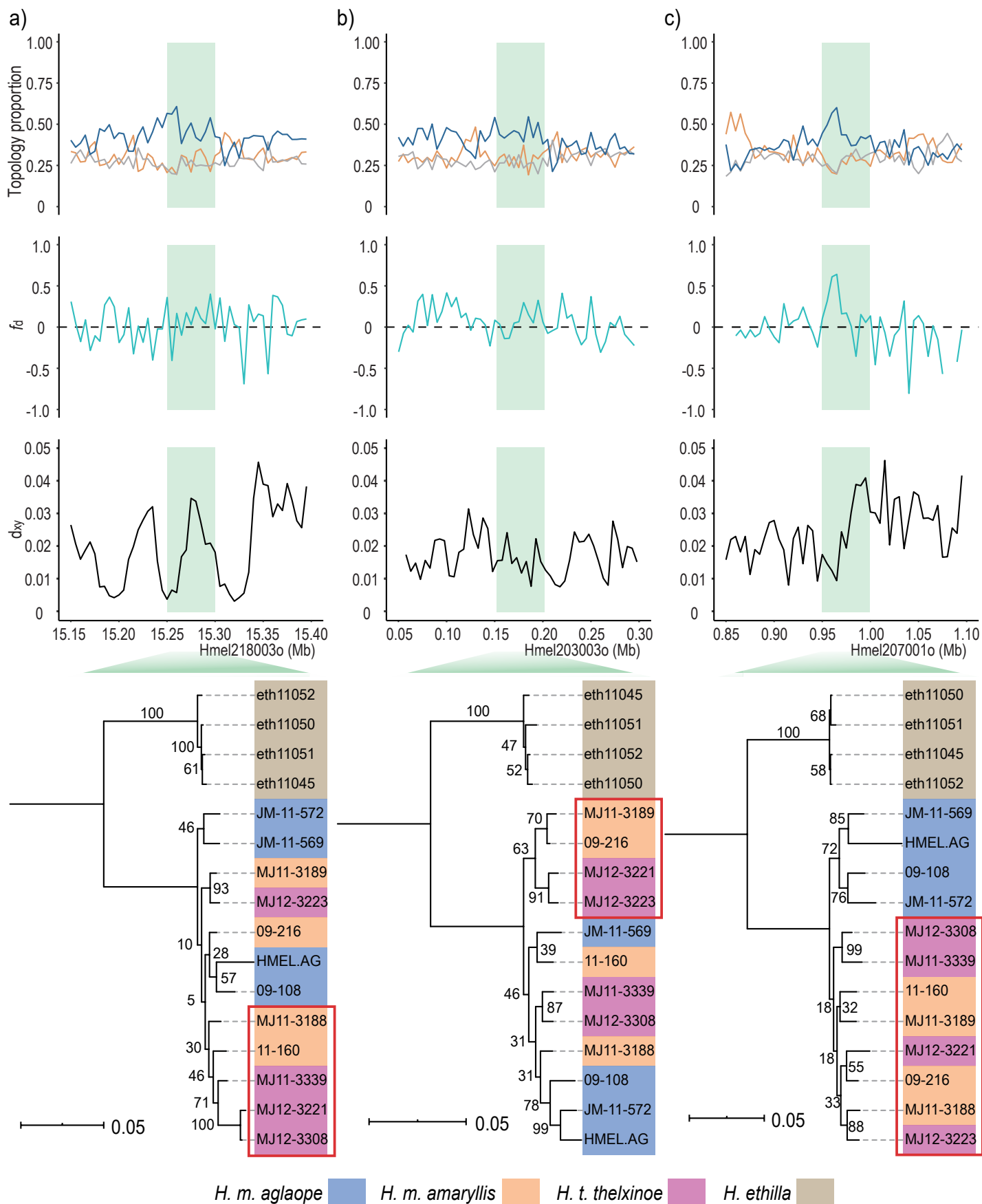
(a) The schematic of human demographic scenario A for ERICA evaluation. The chimpanzee was used as the outgroup. Branch widths are scaled according to the effective population size, and dashed lines indicate divergence times. The arrow indicates gene flow from Neanderthal to Europeans. 8 individuals were sampled for each current population and 2 individuals were sampled for Neanderthals, to be consistent with genomatnn evaluation. YRI, Yoruba in Ibadan, Nigeria; CEU, Northwestern Europeans; Nea, Neanderthal; Chi, chimpanzee. (b) The ROC curves of ERICA, genomatnn, IBDmix, and Sprime for introgression detection in demographic scenario A with neutral and selective sweep as negative categories and adaptive introgression as a positive category. For the genomatnn algorithm, the black lines indicate evaluations of pre-trained CNNs using scenario A from the original study; the green lines indicate evaluations of CNNs trained using the data of scenario A with 8 samples per current population, the orange lines indicate evaluations of CNNs trained using the data of scenario B with the same sample size. The solid circles indicate the TPRs with FPRs equal to 5%, and the hollow circles indicate the TPRs and FPRs obtained using the software default thresholds for classification. Each scenario contained 1,000 simulations. (c) The heatmaps show the sensitivities of ERICA and genomatnn for introgression detection in demographic scenario A at different selection coefficient (s) and starting times of selection. Two thresholds are used: one to make the FPR of neutral datasets $\leq 5\%$ and the other to be the default value of the software. (d) The ROC curves of ERICA, genomatnn, IBDmix, and Sprime for introgression detection in demographic scenario B with neutral and selective sweep as negative categories and adaptive introgression as a positive category. For the genomatnn algorithm, the black lines indicate evaluations of the pre-trained CNNs of scenario B; the green lines indicate evaluations of CNNs trained using the data of scenario B with 8 samples per current population; the orange lines indicate evaluations of CNNs trained using the data of scenario A. (e) The heatmaps show the sensitivities of ERICA and genomatnn for introgression detection in demographic scenario B at different selection coefficient (s) and starting times of selection. Two thresholds are used: one to make the FPR of neutral datasets $\leq 5\%$ and the other to be the default value of the software.



Supplementary Figure S24. Genome-wide evolutionary inference for *Heliconius* butterflies using ERICA.

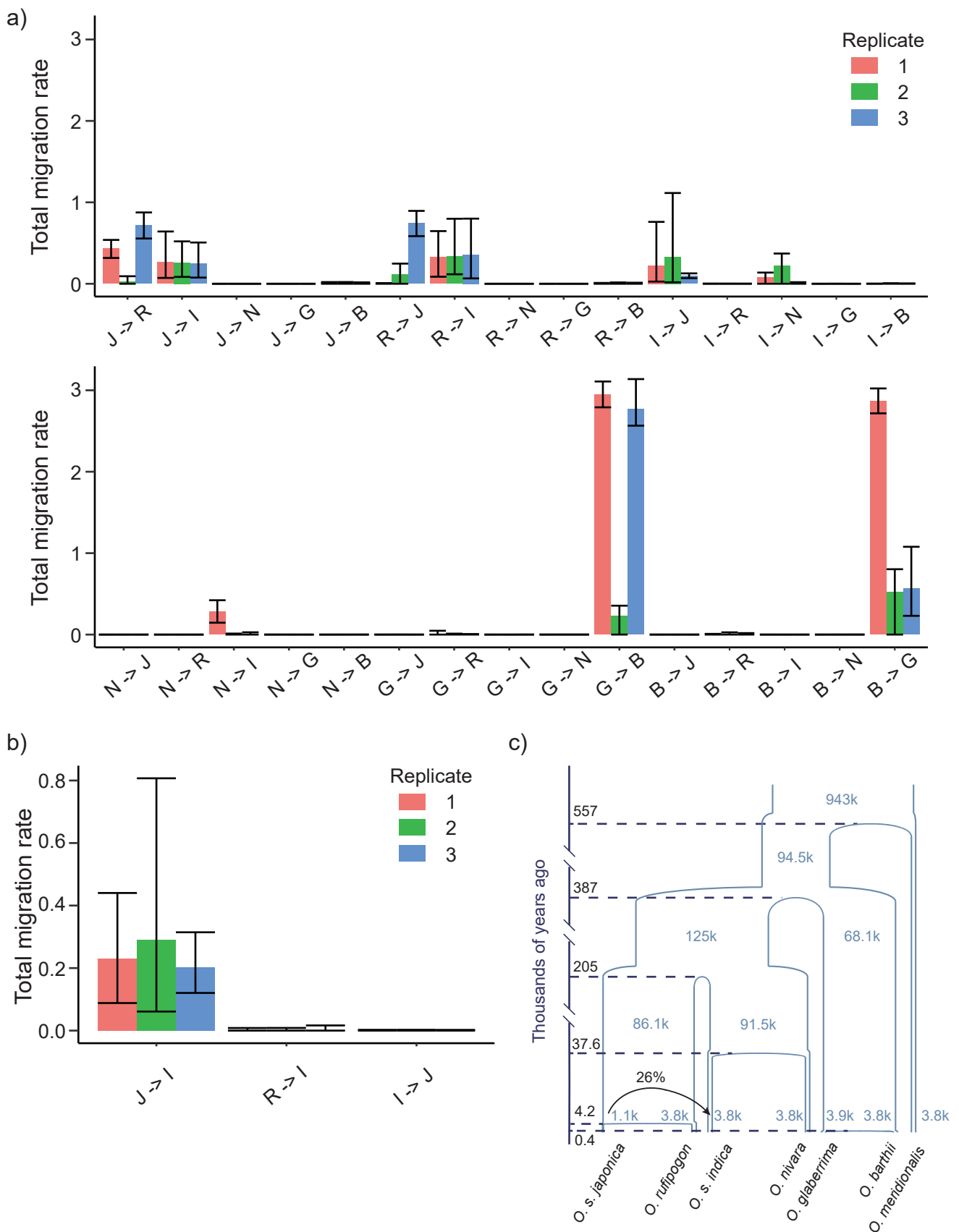
The proportions of three topologies for each 50-kb adjacent window are predicted across the *Heliconius* genome. The ERICA results were visualized using ERICAVisualization.py (<http://erica.cibr.ac.cn/>).

— (((*H. m. amaryllis*, *H. t. thelxinoe*), *H. m. aglaope*), *H. ethilla*)
— (((*H. m. aglaope*, *H. m. amaryllis*), *H. t. thelxinoe*), *H. ethilla*)
— (((*H. m. aglaope*, *H. t. thelxinoe*), *H. m. amaryllis*), *H. ethilla*)



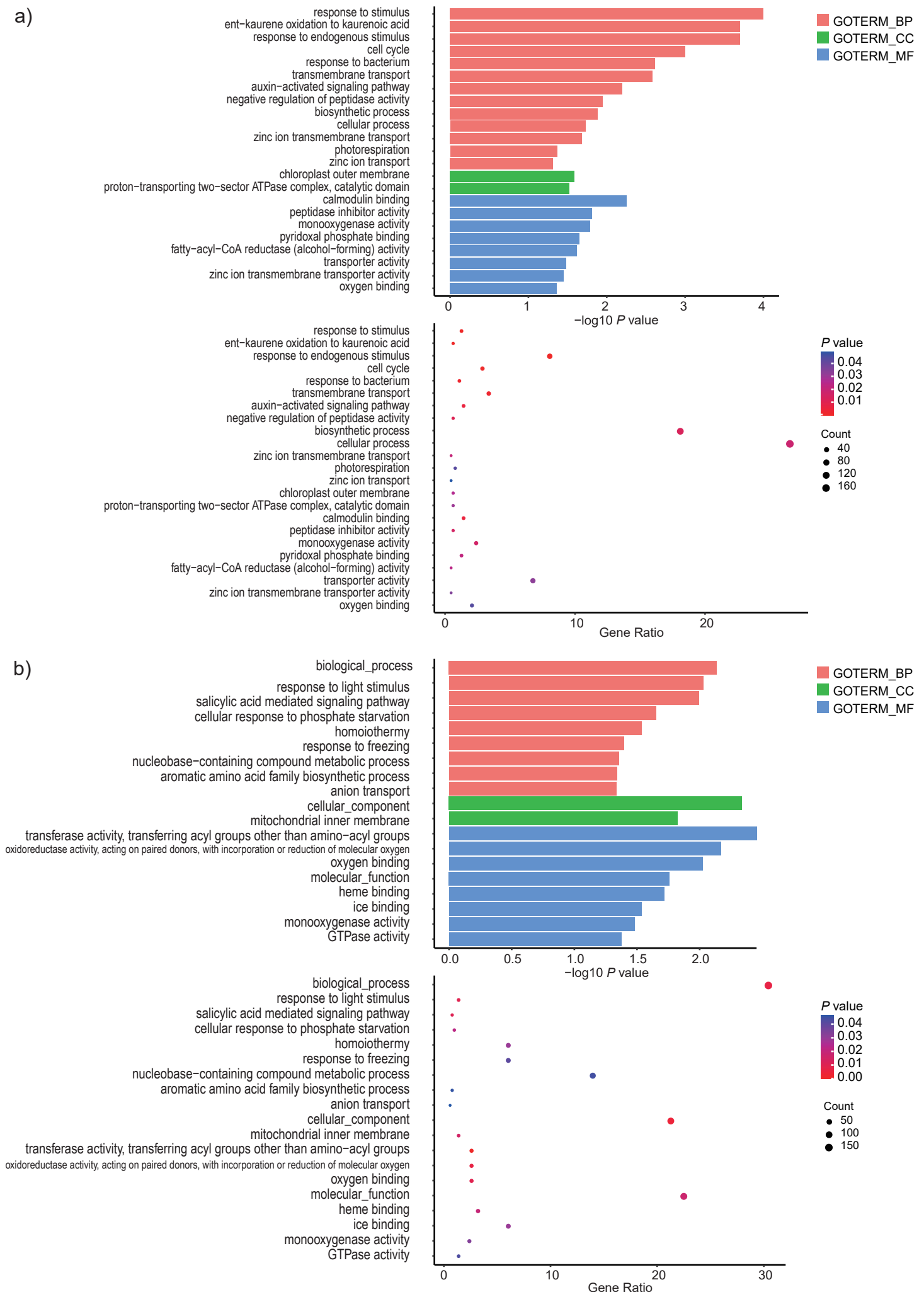
Supplementary Figure S25. Signatures of introgression along three ERICA candidate loci.

The three candidate loci were detected by ERICA, but not by the D and f_d methods. Fine-scale results of ERICA, f_d -statistic, and d_{xy} scans are plotted for each candidate locus. A maximum-likelihood tree was constructed for the highlighted 50-kb window in each locus to infer the direction of gene flow.

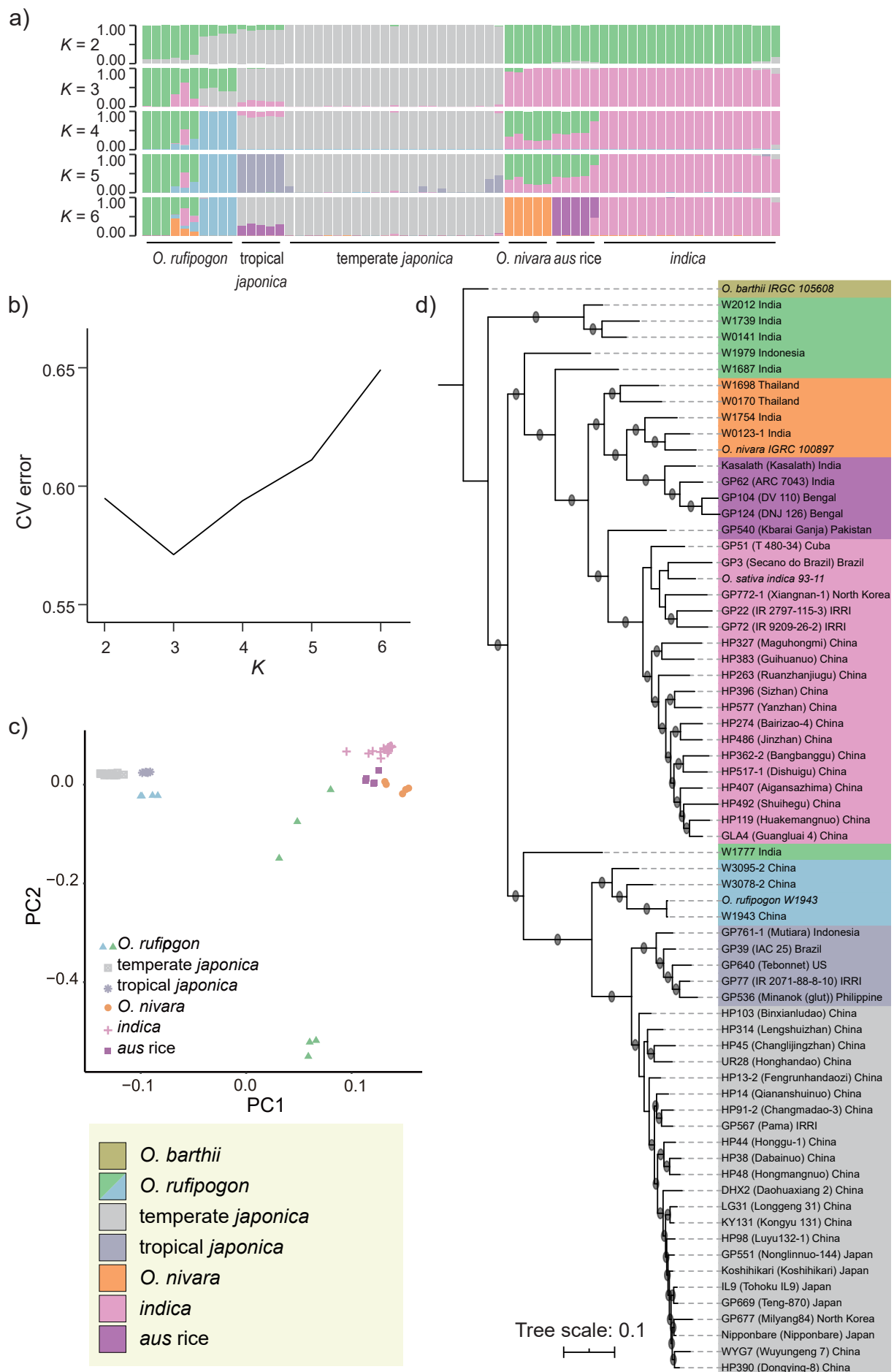


Supplementary Figure S26. Demographic models and migration rates estimated using G-PhoCS.

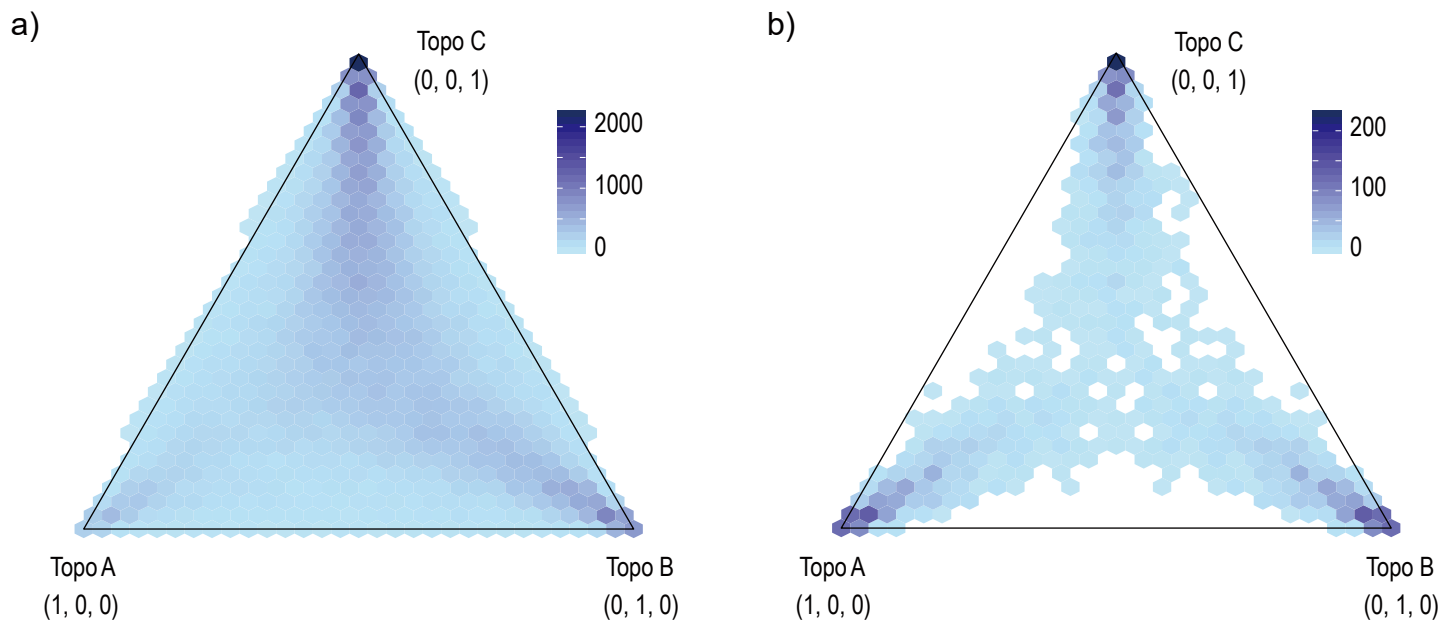
(a) All of the possible migrations between current species were tested independently. The bar plots show the total migration rates with a 95% highest posterior density (HPD). Each potential migration band was tested in three repeated analyses. Significant bidirectional gene flow was detected between japonica and indica, and directional gene flow was detected from *O. rufipogon* to *indica*. The labels are J for *japonica*, R for *O. rufipogon*, I for *indica*, N for *O. nivara*, G for *O. glaberrima* and B for *O. barthii*. (b) A full model analysis containing three detected migration bands was conducted, but the results only suggest significant gene flow from *japonica* to *indica*. J, *japonica*; R, *O. rufipogon*; I, *indica*. (c) The full demographic model was estimated using G-PhoCS. Branch widths are scaled according to the effective population size, and dashed lines indicate divergence times. An arrow is included between *japonica* and *indica* to suggest the direction and strength of gene flow.



Supplementary Figure S27. Gene Ontology (GO) distributions of genes located in introgressed loci. GO enrichment analyses were performed for genes located in introgressed loci between *japonica/indica* (a) and between tropical *japonica* and *O. nivara/indica* (b). Dot colors indicate *P* values and dot sizes indicate the hit counts of GO terms.



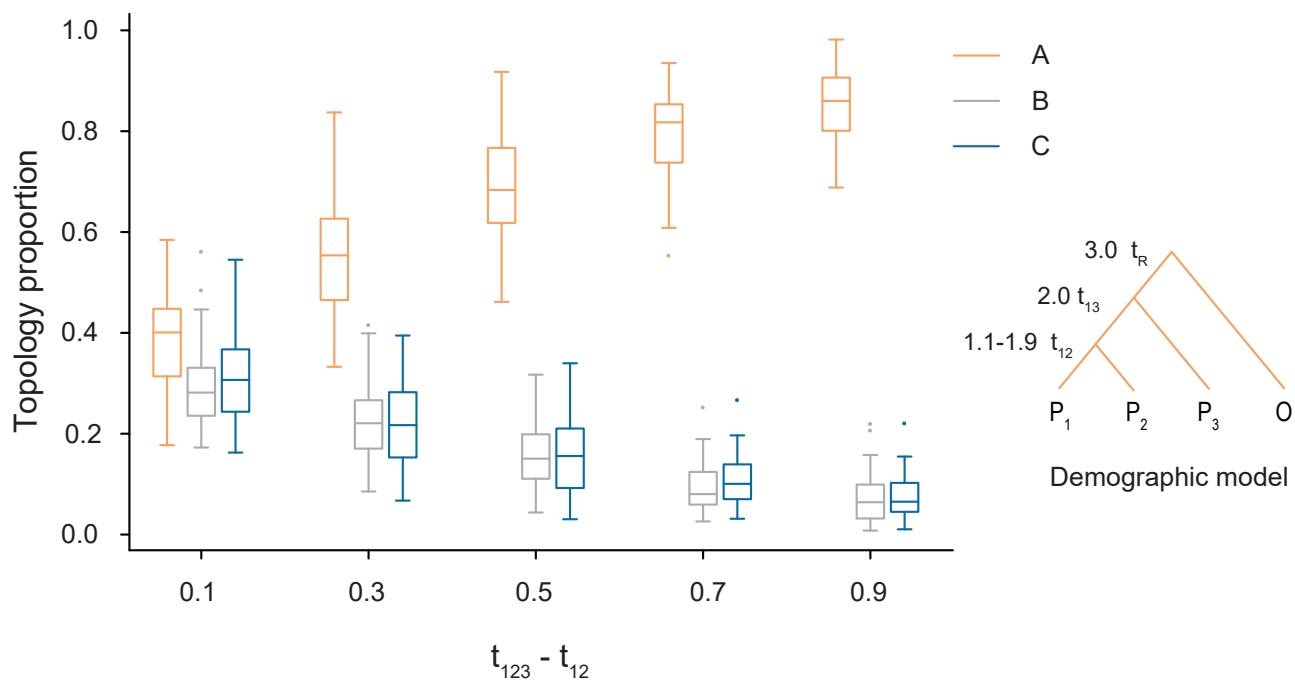
Supplementary Figure S28. Genetic structure and genome-wide phylogeny of the rice population genomic dataset. ADMIXTURE and PCA analyses were applied to 66 rice samples to characterize the population substructure of the samples. Both the population structure analysis (a-b) and principal component analysis (PCA) (c) showed that individual samples from different rice species and accessions form clusters separately, but with gene flow among a subset of them. The color fractions in each column represent the genetic composition of each individual with K clusters (a) and K = 3 as a suggested true genetic population by cross-validation (CV) error (b). The maximum-likelihood phylogenetic tree was constructed based on the genome-wide SNP data, with the scale bar representing the percentage of substitution per site and the circles indicating 100% bootstrap support (d).



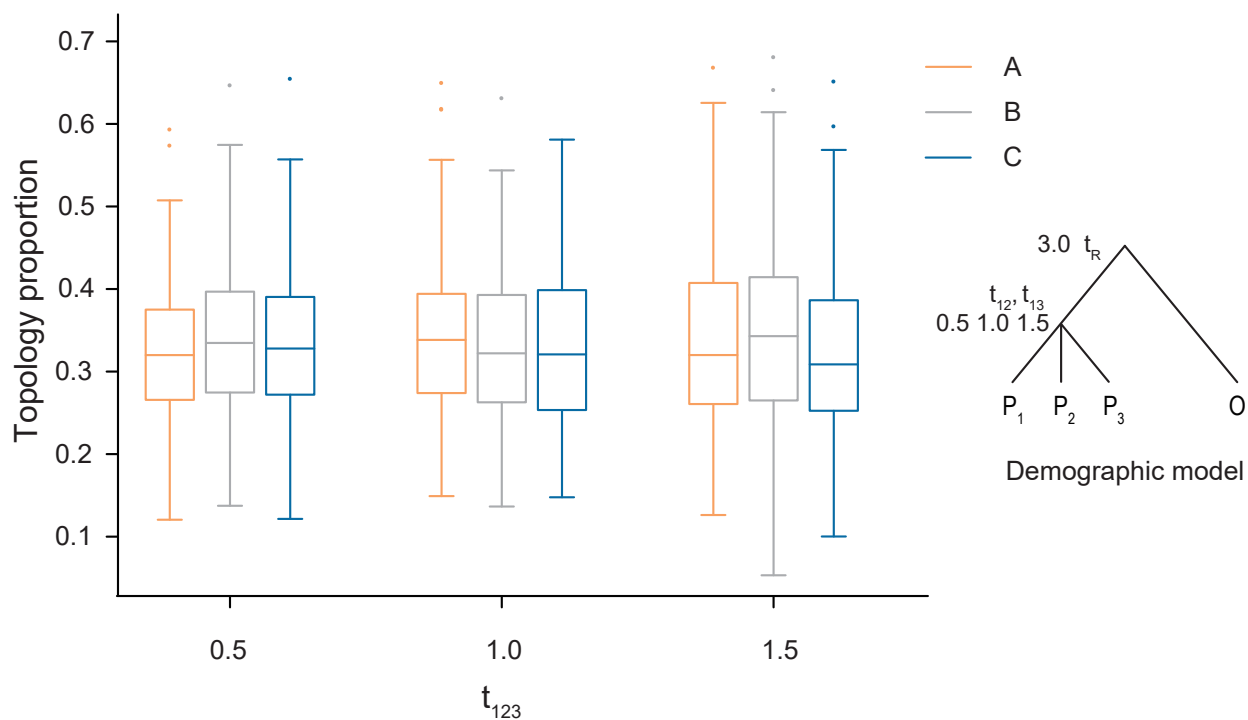
Supplementary Figure S29. The distributions of four-taxon training dataset.

(a) The distributions of the three-dimensional vector labels of the four-taxon training dataset are shown in three-dimensional space and translated into two dimensional coordinates using orthographic projection. The color gradient represents the number of data points with the given labels. (b) Similarly, the distributions of vector labels under the evolutionary scenarios without gene flow are also shown in three-dimensional space.

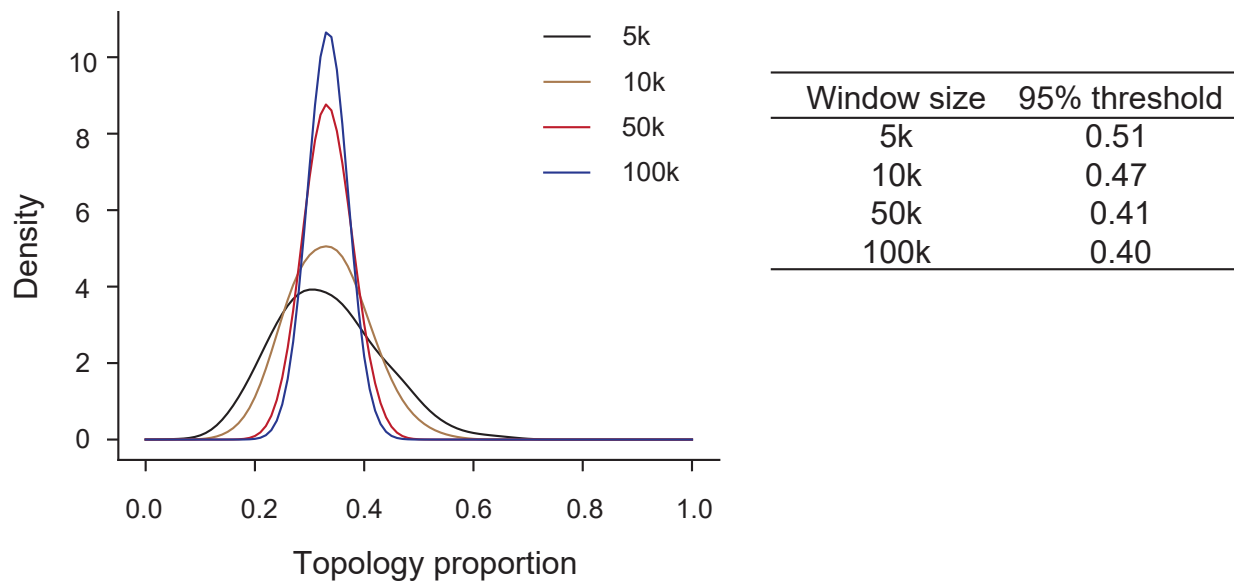
a)



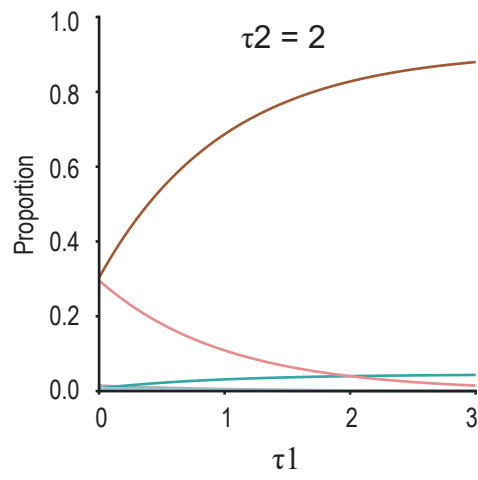
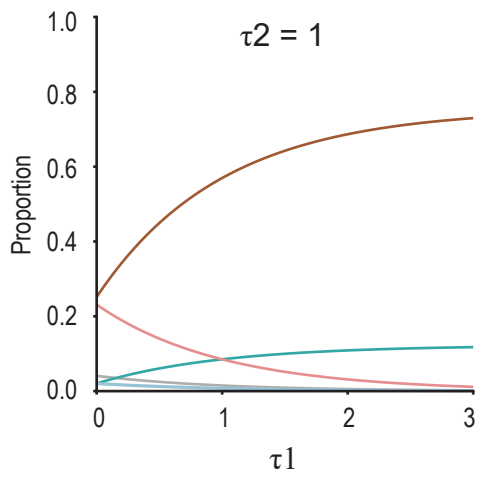
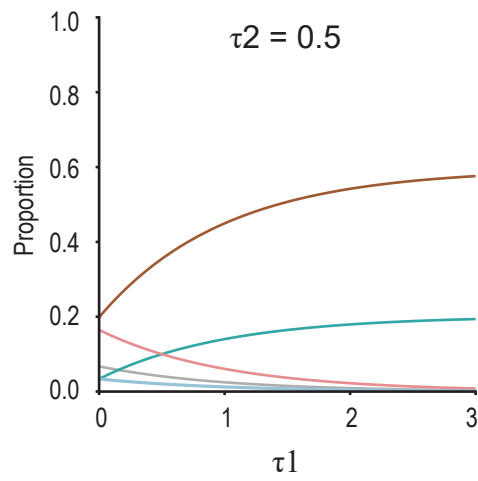
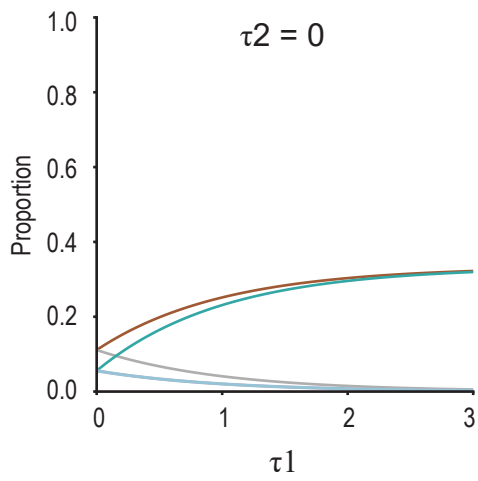
b)



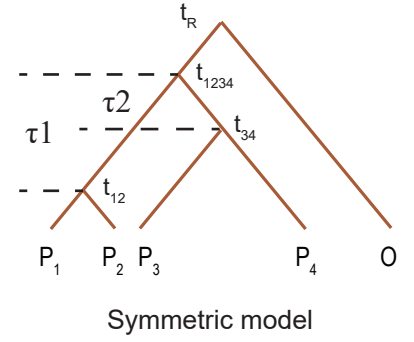
c)



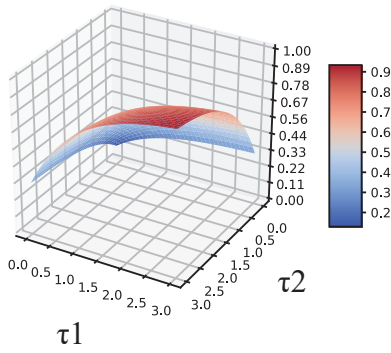
d)



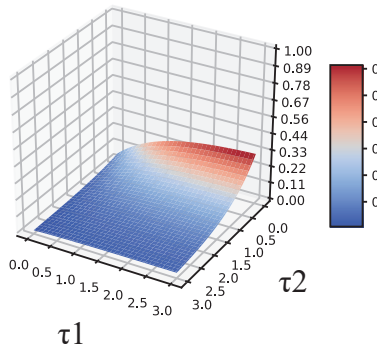
- M
- A, J
- F, I
- N, O
- B, C, D, E, G, H, K, L



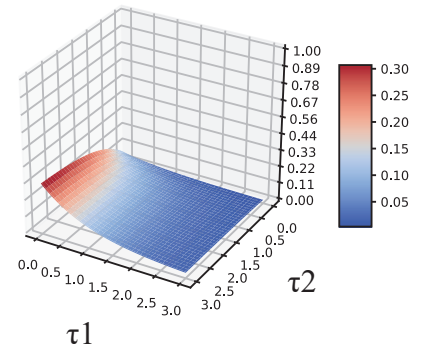
e)



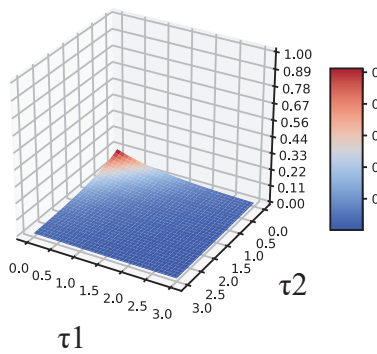
M



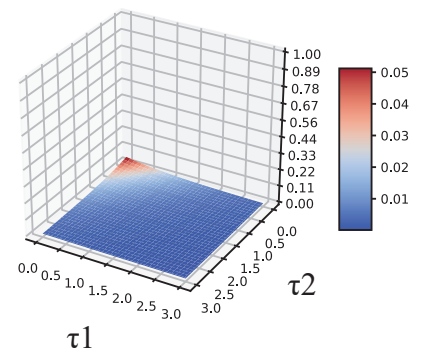
A, J



F, I

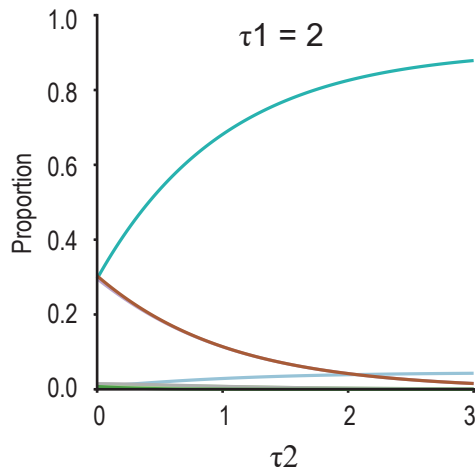
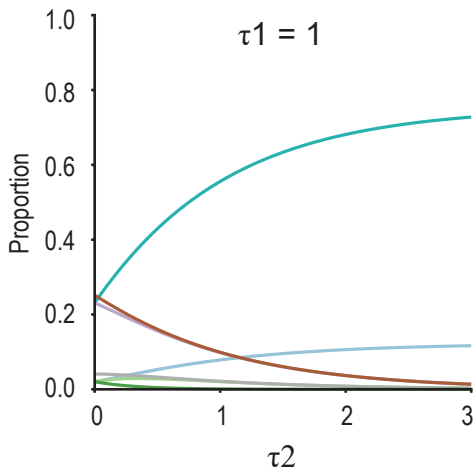
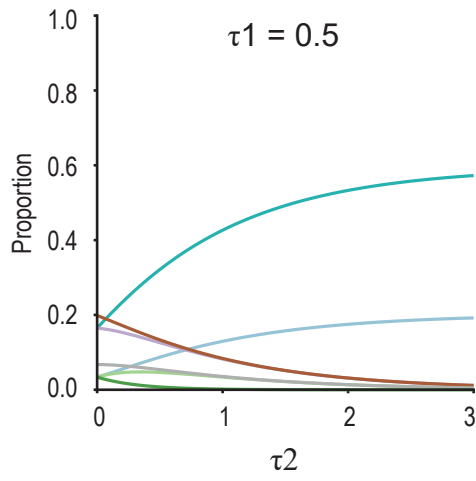
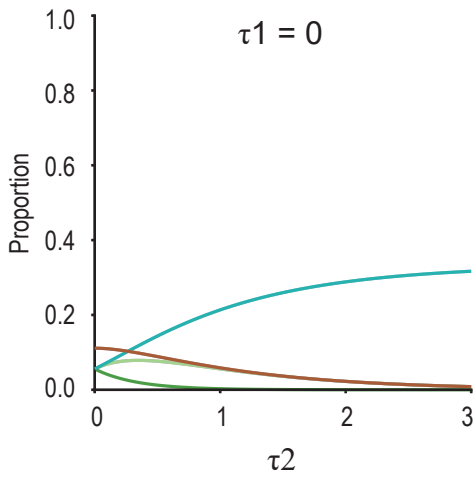


N, O

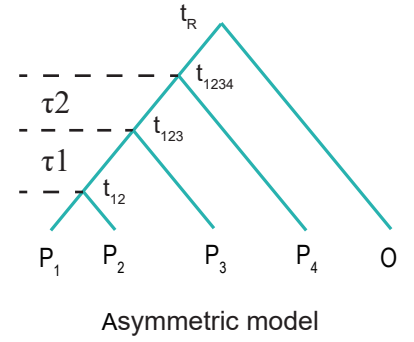


B, C, D, E,
G, H, K, L

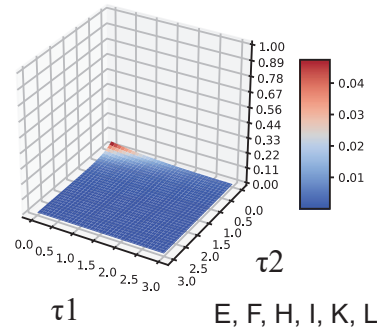
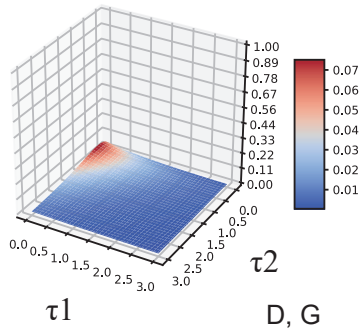
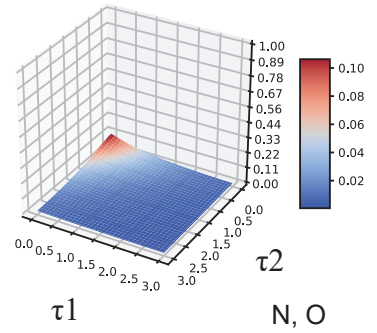
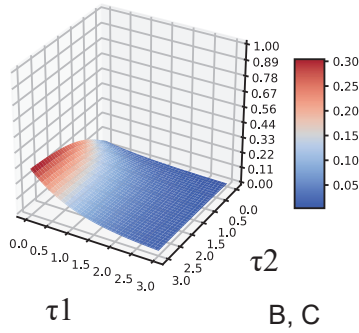
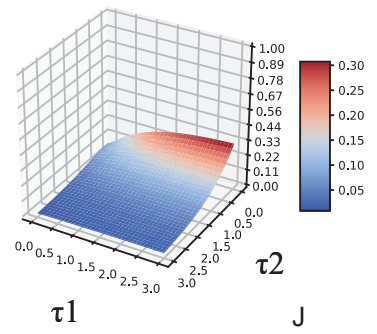
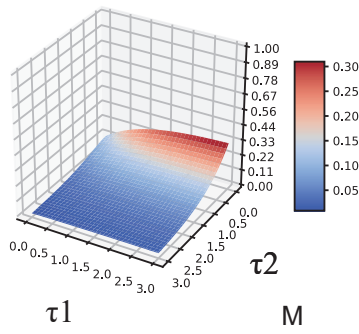
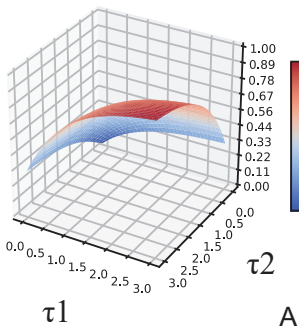
f)



- A
- M
- J
- B, C
- N, O
- D, G
- E, F, H, I, K, L



g)



Supplementary Figure S30. The impact of ILS depends on the divergence time.

(a) The probabilities of discordant topologies increase when the divergence time decreases. The split times of P3 and the outgroup were fixed in the model (2.0 and 3.0 in $4N$ generations, respectively). The divergence time of ancestral populations ranged from 0.1 to 0.9. 50 replicates were simulated for each model. (b) The three topologies have the same probabilities when the three ingroup species split at the same time. Models with different split times were examined, revealing that there was no significant difference between groups (Kruskal-Wallis test P value = 0.3264). 200 replicates were simulated for each model. (c) Kernel density estimation and 95% quantile for different window sizes. All results from panel B were merged and used in the calculation. (d) Theoretical distribution of discordant topologies in the symmetric five-taxon model. There are four groups for probabilities of alternative topologies, and they are determined by the τ_1 and τ_2 of speciation events. Note that we assume the species phylogeny is topology M (((P1, P2), (P3, P4)), O). (e) The landscape of each topology group in the symmetric five taxon model. (f) Theoretical distribution of discordant topologies in the asymmetric five-taxon model and six groups of alternative topologies with species phylogeny A (((((P1, P2), P3), P4), O). (g) The landscape of each topology group in the asymmetric five-taxon model.

Supplementary Table S1. Parameters used in simulation studies of other introgression-detection algorithms.

Method	Population size (N)	Mutation rate		Recombination rate		Migration rate		Window size
		μ (per site per generation)	$4N\mu$	r (per site per generation)	$4Nr$	m (fraction per generation)	$4Nm$	
<i>f_d</i> (Martin et al. <i>Mol. Biol. Evol.</i> 2015)	*		0.01		0.001, 0.01, 0.1	*	*	1 kb, 5 kb, 10 kb
<i>D_{FOIL}</i> (Pease et al. <i>Syst. Biol.</i> 2015)	1 M	7×10^{-9}	0.028	1×10^{-8} , 1×10^{-7}	0.04, 0.4		0.1, 1, 5, 50, 500	100 kb
Twisst (Martin et al. <i>Genetics</i> 2017)	100 k	2.5×10^{-8}	0.01		0.001, 0.01		0.1, 5	1 Mb
FILET (Schrider et al. <i>PLoS Genet.</i> 2018)	26 k to 9.3 M	3.5×10^{-9}			0.2		0.013, 0.086	10 kb
IBDmix (Chen et al. <i>Cell</i> 2020)	7 k	1.25×10^{-8}		1×10^{-8}		2×10^{-4}		*
Sprime (Browning et al. <i>Cell</i> 2018)	15 k	1.2×10^{-8}		1×10^{-8}		1×10^{-5}		50 kb
genomatnn (Gower et al. <i>Elife</i> 2021)	1 k to 48 k	1.29×10^{-8}				0.02 (one generation)		100 kb

Supplementary Table S2. Evolutionary scenarios of the training and test datasets in the four-taxon and five-taxon models.

Evolutionary scenarios of the training and testing datasets in the four-taxon model								
part 1 TopoB (((P1,P3),P2),O)								
ms -ej t13 3 1 -ej t123 2 1 -ej 3 4 1								
t13	t123	combinations	replicates per combination (training dataset)	replicates per combination (test dataset)				
range: [0.2, 1.8], step size: 0.2	range: [t13 + 0.2, 2], step size: 0.2	45	120	6				
part 2 TopoA (((P1,P2),P3),O) with introgression from P3 to P2								
ms -ej t12 2 1 -ej t123 3 1 -ej 3 4 1 -es t_intro 2 p -ej t_intro 5 3								
t12	t123	t_intro	probability (p)	combinations	replicates per combination (training dataset)	replicates per combination (test dataset)		
range: [0.4, 2.0], step size: 0.4	range: [t12 + 0.1, 2.1], step size: 0.6	range: [0.1, t12], step size: 0.4	range: [0, 0.9], step size: 0.1	240	120	6		
part 3 TopoA (((P1,P2),P3),O) with introgression from P2 to P3								
ms -ej t12 2 1 -ej t123 3 1 -ej 3 4 1 -es t_intro 3 p -ej t_intro 5 2								
t12	t123	t_intro	probability (p)	combinations	replicates per combination (training dataset)	replicates per combination (test dataset)		
range: [0.4, 2.0], step size: 0.4	range: [t12 + 0.1, 2.1], step size: 0.6	range: [0.1, t12], step size: 0.4	range: [0, 0.9], step size: 0.1	240	120	6		
part 4 TopoA (((P1,P2),P3),O) with introgression from P3 to P1 and introgression from P3 to P2								
ms -ej t12 2 1 -ej t123 3 1 -ej 3 4 1 -es t_intro_23 2 p_23 -ej t_intro_23 5 3 -es t_intro_13 1 p_13 -ej t_intro_13 6 3								
t12	t123	t_intro_23	p_23	t_intro_13	p_13	combinations	replicates per combination (training dataset)	replicates per combination (test dataset)
range: [0.4, 2.0], step size: 0.4	range: [t12 + 0.1, 2.1], step size: 0.6	range: [0.1, t12], step size: 0.4	range: [0, 0.9], step size: 0.1	random [0.1, t12-0.2]	random [0.0, 1.0]	240	120	6
part 5 TopoA (((P1,P2),P3),O) with introgression from P1 to P3 and introgression from P2 to P3								
ms -ej t12 2 1 -ej t123 3 1 -ej 3 4 1 -es t_intro_23 3 p_23 -ej t_intro_23 5 2 -es t_intro_13 3 p_13 -ej t_intro_13 6 1								
t12	t123	t_intro_23	p_23	t_intro_13	p_13	combinations	replicates per combination (training dataset)	replicates per combination (test dataset)
range: [0.4, 2.0], step size: 0.4	range: [t12 + 0.1, 2.1], step size: 0.6	range: [0.1, t12], step size: 0.4	range: [0, 0.9], step size: 0.1	random [0.1, t12-0.2]	random [0.0, 1.0]	240	120	6
Evolutionary scenarios of the training and testing datasets in the five-taxon model								
part 1 TopoA (((((P1,P2),P3),P4),O)								
ms -ej t12 2 1 -ej t123 3 1 -ej t1234 4 1 -ej 3 5 1								
t12	t123	t1234	combinations	replicates per combination (training dataset)	replicates per combination (test dataset)			
range: [0.2, 1.7], step size: 0.3	range: [t12 + 0.2, 2.2], step size: 0.3	range: [t123 + 0.2, 2.7], step size: 0.3	110	120	12			
part 2 TopoA (((((P1,P2),P3),P4),O) with one introgression								
direction of introgression	t12	t123	t1234	t_intro	p	combinations	replicates per combination (training dataset)	replicates per combination (test dataset)
one of [P1 -> P3, P3 -> P1, P2 -> P3, P3 -> P2, P1 -> P4, P4 -> P1, P2 -> P4, P4 -> P2]	random [0.4, 1.5]	random [t12+0.2, 2.1]	random [t123+0.2, 2.7]	random [0.1, t12-0.1]	range: [0, 0.9], step size: 0.1	80	120	12
one of [P12 -> P4, P4 -> P12]	random [0.4, 1.5]	random [t12+0.2, 2.1]	random [t123+0.2, 2.7]	random [t12+0.1, t123-0.1]	range: [0, 0.9], step size: 0.1	20	120	12
one of [P3 -> P4, P4 -> P3]	random [0.4, 1.5]	random [t12+0.2, 2.1]	random [t123+0.2, 2.7]	random [0.1, t123-0.1]	range: [0, 0.9], step size: 0.1	20	120	12

part 3 TopoA (((P1,P2),P3),P4),O) with two introgression

direction of introgression 1	direction of introgression 2	t12	t123	t1234	t_intro_1	p_1	t_intro_2	p_2	combinations	replicates per combination (training dataset)	replicates per combination (test dataset)
one of [P1 -> P3, P3 -> P1, P2 -> P3, P3 -> P2, P1 -> P4, P4 -> P1, P2 -> P4, P4 -> P2]	one of [P1 -> P3, P3 -> P1, P2 -> P3, P3 -> P2, P1 -> P4, P4 -> P1, P2 -> P4, P4 -> P2]	random [1.0, 1.5]	random [t12+0.3, 2.2]	random [t123+0.2, 2.7]	random [0.5, t12-0.1]	random [0.0, 1.0]	random [0.1, min(t_intro_1, t12)-0.1]	random [0.0, 1.0]	64	240	24
one of [P1 -> P3, P3 -> P1, P2 -> P3, P3 -> P2, P1 -> P4, P4 -> P1, P2 -> P4, P4 -> P2]	one of [P12 -> P4, P4 -> P12]	random [1.0, 1.5]	random [t12+0.3, 2.2]	random [t123+0.2, 2.7]	random [0.5, t12-0.1]	random [0.0, 1.0]	random [max(t_intro_1, t12)+0.1, t123 - 0.1]	random [0.0, 1.0]	16	240	24
one of [P1 -> P3, P3 -> P1, P2 -> P3, P3 -> P2, P1 -> P4, P4 -> P1, P2 -> P4, P4 -> P2]	one of [P3 -> P4, P4 -> P3]	random [1.0, 1.5]	random [t12+0.3, 2.2]	random [t123+0.2, 2.7]	random [0.5, t12-0.1]	random [0.0, 1.0]	random [0.1, min(t_intro_1, t123)-0.1]	random [0.0, 1.0]	16	240	24
one of [P12 -> P4, P4 -> P12]	one of [P1 -> P3, P3 -> P1, P2 -> P3, P3 -> P2, P1 -> P4, P4 -> P1, P2 -> P4, P4 -> P2]	random [1.0, 1.5]	random [t12+0.3, 2.2]	random [t123+0.2, 2.7]	random [t12+0.1, t123-0.2]	random [0.0, 1.0]	random [0.1, min(t_intro_1, t12)-0.1]	random [0.0, 1.0]	16	240	24
one of [P12 -> P4, P4 -> P12]	one of [P12 -> P4, P4 -> P12]	random [1.0, 1.5]	random [t12+0.3, 2.2]	random [t123+0.2, 2.7]	random [t12+0.1, t123-0.2]	random [0.0, 1.0]	random [max(t_intro_1, t12)+0.1, t123 - 0.1]	random [0.0, 1.0]	4	240	24
one of [P12 -> P4, P4 -> P12]	one of [P3 -> P4, P4 -> P3]	random [1.0, 1.5]	random [t12+0.3, 2.2]	random [t123+0.2, 2.7]	random [t12+0.1, t123-0.2]	random [0.0, 1.0]	random [0.1, min(t_intro_1, t123)-0.1]	random [0.0, 1.0]	4	240	24
one of [P3 -> P4, P4 -> P3]	one of [P1 -> P3, P3 -> P1, P2 -> P3, P3 -> P2, P1 -> P4, P4 -> P1, P2 -> P4, P4 -> P2]	random [1.0, 1.5]	random [t12+0.3, 2.2]	random [t123+0.2, 2.7]	random [0.5, t123-0.2]	random [0.0, 1.0]	random [0.1, min(t_intro_1, t12)-0.1]	random [0.0, 1.0]	16	240	24
one of [P3 -> P4, P4 -> P3]	one of [P12 -> P4, P4 -> P12]	random [1.0, 1.5]	random [t12+0.3, 2.2]	random [t123+0.2, 2.7]	random [0.5, t123-0.2]	random [0.0, 1.0]	random [max(t_intro_1, t12)+0.1, t123 - 0.1]	random [0.0, 1.0]	4	240	24
one of [P3 -> P4, P4 -> P3]	one of [P3 -> P4, P4 -> P3]	random [1.0, 1.5]	random [t12+0.3, 2.2]	random [t123+0.2, 2.7]	random [0.5, t123-0.2]	random [0.0, 1.0]	random [0.1, min(t_intro_1, t123)-0.1]	random [0.0, 1.0]	4	240	24

part 4 TopoM (((P1,P2),(P3,P4)),O)

ms -ej t12 2 1 -ej t34 4 3 -ej t1234 3 1 -ej 3 5 1

t12	t34	t1234	combinations	replicates per combination (training dataset)	replicates per combination (test dataset)
range: [0.2, 1.7], step size: 0.3	range: [0.2, 1.7], step size: 0.3	range: [max(t12, t34)+0.2, 2.7], step size: 0.5	110	30	3

part 5 TopoM (((P1,P2),(P3,P4)),O) with one introgression

direction of introgression	t12	t34	t1234	t_intro	p	combinations	repeats number (training dataset)	repeats number (test dataset)
one of [P1 -> P3, P3 -> P1, P2 -> P3, P3 -> P2, P1 -> P4, P4 -> P1, P2 -> P4, P4 -> P2]	random [0.5, 1.7]	random [0.5, 1.7]	random [max(t12, t34)+0.2, 2.5]	random [0.1, min(t12, t34)-0.1]	range: [0, 0.9], step size: 0.1	80	30	3
one of [P12 -> P4, P4 -> P12, P12 -> P3, P3 -> P12]	random [0.5, 1.7]	random [t12+0.2, 2.1]	random [max(t12, t34)+0.2, 2.5]	random [t12+0.1, t34-0.1]	range: [0, 0.9], step size: 0.1	40	30	3
one of [P34 -> P1, P1 -> P34, P34 -> P2, P2 -> P34]	random [0.5, 1.7]	random [0.2, t12-0.2]	random [max(t12, t34)+0.2, 2.5]	random [t34+0.1, t12-0.1]	range: [0, 0.9], step size: 0.1	40	30	3

part 6 TopoM (((P1,P2),(P3,P4)),O) with two introgression

direction of introgression 1	direction of introgression 2	t12	t34	t1234	t_intro_1	p_1	t_intro_2	p_2	combinations	repeats number (training dataset)	repeats number (test dataset)
one of [P1 -> P3, P3 -> P1, P2 -> P3, P3 -> P2, P1 -> P4, P4 -> P1, P2 -> P4, P4 -> P2]	one of [P1 -> P3, P3 -> P1, P2 -> P3, P3 -> P2, P1 -> P4, P4 -> P1, P2 -> P4, P4 -> P2]	random [1.0, 1.7]	random [1.0, 1.7]	random [max(t12, t34)+0.2, 2.5]	random [0.5, min(t12, t34)-0.1]	range: [0, 0.9], step size: 0.1	random [0.1, min(t12, t34, t_intro_1)-0.1]	range: [0, 0.9], step size: 0.1	64	30	3
one of [P12 -> P4, P4 -> P12, P12 -> P3, P3 -> P12]	one of [P1 -> P3, P3 -> P1, P2 -> P3, P3 -> P2, P1 -> P4, P4 -> P1, P2 -> P4, P4 -> P2]	random [1.0, 1.7]	random [t12+0.2, 2.1]	random [max(t12, t34)+0.2, 2.5]	random [t12+0.1, t34-0.1]	range: [0, 0.9], step size: 0.1	random [0.1, min(t12, t34, t_intro_1)-0.1]	range: [0, 0.9], step size: 0.1	32	30	3
one of [P34 -> P1, P1 -> P34, P34 -> P2, P2 -> P34]	one of [P1 -> P3, P3 -> P1, P2 -> P3, P3 -> P2, P1 -> P4, P4 -> P1, P2 -> P4, P4 -> P2]	random [1.0, 1.7]	random [0.5, t12-0.2]	random [max(t12, t34)+0.2, 2.5]	random [t34+0.1, t12-0.1]	range: [0, 0.9], step size: 0.1	random [0.1, min(t12, t34, t_intro_1)-0.1]	range: [0, 0.9], step size: 0.1	32	30	3

Supplementary Table S3. Parameters used in data simulation for ERICA model training and evaluation of topology inference.

Datasets	Population size (N)	Mutation rate		Recombination rate		Sample size per population	Error rates/missing rates
		μ (per site per generation)	$4N\mu$	r (per site per generation)	$4Nr$		
Training dataset	1 M	2.5×10^{-9}	0.01	2.5×10^{-9}	0.01	8	0
Test dataset (D1)	1 M	2.5×10^{-9}	0.01	2.5×10^{-9}	0.01	8	0
Recombination rate test (D2)	1 M	2.5×10^{-9}	0.01		0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5	8	0
Substitution rate test (D3)	1 M		0.001, 0.005, 0.01, 0.05, 0.1, 0.5	2.5×10^{-9}	0.01	8	0
Population size test (D4)	0.1 M, 0.5 M, 1 M, 5 M, 10 M, 50 M	2.5×10^{-9}		2.5×10^{-9}		8	0
Sample size test (D5)	1 M	2.5×10^{-9}	0.01	2.5×10^{-9}	0.01	1 to 8	0
Sequence error rate test (D6)	1 M	2.5×10^{-9}	0.01	2.5×10^{-9}	0.01	8	0, 0.001, 0.01, 0.02
Sequence missing rate test (D7)	1 M	2.5×10^{-9}	0.01	2.5×10^{-9}	0.01	8	0, 0.01, 0.1, 0.2

**Supplementary Table S4. Sample information and sequencing statistics.
Samples of *Heliconius***

Sample	Taxon	Data Source	Location	Reads (Gb)	Genotype Calls (Qual>50)	Mean Depth (Qual>50)
eth11045	<i>H. ethilla narcaea</i>	PRJNA308754	Brazil	7.4	95.5 Mb	19.16
eth11050		PRJNA308754	Brazil	4.9	94.6 Mb	12.82
eth11051		PRJNA308754	Brazil	5.0	94.7 Mb	12.88
eth11052		PRJNA308754	Brazil	7.4	95.3 Mb	19.38
JM-11-572	<i>H. melpomene aglaope</i>	PRJEB1749	Peru	12.9	90.7 Mb	20.64
09-108		PRJEB1749	Peru	14.2	89.8 Mb	20.17
JM-11-569		PRJEB1749	Peru	15.3	91.4 Mb	24.67
HMEL.AG		PRJNA73595	Peru	9.7	82.6 Mb	23.41
MJ11-3188	<i>H. melpomene amaryllis</i>	PRJEB11772	Peru	9.2	100.4 Mb	22.81
MJ11-3189		PRJEB11772	Peru	10.2	100.7 Mb	25.15
11-160		PRJEB1749	Peru	15.0	101.2 Mb	33.63
09-216		PRJEB1749	Peru	11.2	101.2 Mb	25.90
MJ12-3221	<i>H. timareta thelxinoe</i>	PRJEB11772	Peru	10.0	100.9 Mb	22.41
MJ12-3233		PRJEB11772	Peru	8.7	100.4 Mb	23.34
MJ12-3308		PRJEB11772	Peru	8.4	100.3 Mb	22.84
MJ11-3339		PRJEB11772	Peru	9.5	99.9 Mb	23.77

Samples of *Oryza*

Sample	Taxon	Cultivar	Original area	Genome assembly (bp)	Genotype Calls
Nipponbare ^b	<i>temperate japonica</i>	Nipponbare	Japan	335,723,072	7,469,786
GP551		Nonglinnuo-144	Japan	341,152,570	7,373,501
GP567		Pama	IRRI	384,964,699	7,165,802
GP669		Teng-870	Japan	409,839,250	7,449,146
GP677		Milyang84	North Korea	373,744,997	7,423,196
HP13-2		Fengrunhandaozi	China	379,343,915	7,283,102
HP14 ^b		Qiananshuinuo	China	383,588,910	7,412,190
HP38		Dabainuo	China	340,659,679	6,967,652
HP44 ^b		Honggu-1	China	416,546,134	7,373,480
HP45 ^b		Changlijingzhan	China	372,774,091	7,203,813
HP48 ^b		Hongmangnuo	China	388,430,888	7,130,686
HP91-2		Changmadao-3	China	381,770,358	7,013,912
HP98		Luyu132-1	China	390,940,925	7,247,138
HP103 ^b		Binxianludao	China	383,259,616	7,278,823
HP314 ^b		Lengshuizhan	China	411,705,109	7,151,723
HP390		Dongying-8	China	474,246,301	7,543,118
WYG7		Wuyungeng 7	China	343,912,582	7,002,181
KY131		Kongyu 131	China	349,532,929	7,549,926

DHX2		Daohuaxiang 2	China	350,648,452	7,358,983
IL9		Tohoku IL9	Japan	346,336,995	7,540,899
Koshihikari		Koshihikari	Japan	349,267,255	7,619,083
LG31		Longgeng 31	China	351,272,523	7,464,206
UR28 ^b		Honghandao	China	345,011,652	7,015,456
GP39 ^b		IAC 25	Brazil	376,584,850	7,081,162
GP77 ^b		IR 2071-88-8-10	IRRI	405,763,174	6,949,046
GP536 ^b	tropical japonica	Minanok (glut)	Philippine	422,328,976	7,005,583
GP640 ^b		Tebonnet	US	336,127,868	6,894,966
GP761-1 ^b		Mutiara	Indonesia	393,958,316	7,185,814
Osind ^{a, b}		93-11	NA	374,545,499	6,596,576
GLA4 ^b		Guangluai 4	China	344,078,917	6,482,238
GP3		Secano do Brazil	Brazil	398,466,374	6,614,052
GP22 ^b		IR 2797-115-3	IRRI	350,106,709	6,383,907
GP51		T 480-34	Cuba	391,053,916	6,533,527
GP72 ^b		IR 9209-26-2	IRRI	406,076,252	6,485,231
GP772-1		Xiangnan-1	North Korea	423,625,978	6,745,465
HP119 ^b		Huakemangnuo	China	361,592,368	6,599,654
HP263		Ruanzhanjiugu	China	424,117,304	6,414,622
HP274 ^b	indica	Bairizao-4	China	433,598,500	6,390,203
HP327		Maguhongmi	China	420,487,662	6,527,385
HP362-2 ^b		Bangbanggu	China	427,490,131	6,518,467
HP383		Guihuanuo	China	444,424,740	6,602,062
HP396 ^b		Sizhan	China	467,040,233	6,541,891
HP407		Aigansazhima	China	361,446,731	6,504,226
HP486		Jinzhan	China	451,321,993	6,532,757
HP492		Shuihegu	China	453,178,682	6,440,712
HP517-1		Dishuigu	China	412,870,524	6,600,902
HP577		Yanzhan	China	341,152,570	6,535,221
GP104		DV 110	Bengal	379,082,993	6,570,415
GP124		DNJ 126	Bengal	418,184,594	6,545,221
GP540	aus rice	Kbarai Ganja	Pakistan	347,887,577	6,361,364
GP62		ARC 7043	India	418,576,242	6,551,308
Kasalath		Kasalath	India	342,788,933	6,480,186
Orufi ^{a, b}		W1943	NA	339,177,042	6,051,109
W1943 ^b		NA	China	325,555,038	6,496,025
W3078-2 ^b		NA	China	360,204,359	6,647,361
W3095-2 ^b		NA	China	388,419,790	6,744,189
W0141	O. rufipogon	NA	India	363,990,130	6,090,016
W1687		NA	India	401,782,274	6,278,939
W1739		NA	India	351,661,657	6,334,346
W1777		NA	India	368,695,535	6,212,301
W1979		NA	Indonesia	380,890,955	5,934,654
W2012		NA	India	341,288,019	6,159,846

Oniva ^{a, b}		IRGC:100897	NA	337,950,324	5,777,773
W0123-1 ^b		NA	India	353,583,803	6,450,433
W0170 ^b	<i>O. nivara</i>	NA	Thailand	360,890,596	6,355,016
W1698 ^b		NA	Thailand	361,700,642	6,343,568
W1754 ^b		NA	India	353,556,105	6,362,737
Obart ^{a, b}	<i>O. barthii</i>	IRGC:105608	NA	308,272,304	4,985,453

a Chromosome-level assembly from OGE/IOMAP

b Samples used in ERICA analyses

Supplementary Table S5. The raw estimates and the calibrated values of G-PhoCS analyses

Single migration model												
	Replicate 1			Replicate 2				Replicate 3				
	Scaled estimates (x 10 ³)	95% HPD interval	m_{total}^a	95% HPD interval	Scaled estimates (x 10 ³)	95% HPD interval	m_{total}	95% HPD interval	Scaled estimates (x 10 ³)	95% HPD interval	m_{total}	95% HPD interval
mJ->R ^b	0.828	0.608-1.034	0.431	0.317-0.539	0.122	0-0.499	0.023	0-0.092	0.868	0.673-1.061	0.717	0.556-0.876
mJ->I ^c	9.846	2.674-23.879	0.265	0.072-0.642	9.754	3.181-19.646	0.258	0.084-0.521	8.617	2.641-17.632	0.248	0.076-0.507
mJ->N	0.001	0-0.003	0.000	0-0	0.001	0-0.003	0.000	0-0	0.000	0-0.003	0.000	0-0
mJ->G	0.059	0-0.135	0.000	0-0	0.075	0-0.152	0.000	0-0	0.061	0-0.126	0.000	0-0
mJ->B	1.806	0-7.121	0.005	0-0.021	2.020	0-7.204	0.006	0-0.021	1.470	0-5.968	0.004	0-0.017
mR->J	0.008	0-0.055	0.001	0-0.008	0.367	0-0.817	0.111	0-0.247	0.869	0.681-1.043	0.745	0.585-0.895
mR->I ^c	130.449	34.554-258.685	0.326	0.086-0.647	166.369	57.597-395.167	0.336	0.116-0.798	151.662	28.057-340.368	0.356	0.066-0.8
mR->N	0.001	0-0.003	0.000	0-0	0.000	0-0.002	0.000	0-0	0.000	0-0.003	0.000	0-0
mR->G	0.054	0-0.127	0.000	0-0	0.053	0-0.12	0.000	0-0	0.060	0-0.149	0.000	0-0
mR->B	2.452	0-3.635	0.007	0-0.01	1.324	0-5.815	0.004	0-0.016	0.949	0-4.407	0.003	0-0.013
mI->J ^c	8.573	1.006-28.812	0.226	0.027-0.76	16.306	0.89-54.93	0.331	0.018-1.115	0.270	0.193-0.358	0.096	0.069-0.128
mI->R	0.003	0-0.018	0.000	0-0.002	0.002	0-0.015	0.000	0-0.002	0.002	0-0.011	0.000	0-0.002
mI->N	0.151	0-0.249	0.084	0-0.138	0.262	0-0.447	0.218	0-0.371	0.007	0-0.041	0.003	0-0.019
mI->G	0.166	0-0.253	0.000	0-0.001	0.174	0-0.42	0.000	0-0.001	0.084	0-0.256	0.000	0-0.001
mI->B	0.474	0-0.703	0.001	0-0.002	0.578	0-2.396	0.002	0-0.007	0.366	0-1.676	0.001	0-0.004
mN->J	0.000	0-0.002	0.000	0-0	0.000	0-0.002	0.000	0-0	0.000	0-0.002	0.000	0-0
mN->R	0.001	0-0.003	0.000	0-0	0.000	0-0.002	0.000	0-0	0.001	0-0.003	0.000	0-0
mN->I	0.323	0.169-0.488	0.279	0.146-0.422	0.005	0-0.029	0.002	0-0.012	0.011	0-0.063	0.005	0-0.028
mN->G	0.074	0-0.209	0.000	0-0.001	0.072	0-0.187	0.000	0-0	0.104	0-0.209	0.000	0-0.001
mN->B	0.103	0-0.15	0.000	0-0	0.075	0-0.198	0.000	0-0.001	0.087	0-0.237	0.000	0-0.001
mG->J	0.082	0-0.246	0.000	0-0.001	0.120	0-0.345	0.000	0-0.001	0.107	0-0.239	0.000	0-0.001
mG->R	2.468	0-17.262	0.007	0-0.047	0.839	0-4.207	0.002	0-0.011	0.537	0-2.397	0.001	0-0.007
mG->I	0.075	0-0.199	0.000	0-0.001	0.071	0-0.155	0.000	0-0	0.111	0-0.316	0.000	0-0.001
mG->N	0.212	0-0.254	0.001	0-0.001	0.088	0-0.302	0.000	0-0.001	0.282	0-0.341	0.001	0-0.001
mG->B	1.118	1.059-1.179	2.949	2.791-3.108	3.633	0-5.712	0.226	0-0.355	1.135	1.052-1.287	2.768	2.565-3.137
mB->J	0.097	0-0.217	0.000	0-0.001	0.083	0-0.222	0.000	0-0.001	0.090	0-0.247	0.000	0-0.001
mB->R	0.636	0-3.054	0.002	0-0.009	2.496	0-10.364	0.007	0-0.027	1.538	0-6.888	0.004	0-0.018
mB->I	0.091	0-0.182	0.000	0-0	0.057	0-0.147	0.000	0-0	0.074	0-0.185	0.000	0-0
mB->N	0.082	0-0.239	0.000	0-0.001	0.102	0-0.286	0.000	0-0.001	0.180	0-0.379	0.000	0-0.001
mB->G	1.116	1.057-1.176	2.868	2.717-3.023	1.838	0-2.817	0.524	0-0.803	4.081	1.666-7.797	0.564	0.23-1.078

Significant migration model

	Replicate 1			Replicate 2				Replicate 3				
	Scaled estimates (x 10 ³)	95% HPD interval	m_{total}	95% HPD interval	Scaled estimates (x 10 ³)	95% HPD interval	m_{total}	95% HPD interval	Scaled estimates (x 10 ³)	95% HPD interval	m_{total}	95% HPD interval
mJ->I ^c	6.764	2.579-12.883	0.231	0.088-0.44	8.157	1.717-22.703	0.290	0.061-0.807	4.622	2.753-7.187	0.202	0.12-0.314
mR->I	0.041	0-0.244	0.001	0-0.008	0.050	0-0.24	0.002	0-0.009	0.051	0-0.377	0.002	0-0.016
mI->J	0.013	0-0.073	0.000	0-0.002	0.012	0-0.072	0.000	0-0.003	0.009	0-0.053	0.000	0-0.002

Full model with migration from *japonica* to *indica*

	Replicate 1				Replicate 2				Replicate 3				Mean
	Scaled estimates (x 10 ⁻⁴)	95% HPD interval	Effective population size	95% HPD interval	Scaled estimates (x 10 ⁻⁴)	95% HPD interval	Effective population size	95% HPD interval	Scaled estimates (x 10 ⁻⁴)	95% HPD interval	Effective population size	95% HPD interval	Effective population size
$\theta_{O. s. japonica}^d$	0.282	0.059-0.513	1,083	228-1,974	0.278	0.082-0.494	1,070	317-1,902	0.309	0.082-0.533	1,187	313-2,050	1,113
$\theta_{O. rufipogon}$	0.997	0.000-3.005	3,835	0-11,557	1.011	0.000-3.036	3,887	0-11,678	0.989	0.000-2.984	3,804	0-11,475	3,842
$\theta_{O. s. indica}$	0.981	0.000-2.973	3,773	0-11,436	1.008	0.000-3.012	3,878	0-11,584	0.998	0.000-2.986	3,840	0-11,485	3,830
$\theta_{O. nivara}$	0.995	0.000-2.987	3,827	0-11,489	1.003	0.000-2.983	3,858	0-11,473	0.990	0.000-2.957	3,808	0-11,373	3,831
$\theta_{O. glaberrima}$	1.013	0.000-3.037	3,894	0-11,680	1	0.000-2.976	3,847	0-11,447	0.997	0.000-3.005	3,836	1-11,558	3,859
$\theta_{O. barthii}$	1.01	0.000-3.032	3,885	0-11,661	0.991	0.000-2.972	3,812	0-11,429	0.985	0.000-2.951	3,787	0-11,352	3,828
$\theta_{O. meridionalis}$	0.997	0.000-2.988	3,833	0-11,493	1	0.000-2.978	3,847	0-11,453	0.995	0.000-2.983	3,827	0-11,472	3,836
θ_{JR}	22.275	20.555-24.077	85,673	79,058-92,603	22.532	21.049-24.176	86,660	80,957-92,986	22.351	20.770-24.002	85,966	79,884-92,316	86,100
θ_{IN}	22.073	18.607-25.756	84,896	71,563-99,061	26.272	23.518-29.136	101,045	90,453-112,063	23.046	20.954-25.553	88,640	80,591-98,281	91,527
θ_{GB}	17.777	16.837-18.669	68,373	64,757-71,804	17.604	16.672-18.516	67,708	64,122-71,214	17.753	16.813-18.742	68,280	64,664-72,084	68,121
θ_{JRIN}	33.645	29.608-37.874	129,402	113,878-145,670	31.286	27.611-35.292	120,329	106,195-135,738	32.887	29.134-36.581	126,488	112,053-140,695	125,406
θ_{JRINGB}	23.452	19.071-27.410	90,201	73,351-105,424	26.167	21.63-31.079	100,641	83,191-119,534	24.088	19.235-28.260	92,648	73,982-108,693	94,496
θ_{Root}	245.927	237.524-254.542	945,874	913,553-979,009	244.622	236.58-253.321	940,853	909,923-974,310	245.201	237.469-253.164	943,079	913,342-973,708	943,269
	Scaled estimates (x 10 ⁻⁴)	95% HPD interval	Calibrated speciation time (year)	95% HPD interval	Scaled estimates (x 10 ⁻⁴)	95% HPD interval	Calibrated speciation time (year)	95% HPD interval	Scaled estimates (x 10 ⁻⁴)	95% HPD interval	Calibrated speciation time (year)	95% HPD interval	Calibrated speciation time (year)
τ_{JR}^c	0.269	0.057-0.503	4,137	871-7,732	0.265	0.070-0.479	4,077	1,077-7,375	0.287	0.084-0.520	4,422	1,291-7,998	4,212
τ_{IN}	2.937	2.038-3.669	45,189	31,360-56,445	1.732	1.298-2.107	26,646	19,962-32,422	2.658	2.233-3.051	40,892	34,357-46,934	37,576
τ_{GB}	0.029	0.000-0.085	440	0-1,311	0.03	0.000-0.088	462	1-1,358	0.029	0.000-0.086	442	0-1,317	448
τ_{JRIN}	13.303	12.495-14.187	204,658	192,237-218,254	13.69	12.904-14.469	210,622	198,518-222,605	13.422	12.640-14.172	206,498	194,465-218,031	207,259
τ_{JRINGB}	25.321	24.462-26.178	389,552	376,345-402,734	24.898	23.897-25.911	383,038	367,649-398,625	25.190	24.255-26.160	387,543	373,160-402,458	386,711
τ_{Root}	35.85	34.283-37.402	551,542	527,435-575,412	36.817	35.266-38.344	566,412	542,554-589,906	35.985	34.460-37.387	553,617	530,158-575,182	557,190
	Scaled estimates (x 10 ³)	95% HPD interval	m_{total}	95% HPD interval	Scaled estimates (x 10 ³)	95% HPD interval	m_{total}	95% HPD interval	Scaled estimates (x 10 ³)	95% HPD interval	m_{total}	95% HPD interval	m_{total}
$m_{J \rightarrow I}^c$	9.846	2.674-23.879	0.265	0.072-0.642	9.754	3.181-19.646	0.258	0.084-0.521	8.617	2.641-17.632	0.248	0.076-0.507	0.257

^a total migration rate (m_{total}) = migration rate (m) x scaled speciation time (τ)

^b J for *O. s. japonica*, R for *O. rufipogon*, I for *O. s. indica*, N for *O. nivara*, G for *O. glaberrima* and B for *O. barthii*

^c All of three replicates have a lower limit of the HPD intervals greater than 0

^d $\theta = 4N\mu g$, where N is the effective population size, μ is the mutation rate (6.5×10^{-9} per site per generation), g is the generation time (1 year)

^e $\tau = T\mu/g$, where T is the absolute speciation time

Supplementary Table S6. The sensitivities of ERICA and other algorithms for detecting introgression under different evolutionary scenarios.

	Method	Split time			Population size		Window size		Error rate	
		D11	D12	D13	D14	D15	D16	D17	D18	D19
Neutral Introgression (Four-taxon)	ERICA	68.7	44.8	7.7	64.4	47.6	81.0	65.1	67.1	67.9
	D	64.9	49.6	17.9	62.3	49.3	74.8	66.3	60.7	64.4
	f_d	62.2	42.9	12.5	56.4	44.0	72.9	60.2	59.4	62.4
	IBDmix	72.5	57.6	17.7	68.4	45.6	90.1	69.6	55.2	26.1
	Sprime	39.6	28.8	16.2	36.2	28.8	54.8	54.2	27.4	26.4
Adaptive Introgression (Four-taxon)	ERICA	100	99.9	58.0	100	96.0	100	100	100	100
	D	97.6	87.8	44.7	94.7	77.1	99.8	98.3	94.7	97.5
	f_d	97.0	83.6	31.5	93.7	71.1	99.6	97.7	94.9	97.2
	IBDmix	61.8	44.2	16.2	49.7	31.1	98.2	73.8	33.4	22.8
	Sprime	100	95.2	35.4	96.6	70.0	100	98.0	61.4	89.0
Neutral Introgression (Five-taxon)	ERICA	43.1	19.1	0.4	36.4	14.6	49.5	32.2	44.0	47.4
	D_{FOIL}	26.8	9.2	0.0	17.6	1.4	51.6	29.3	14.6	18.4
	IBDmix	71.1	54.2	16.0	63.2	45.6	87.9	72.8	56.5	21.6
	Sprime	17.4	6.0	6.2	21.0	11.7	22.8	20.2	26.4	9.0
Adaptive Introgression (Five-taxon)	ERICA	99.7	77.0	9.5	95.0	37.1	99.7	84.9	99.5	99.5
	D_{FOIL}	71.5	31.7	1.8	47.5	2.9	97.9	66.7	43.6	54.0
	IBDmix	59.7	37.3	18.1	53.8	29.0	95.9	77.2	34.5	18.6
	Sprime	82.6	60.7	19.8	75.5	32.1	88.7	71.6	58.2	48.5

Evolutionary scenarios used:

Dataset D11: population size (N): 1 M; split time: t_{12} : 1, t_{123} : 2, and t_R : 3 (in units of $4N$ generations) for four-taxon case, and t_{12} : 1, t_{34} : 1.5, t_{1234} : 2, and t_R : 3 for five-taxon case; window size: 5 kb; error rate: 0.

Dataset D12: N : 1 M; split time: t_{12} : 0.5, t_{123} : 1, and t_R : 1.5 (in units of $4N$ generations) for four-taxon case, and t_{12} : 0.5, t_{34} : 0.75, t_{1234} : 1, and t_R : 1.5 for five-taxon case; window size: 5 kb; error rate: 0.

Dataset D13: N : 1 M; split time: t_{12} : 0.1, t_{123} : 0.2, and t_R : 0.3 (in units of $4N$ generations) for four-taxon case, and t_{12} : 0.1, t_{34} : 0.15, t_{1234} : 0.2, and t_R : 0.3 for five-taxon case; window size: 5 kb; error rate: 0.

Dataset D14: N : 0.5 M; split time: the same as D11; window size: 5 kb; error rate: 0.

Dataset D15: N : 0.1 M; split time: the same as D11; window size: 5 kb; error rate: 0.

Dataset D16: N : 0.5 M; split time: the same as D11; window size: 50 kb; error rate: 0.

Dataset D17: N : 0.1 M; split time: the same as D11; window size: 50 kb; error rate: 0.

Dataset D18: N : 1 M; split time: the same as D11; window size: 5 kb; error rate: 2%.

Dataset D19: N : 1 M; split time: the same as D11; window size: 5 kb; missing rate: 10%.

Supplementary Table S7. Time and memory cost of ERICA models.

Method	100 kb	1 Mb	10 Mb	100 Mb
ERICA (four-taxon) ^a	0:29 (3.3 Gb)	0:26 (3.4 Gb)	1:37 (4.5 Gb)	14:01 (31.2 Gb)
ERICA (five-taxon) ^a	0:45 (3.3 Gb)	0:53 (3.5 Gb)	2:37 (5.3 Gb)	20:40 (38.7 Gb)
ABBA-BABA test ^b	0:08	0:20	3:13	31:41
D_{FOIL} ^c	0:18	2:56	30:54	328:44

Times are minutes: seconds.

a. The ERICA models ran on one NVIDIA Tesla V100 SXM2 32GB GPU.

b. The ABBA-BABA test ran on one Intel Xeon Gold 6130 CPU using 20 threads.

c. The D_{FOIL} test ran on one Intel Xeon Gold 6130 CPU using one thread.

Supplementary Table S8. Top 15 outlier regions supporting ((*H. m. amaryllis*, *H. t. thelxinoe*), *H. m. aglaope*), *H. ethilla*) in the *Heliconius* genome.

Chromosome	Position Start	Proportion of Topo C	D	f_d	d_{xy}
Hmel218003o	750001 ^a	0.6729	0.7958 ± 0.0525***	0.5777 ± 0.0382***	0.0118 ± 0.0013***
Hmel218003o	800001 ^a	0.6295	0.8629 ± 0.0513***	0.5417 ± 0.1577***	0.0113 ± 0.0011***
Hmel218003o	700001 ^a	0.5871	0.6286 ± 0.1008***	0.4276 ± 0.082***	0.0122 ± 0.0017***
Hmel215003o	1400001	0.5621	0.6803 ± 0.0993***	0.3783 ± 0.0941***	0.0201 ± 0.0034***
Hmel218003o	15250001 ^b	0.4892	0.0803 ± 0.0422	0.0774 ± 0.0517	0.0205 ± 0.0054**
Hmel206001o	1500001	0.4852	0.1795 ± 0.0690**	0.1643 ± 0.0884	0.0277 ± 0.0048
Hmel215003o	6200001	0.4716	0.2106 ± 0.0612***	0.1756 ± 0.0567**	0.0358 ± 0.0015
Hmel203003o	7450001	0.4688	0.2397 ± 0.0881**	0.2379 ± 0.1123*	0.0271 ± 0.0026
Hmel215003o	1500001	0.4662	0.3764 ± 0.1430**	0.2388 ± 0.0963*	0.0245 ± 0.0050
Hmel211001o	6400001	0.4613	0.3218 ± 0.1401*	0.3281 ± 0.1344*	0.0227 ± 0.0049*
Hmel203003o	150001 ^b	0.4611	0.1456 ± 0.0885	0.0595 ± 0.0592	0.0158 ± 0.0013***
Hmel218003o	13950001	0.4597	0.2434 ± 0.0636***	0.2297 ± 0.0670***	0.0335 ± 0.0053
Hmel207001o	950001 ^b	0.4559	0.1774 ± 0.1179	0.1990 ± 0.1346	0.0239 ± 0.0055
Hmel207001o	12600001	0.4559	0.2351 ± 0.0664***	0.1797 ± 0.0657**	0.0200 ± 0.0066
Hmel215003o	8750001	0.4532	0.1413 ± 0.0226***	0.1633 ± 0.0245***	0.0332 ± 0.0013

^a *B/D* locus

^b Candidate introgressed loci which were not detected by D and f_d .

Supplementary Table S9. Putative regions of introgression between Asian cultivated rice and between tropical *japonica* rice accessions.

Introgressed regions between *japonica* and *indica*

Chr	Start	End	Overlapped selective sweep regions ^a (Mb)	Associated traits ^a	Annotated genes	Description	
1	1200001	1350000	1.2-1.4				
1	6050001	6100000					
1	9000001	9100000			OsGGR2 (LOC_Os01g16020) (9.0Mb) OsPAPST1 (LOC_Os01g16040) (9.0Mb) RBP-P (LOC_Os01g16090) (9.1Mb)	Geranylgeranyl reductase, α-tocopherol synthesis Mitochondrial carrier protein, Early leaf development RNA binding protein, RNA metabolism, Regulation of genes important for plant growth and development	
1	21050001	21150000					
1	42550001	42650000			YGL8 (LOC_Os01g73450) (42.6Mb)	Chloroplast-targeted UMP kinase, Regulation of leaf colour	
2	1	100000			OsHSP40 (LOC_Os02g01030) (0Mb) DPW3 (LOC_Os02g01070) (0Mb) OsHPR1 (LOC_Os02g01150) (0.1Mb)	Heat shock protein (HSP40), Putative tetratricopeptide repeat (TPR)-containing protein, Growth and development, salt tolerance, abiotic stress tolerance Alpha integrin-like protein, Pollen wall formation Peroxisomal hydroxypyruvate reductase, NADH-dependent HPR, Photorespiratory metabolism	
2	3450001	3600000	3.5-3.8				
2	9700001	9850000					
2	10900001	10950000					
2	22900001	23000000					
3	6350001	6450000					
3	25150001	25250000			OsSh1 (LOC_Os03g44710) (25.2Mb) ^b	YABBY transcription factor, Control of seed shattering	
3	25450001	25500000	25.5-27.1	Panicle length			
3	25900001	26000000	25.5-27.1	Panicle length	OsTUB8 (LOC_Os03g45920) (25.9Mb)	Beta-tubulin, Microtubules formation	
3	26200001	26300000	25.5-27.1	Panicle length	OsSWAP70A (LOC_Os03g46340) (26.2Mb)	Rac/Rop guanine nucleotide exchange factor, Regulation of immune responses	
3	27550001	27650000					
3	28850001	28950000			PIBP1 (LOC_Os03g50560) (28.9Mb)	RRM (RNA recognition motif) protein, Transcription factor, Regulation of blast resistance	
3	29050001	29150000			OsHK4 (LOC_Os03g50860) (29.1Mb) OsSIZ2 (LOC_Os03g50980) (29.1Mb)	Histidine kinase, Cytokinin signaling SUMO (small ubiquitin-related modifier) E3-ligase, Abiotic stress response, Stress adaptation	

4	15450001	15550000	15.2-17.5	Germination rate		
4	16100001	16150000	15.2-17.5	Germination rate		
4	16300001	16400000	15.2-17.5	Germination rate		
4	16750001	16950000	15.2-17.5	Germination rate		
4	17250001	17350000	15.2-17.5	Germination rate		
4	21150001	21200000	21.1-21.3			
4	23400001	23450000	23.3-23.8	Hull color		
4	25650001	25750000	25.8-27.5			
4	27000001	27100000	25.8-27.5		STRK1 (LOC_Os04g45730) (27.1Mb)	Receptor-like cytoplasmic kinase, Salt tolerance, Oxidative stress tolerance
5	4150001	4250000	4.1-4.2		OsTAR1 (LOC_Os05g07720) (4.2Mb)	Tryptophan aminotransferase, Indole-3-acetic acid (IAA) biosynthesis, Grain development
5	8600001	8650000				
5	9300001	9400000	9.1-13.5			
5	9650001	9750000	9.1-13.5			
5	10300001	10450000	9.1-13.5			
5	10800001	10900000	9.1-13.5			
5	11350001	11500000	9.1-13.5			
5	11600001	11950000	9.1-13.5		EBR1 (LOC_Os05g19970) (11.7Mb)	RING-type E3 ubiquitin ligase, Control of innate immunity and broad-spectrum disease resistance
5	12100001	12150000	9.1-13.5			
5	12250001	12350000	9.1-13.5			
5	12550001	12650000	9.1-13.5			
5	13250001	13500000	9.1-13.5			
5	16850001	16950000				
5	23600001	23650000			OsSTN8 (LOC_Os05g40180) (23.6Mb) OsSNAT1 (LOC_Os05g40260) (23.6Mb)	Serine/threonine protein kinase, Photosystem II (PSII) repair, Phosphorylation of PSII core protein Serotonin N-acetyltransferase, Melatonin synthesis
5	29650001	29700000	29.7-29.9		NRRb (LOC_Os05g51690) (29.7Mb) OsNDPK3 (LOC_Os05g51700) (29.7Mb)	NRR alternative splicing variant, Regulation of root development in response to macronutrient deficiency Nucleoside diphosphate kinase
6	900001	1050000				
6	3950001	4000000	3.7-4.0	Stigma exertion		
6	5050001	5100000	5- 5.6	Stigma color	OsHAL3 (LOC_Os06g09910) (5.1Mb)	Halotolerance protein, Flavin mononucleotide (FMN)-binding protein, Regulation of cell division, Light-regulated growth

6	5200001	5300000	5- 5.6	Stigma color	OsRLCK204 (LOC_Os06g10230) (5.3Mb) OsC1 (LOC_Os06g10350) (5.3Mb) ^c	Receptor-like kinase, Heat tolerance R2R3-Myb factor, Homolog of maize C1, Determinant of anthocyanin biosynthesis in rice leaves, Apiculus and leaf sheath colour
6	6650001	6750000			OsLsi6 (LOC_Os06g12310) (6.7Mb) ALK (LOC_Os06g12450) (6.7Mb)	Aquaporin NIP III subfamily protein, Aquaporin, Silicon influx transporter, Water transport Soluble starch synthase II-3, Endosperm starch synthesis, Determination of the type of amylopectin structure of starch grain
6	7050001	7100000				
6	7650001	7800000			OsPFPB (LOC_Os06g13810) (7.6Mb) OsGORK (LOC_Os06g14030) (7.8Mb)	Pyrophosphate-fructose 6-phosphate 1-phosphotransferase (PFP) beta subunit, Regulation of carbon metabolism during grain filling Outward Shaker K ⁺ channel, K ⁺ release by guard cells for stomatal closure, K ⁺ translocation toward the shoots
6	9550001	9600000				
6	21650001	22100000			OsMAX1b (LOC_Os06g36920) (21.8Mb) OsAAO1 (LOC_Os06g37080) (21.9Mb) OsAAO2 (LOC_Os06g37150) (22.0Mb) OsKO4 (LOC_Os06g37300) (22.0Mb) OsKO1 (LOC_Os06g37330) (22.0Mb)	Cytochrome P450 family member, Homolog of Arabidopsis MORE AXILLARY GROWTH 1 (MAX1) Ascorbate oxidase, Abiotic stress response L-ascorbate oxidase, Cold, salt, drought stress response Ent-kaurene oxidase, Diterpenoid phytoalexin biosynthesis Ent-kaurene oxidase 1, Gibberellin biosynthesis, Regulation of seed germination
7	300001	450000			OsDLK (LOC_Os07g01490) (0.3Mb) OsPSKR3 (LOC_Os07g01710) (0.4Mb)	Class-XIV kinesin Phytosulfokine (PSK) receptor-like protein
7	2800001	2850000	2.4-3.0	Tiller angle	PROG1 (LOC_Os07g05900) (2.8Mb) ^b	Zinc-finger nuclear transcription factor, Control of prostrate growth, prog1 variants identified in <i>O. sativa</i> disrupt the prog1 function
7	3700001	3800000	3.7-4.1		OsGA20ox3 (LOC_Os07g07420) (3.7Mb)	GA 20-oxidase3, GA metabolism
7	4000001	4050000	3.7-4.1		OsNek3 (LOC_Os07g08000) (4.0Mb)	Serine/threonine protein kinase, Never-in-mitosis A (NIMA)-related kinase, Pollen germination
7	6100001	6200000	6.1-6.3		Rc (LOC_Os07g11020) (6.1Mb) ^c	Basic helix-loop-helix (bHLH) protein, Proanthocyanidin synthesis, Regulation of proanthocyanidin pigmentation
7	27300001	27450000				
8	23650001	23750000	23.8-24.9	Stigma exertion, awn length		
8	24100001	24300000	23.8-24.9	Stigma exertion, awn length	OsDOF24 (LOC_Os08g38220) (24.2Mb)	Transcription factor with DOF (DNA-binding one zinc finger) domain, Repressor of leaf senescence, Negative regulation of JA biosynthesis pathway
8	24500001	24550000	23.8-24.9	Stigma exertion, awn length	OsHAP5D (LOC_Os08g38780) (24.5Mb)	NUCLEAR FACTOR Y (NF-Y) transcription factor, Control of heading date

8	24800001	24900000	23.8-24.9	Stigma exertion, awn length	RCN11 (LOC_Os08g39380) (24.9Mb)	Beta 1, 2-xylosyltransferase, Response to abiotic stresses and phytohormones
8	25100001	25150000				
9	9550001	9650000			OsPP2C1 (LOC_Os09g15670) (9.6Mb)	Protein phosphatase 2C, Abiotic stress response, Early panicle development
9	11550001	11650000				
9	21500001	21650000	21.4-21.5		OsRopGEF7 (LOC_Os09g37270) (21.5Mb) OsSAUR39 (LOC_Os09g37330) (21.6Mb) OsSAUR45 (LOC_Os09g37400) (21.6Mb) LOGL9 (LOC_Os09g37540) (21.6Mb)	Small GTPase Rac/ROP guanine nucleotide exchange factor, Signal transduction SAUR family protein, Negative regulator of auxin synthesis and transport Small auxin-up RNA (SAUR), Auxin-responsive SAUR gene family member, Auxin synthesis and transport, Plant growth Lysine decarboxylase-like protein, Response to abiotic stresses
10	13000001	13050000				
10	22400001	22450000				
11	200001	300000				
12	1150001	1250000	1.2-1.3		TL (1.2Mb) OsMYB60 (LOC_Os12g03150) (1.2Mb)	Antisense long noncoding RNA, Maintenance of leaf blade flattening, Suppression of sense gene expression by mediating chromatin modifications R2R3 MYB transcription factor, Leaf development
12	2850001	2950000	2.6-2.9			
12	25100001	25250000			OsWRKY94 (LOC_Os12g40570) (25.1Mb)	WRKY transcription factor, Cold tolerance

^a Huang et al. A map of rice genome variation reveals the origin of cultivated rice. Nature, 490(7421), 497-501.

^b Domestication genes located within introgressed loci

^c Domestication genes located close to introgressed loci

Introgressed regions from *indica* / *O. nivara* to tropical *japonica*

Chr	Start	End	Annotated genes	Description
1	4250001	4300000	PEG19 (LOC_Os01g08570) (4.3Mb)	2-oxoglutarate and iron-dependent oxygenase, Starch accumulation
1	7100001	7150000	OsHPR2 (LOC_Os01g12830) (7.1Mb)	Cytosolic hydroxypyruvate reductase, NADPH-dependent HPR, Photorespiratory metabolism
1	22900001	22950000		
1	23850001	24000000	OsLUGL (LOC_Os01g42260) (23.9Mb)	LEUNIG-like (LUGL) transcriptional regulator, Regulation of floral organ development
1	24850001	24900000		

1	26400001	26450000		
1	32050001	32350000	OsTCP5 (LOC_Os01g55750) (32.1Mb) OsGH3-2 (LOC_Os01g55940) (32.2Mb) STRIPE2 (LOC_Os01g55974) (32.2Mb)	TCP family transcription factor, Strigolactone and cytokinin controlled mesocotyl elongation in darkness; Indole-3-acetic acid (IAA)-amido synthetase, Disease resistance, Abiotic stress tolerance; Putative dCMP deaminase, Chloroplast development
1	32500001	32800000	OsTYDC (LOC_Os01g56380) (32.5Mb)	Aromatic L-amino acid decarboxylase (AADC), Conversion of tyrosine into tyramine
2	4450001	4500000		
2	4750001	4800000		
2	16800001	16850000		
3	10350001	10450000		
3	10600001	10650000	OsPEX11-2 (LOC_Os03g19000) (10.6Mb); OsPEX11-3 (LOC_Os03g19010) (10.6Mb)	Peroxisomal protein; Peroxisomal protein, Peroxisomal biogenesis factor 11, Salt stress tolerance, Antioxidant defense
3	10900001	10950000	OsNAS1 (LOC_Os03g19427) (10.9Mb); OsNAS2 (LOC_Os03g19420) (10.9Mb)	Nicotianamine synthase 1; Nicotianamine synthase 2, Nicotianamine biosynthesis
3	11200001	11250000		
3	11450001	11750000	OsTGA3 (LOC_Os03g20310) (11.5Mb); OsCaM1 (LOC_Os03g20370) (11.5Mb); OsCK1 (LOC_Os03g20380) (11.5Mb); WRKY55 (LOC_Os03g20550) (11.7Mb); OsLEA17 (LOC_Os03g20680) (11.7Mb)	TGA transcription factor, bZIP transcription factor, Defense response; Calmodulin, Ca ²⁺ sensor, Ca ²⁺ signalling, Thermotolerance; Serine/threonine protein kinase, Cold stress tolerance, Abiotic stresses; Similar to WRKY transcription factor 55; Late embryogenesis abundant (LEA) protein, Abiotic stress tolerance
3	15200001	15250000	OsTT1 (LOC_Os03g26970) (15.4Mb) ^a	Proteasome alpha 2 subunit, Thermotolerance
3	15700001	15750000		
4	8300001	8350000		
5	19600001	19850000	OsPDR2 (LOC_Os05g33390) (19.6Mb); OsPPDKB (LOC_Os05g33570) (19.7Mb)	Homolog of Arabidopsis Phosphate Deficiency Response2 (AtPDR2), P5 type ATPase, Regulation of Pi homeostasis; Chloroplastic pyruvate orthophosphate dikinase (PPDK, EC 2.7.9.1), Starch metabolism and structure
5	20000001	20050000		
5	20350001	20400000		
5	27500001	27550000	OsMATE2 (LOC_Os05g48040) (27.5Mb)	Multidrug and toxic compound extrusion (MATE) protein family transporter, Arsenic stress response, Regulation of plant growth and development, Disease resistance

5	27650001	27750000	OsPUP7 (LOC_Os05g48300) (27.7Mb); OsUBC35 (LOC_Os05g48390) (27.7Mb)	Purine permease, Transport of cytokinin, Growth and development control; Ubiquitin-conjugating E2 enzyme, Phosphate homeostasis, Negative regulator of the inorganic phosphate (Pi)-starvation response-signaling pathway NAC family transcription factor, Secondary cell wall biosynthesis, Positive regulator of cellulose biosynthesis; Auxin response factor, Transcription factor, Regulation of lateral root growth;
5	28000001	28350000	OsSND2 (LOC_Os05g48850) (28.0Mb); OsARF15 (LOC_Os05g48870) (28.0Mb); DWL2 (LOC_Os05g48990) (28.1Mb); WRKY49 (LOC_Os05g49100) (28.2Mb); OsMAPK20-5 (LOC_Os05g49140) (28.2Mb); WRKY43 (LOC_Os05g49210) (28.2Mb); WSL6 (LOC_Os05g49220) (28.2Mb)	WUSCHEL-related homeobox transcription factor (WOX9) homolog, Regulation of the uniform growth of shoots; DNA-binding WRKY domain containing protein; Group D mitogen-activated protein kinase (MAPK), Defense responses to BPH, Control of excessive defense responses, Protection from defense-response-related autotoxicity Similar to WRKY transcription factor 43; Era-type GTP-binding protein, Chloroplast development Similar to WRKY transcription factor 19.
5	28450001	28500000	WRKY19 (LOC_Os05g49620) (28.5Mb)	
6	4300001	4350000		
6	30100001	30200000	OsMADS16 (LOC_Os06g49840) (30.2Mb)	Transcription factor, Development of lodicules and stamens
7	19150001	19200000	OsSPL13 (LOC_Os07g32170) (19.1 Mb) ^a	Plant-specific transcription factor, Control of grain size
7	19550001	19600000		
8	10350001	10400000		
8	11250001	11300000		
8	19800001	19850000		
8	21400001	21450000	OsCAF1A (LOC_Os08g34170) (21.4Mb)	Component of the CCR4-NOT complex, Deadenylation, Deadenylation (poly(A) tail shortening), Development and stress response
8	21700001	21750000	OsTPS8 (LOC_Os08g34580) (21.7Mb)	Class II trehalose-phosphate-synthase (TPS) gene family member, Salt stress tolerance, Control of yield-related traits
8	26300001	26350000	OsPIN5b (LOC_Os08g41720) (26.3Mb)	Auxin efflux carrier, Auxin homeostasis, transport and distribution
10	50001	200000	OsSCP45 (LOC_Os10g01134) (1.1Mb)	Serine carboxypeptidase, Regulation of grain filling and seed germination
10	300001	600000	OsCMT2 (LOC_Os10g01570) (0.4Mb)	Chromomethylase, Plant development, Silencing of transposable elements (TEs)
10	3700001	3800000		
10	6600001	6650000		
10	7400001	7450000		

10	7550001	7600000	OsFRDL2 (LOC_Os10g13940) (7.6Mb)	Multidrug and toxic compound extrusion (MATE) protein, Al-induced secretion of citrate
10	7800001	7850000		
10	8550001	8600000		
10	8750001	8800000		
10	8900001	9100000	OsLHP1 (LOC_Os10g17770) (9.0Mb)	Homolog of Arabidopsis LHP1
10	10950001	11100000		
10	13750001	13800000		
10	19750001	19900000	OsNIP3 (LOC_Os10g36924) (19.8Mb); OsGDCH (LOC_Os10g37180) (19.9Mb)	Boric acid channel, Regulation of boron distribution; H subunit of glycine decarboxylase complex (GDC), Photorespiration
11	18250001	18300000		
12	14200001	14250000		
12	16150001	16200000	OsDR11 (LOC_Os12g27520) (16.2Mb)	LAMMER kinase-type protein, Positive regulator of disease resistance
12	20000001	20050000		
12	20750001	20800000		
12	21050001	21100000		

^a Previously reported introgressed genes located close to the introgressed loci

Ontology enrichment of introgressed genes between *japonica* and *indica*

Gene Ontology	Term	Name	<i>P</i> value	Count	Annotated genes	Background count	Background annotated genes	Fold change
GOTERM_BP	GO:0050896	response to stimulus	0.0001	8	616	53	32072	7.9
GOTERM_BP	GO:0010241	ent-kaurene oxidation to kaurenoic acid	0.0002	4	616	4	32072	52.1
GOTERM_BP	GO:0009719	response to endogenous stimulus	0.0002	50	616	1486	32072	1.8
GOTERM_BP	GO:0007049	cell cycle	0.0010	18	616	365	32072	2.6
GOTERM_BP	GO:0009617	response to bacterium	0.0024	7	616	66	32072	5.5
GOTERM_BP	GO:0055085	transmembrane transport	0.0026	21	616	507	32072	2.2
GOTERM_BP	GO:0009734	auxin-activated signaling pathway	0.0063	9	616	137	32072	3.4
GOTERM_BP	GO:0010466	negative regulation of peptidase activity	0.0112	4	616	22	32072	9.5
GOTERM_BP	GO:0009058	biosynthetic process	0.0130	112	616	4709	32072	1.2
GOTERM_BP	GO:0009987	cellular process	0.0187	164	616	7318	32072	1.2
GOTERM_BP	GO:0071577	zinc ion transmembrane transport	0.0206	3	616	10	32072	15.6
GOTERM_BP	GO:0009853	photorespiration	0.0430	5	616	66	32072	3.9
GOTERM_BP	GO:0006829	zinc ion transport	0.0489	3	616	17	32072	9.2
GOTERM_CC	GO:0009707	chloroplast outer membrane	0.0257	4	616	31	32072	6.7
GOTERM_CC	GO:0033178	proton-transporting two-sector ATPase complex, catalytic domain	0.0299	4	616	33	32072	6.3
GOTERM_MF	GO:0005516	calmodulin binding	0.0056	9	616	134	32072	3.5
GOTERM_MF	GO:0030414	peptidase inhibitor activity	0.0153	4	616	25	32072	8.3
GOTERM_MF	GO:0004497	monooxygenase activity	0.0164	15	616	375	32072	2.1
GOTERM_MF	GO:0030170	pyridoxal phosphate binding	0.0223	8	616	141	32072	3.0
GOTERM_MF	GO:0080019	fatty-acyl-CoA reductase (alcohol-forming) activity	0.0240	3	616	11	32072	14.2
GOTERM_MF	GO:0005215	transporter activity	0.0329	42	616	1574	32072	1.4
GOTERM_MF	GO:0005385	zinc ion transmembrane transporter activity	0.0356	3	616	14	32072	11.2
GOTERM_MF	GO:0019825	oxygen binding	0.0435	13	616	351	32072	1.9
Plant Experimental Conditions Ontology	Term	Name	<i>P</i> value	Count	Annotated genes	Background count	Background annotated genes	Fold change
PECO	PECO:0007242	iron nutrient exposure	0.0058	2	390	7	22713	16.6
PECO	PECO:0007172	flood water exposure	0.0146	2	390	11	22713	10.6
PECO	PECO:0007257	Xanthomonas oryzae pv. oryzae exposure	0.0377	8	390	221	22713	2.1
Plant Trait Ontology	Term	Name	<i>P</i> value	Count	Annotated genes	Background count	Background annotated genes	Fold change
TO	TO:0000357	plant growth and development trait	0.0094	1	2	18	3834	106.5
TO	TO:0000599	enzyme activity trait	0.0109	1	2	21	3834	91.3

Ontology enrichment of introgressed genes from *indica* / *O. nivara* to tropical *japonica*

Gene Ontology	Term	Name	<i>P</i> value	Count	Annotated genes	Background count	Background annotated genes	Fold change
GOTERM_BP	GO:0008150	biological_process	0.0073	150	491	8141	32072	1.2
GOTERM_BP	GO:0009416	response to light stimulus	0.0094	7	491	113	32072	4.0
GOTERM_BP	GO:0009863	salicylic acid mediated signaling pathway	0.0102	4	491	27	32072	9.7
GOTERM_BP	GO:0016036	cellular response to phosphate starvation	0.0223	5	491	67	32072	4.9
GOTERM_BP	GO:0042309	homiothermy	0.0291	30	491	1295	32072	1.5
GOTERM_BP	GO:0050826	response to freezing	0.0402	30	491	1335	32072	1.5
GOTERM_BP	GO:0006139	nucleobase-containing compound metabolic process	0.0440	69	491	3614	32072	1.2
GOTERM_BP	GO:0009073	aromatic amino acid family biosynthetic process	0.0456	4	491	50	32072	5.2
GOTERM_BP	GO:0006820	anion transport	0.0464	3	491	21	32072	9.3
GOTERM_CC	GO:0005575	cellular_component	0.0046	105	491	5319	32072	1.3
GOTERM_CC	GO:0005743	mitochondrial inner membrane	0.0151	7	491	126	32072	3.6
GOTERM_MF	GO:0016747	transferase activity, transferring acyl groups other than amino-acyl groups	0.0000	13	491	179	32072	4.7
GOTERM_MF	GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	0.0068	13	491	335	32072	2.5
GOTERM_MF	GO:0019825	oxygen binding	0.0095	13	491	351	32072	2.4
GOTERM_MF	GO:0003674	molecular_function	0.0174	111	491	5946	32072	1.2
GOTERM_MF	GO:0020037	heme binding	0.0192	16	491	530	32072	2.0
GOTERM_MF	GO:0050825	ice binding	0.0291	30	491	1295	32072	1.5
GOTERM_MF	GO:0004497	monooxygenase activity	0.0332	12	491	375	32072	2.1
GOTERM_MF	GO:0003924	GTPase activity	0.0422	7	491	162	32072	2.8
Plant Experimental Conditions Ontology	Term	Name	<i>P</i> value	Count	Annotated genes	Background count	Background annotated genes	Fold change
PECO	PECO:0007270	continuous dark (no light) exposure	0.0044	33	336	1377	22713	1.6
PECO	PECO:0007373	mechanical damage exposure	0.0056	12	336	345	22713	2.4
PECO	PECO:0007226	Magnaporthe grisea exposure	0.0185	138	336	8058	22713	1.2
PECO	PECO:0007048	sodium chloride exposure	0.0297	48	336	2469	22713	1.3
PECO	PECO:0007174	cold temperature exposure	0.0314	40	336	2001	22713	1.4
Plant Trait Ontology	Term	Name	<i>P</i> value	Count	Annotated genes	Background count	Background annotated genes	Fold change
TO	TO:0000224	iron sensitivity	0.0000	2	3	10	3834	255.6
TO	TO:0000495	chlorophyll content	0.0016	2	3	90	3834	28.4
TO	TO:0000075	light sensitivity	0.0023	1	3	3	3834	426.0
TO	TO:0000104	floury endosperm	0.0241	1	3	31	3834	41.2

Supplementary Table S10. ERICA results of the rice population genomic

dataset. (((((tropical japonica, temperate japonica), *O. rufipogon*), *O. nivara*), *O.*

Topo	Proportion	Number of outliers	Candidate introgression
<i>barthii</i>			
A	0.141	11	Null
M	0.146	23	<i>O. rufipogon</i> <-> <i>O. nivara</i>
J	0.061	3	<i>japonica</i> <-> <i>O. nivara</i>
B	0.106	1	tropical japonica <-> <i>O. rufipogon</i>
C	0.094	0	temperate japonica <-> <i>O. rufipogon</i>
N	0.084	0	<i>O. nivara</i> -> temperate japonica
O	0.116	85	<i>O. nivara</i> -> tropical japonica
D	0.048	2	
G	0.032	0	
E	0.024	0	
F	0.031	0	
H	0.037	0	
I	0.023	0	
K	0.037	0	tropical japonica - > <i>O. nivara</i>
L	0.021	0	temperate japonica - > <i>O. nivara</i>

(((tropical japonica, temperate japonica), *O. rufipogon*), *indica*), *O. barthii*)

Topo	Proportion	Number of outliers	Candidate introgression
A	0.125	6	Null
M	0.138	20	<i>O. rufipogon</i> <-> <i>indica</i>
J	0.062	3	<i>japonica</i> <-> <i>indica</i>
B	0.098	3	tropical japonica <-> <i>O. rufipogon</i>
C	0.085	0	temperate japonica <-> <i>O. rufipogon</i>
N	0.088	0	<i>indica</i> -> temperate japonica
O	0.121	94	<i>indica</i> -> tropical japonica
D	0.047	3	
G	0.032	0	
E	0.030	0	
F	0.031	0	
H	0.043	0	
I	0.023	0	
K	0.048	0	tropical japonica - > <i>indica</i>
L	0.029	0	temperate japonica - > <i>indica</i>