Journal (2050), Vol. 00, No. 000, pp. 1–8
DOI: 10.1000/xxx/xxx000

Inferring the direction of between-species gene flow using genomic sequence data

Yuttapong Thawornwattana (orcid: 0000-0003-2745-163X)[1,†], Jun Huang (orcid: 0000-0002-4196-9729)[2,†], Tomáš Flouri (orcid: 0000-0002-8474-9507)[3], James Mallet (orcid: 0000-0002-3370-0367)[1], and Ziheng Yang (orcid: 0000-0003-3351-7981) [3,*]
[1]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA
[2]School of Biomedical Engineering, Capital Medical University, Beijing, 100069, P.R. China
[3]Department of Genetics, Evolution and Environment, University College London, London, WC1E 6BT, UK
[†]Those authors contributed equally to this work.

Received on xxxx, revised on xxxx, accepted on xxxx

Genomic data are informative about the history of species divergence and interspecific gene flow, including the direction, timing, and strength of gene flow. Nevertheless, gene flow in opposite directions generates similar patterns in multilocus sequence data, such as reduced sequence divergence between the hybridizing species, and as a result, inference of the direction of gene flow is challenging. Here we study the amount of information about the direction of gene flow in data of multilocus sequence alignments, when the data are analyzed using likelihood-based methods under the multi-species coalescent with introgression (MSci) model. We analyze the case of two species, and use simulation to examine larger species trees. We found that it is easier to infer gene flow from a small population to a large one than in the opposite direction, and easier to infer inflow (gene flow from outgroup species to an ingroup species) than outflow (gene flow from an ingroup species to an outgroup species). If introgression is assumed to occur in the wrong direction, the time of introgression tends to be correctly estimated, Bayesian test of gene flow is often significant, and the estimated introgression probability may be even greater than the true rate. We discuss factors that cause gene flow to be asymmetrical, including geography, behavior, and incompatibility of introgressed alleles with the host genomic background. We analyze a dataset of *Heliconius* butterflies to demonstrate that typical genomic datasets are informative for inferring the direction of interspecific gene flow, as well as its timing and strength.

BPP | direction of gene flow | Gene flow | introgression | multispecies coalescent

Introduction

Gene flow between species is an important evolutionary process that can facilitate species diversification and adaptation (Arnold and Kunte, 2017; Campbell *et al.*, 2018; Feurtey and Stukenbrock, 2018; Marques *et al.*, 2019; Edelman and Mallet, 2021). It occurs as a result of hybridization followed by backcrossing with one of the two hybridizing parental species, with one parent as the source or donor and the other the target or recipient. The outcome of introgression in each direction is influenced by multiple factors including mating preference and reproductive barriers, ecological selection on hybrids, and incompatibility of the introgressed alleles with the host genomic background (Coyne and Orr, 2004; Peters *et al.*, 2017; Martin and Jiggins, 2017; Moran *et al.*, 2021). Gene flow is thus intrinsically asymmetrical, being more likely in one direction than in the other. Reliable inference of the direction of introgression, as well as its timing and rate, using the abundant genomic data will advance our understanding of this important evolutionary process and its consequences, including the role of gene flow during the speciation process, and the adaptive nature of the introgressed alleles.

Two types of models of interspecific gene flow have been developed, both in the multispecies coalescent (MSC) framework. The MSC-with-introgression (MSci; Flouri *et al.*, 2020) model, also known as multispecies network coalescent (MSNC; Yu *et al.*, 2012; Wen and Nakhleh, 2018; Zhang *et al.*, 2018), assumes that gene flow occurs at a particular time point in the past. The rate of gene flow is measure by the introgression probability (φ or γ), which is the proportion of immigrants in the recipient population at the time of introgression. The MSC-with-migration (MSC-M) model, also known as the isolation-with-migration (IM) model, assumes that gene flow occurs continuously at a certain rate every generation after species divergence (Nielsen and Wakeley, 2001; Hey *et al.*, 2018). The rate of gene flow is measured by the expected proportion (m_{AB}) or number (M_{AB}) of immigrants in every generation, with $m_{AB} = N_B m_{AB}$,

*Correspondence: ythawornwattana@g.harvard.edu, z.yang@ucl.ac.uk

where N_B is the population size of the recipient population. In both models, the rate of gene flow (φ or M) should be considered an 'effective' rate, as it reflects the combined effects of gene flow, genetic drift, and natural selection purging or fixing introgressed alleles, influenced by the local recombination rate.

Interspecific gene flow alters gene genealogies, causing fluctuations over the genome in the genealogical history of sequences sampled from the extant species. Under both the MSC-M and MSci models, the gene tree and coalescent times have probabilistic distributions specified by the model and parameters, including species divergence times, population sizes for extant and extinct species, and the rate of gene flow (see Yang, 2014; Jiao *et al.*, 2021 for reviews). Data of multilocus sequence alignments, which are informative about the gene tree topologies and coalescent times that underlie the sequence data, are thus informative about the direction of gene flow as well as its timing and strength. Here in this paper we study the nature of the inference problem and focus on likelihood methods for inferring gene flow under the MSci model (Flouri *et al.*, 2020). Commonly used heuristic methods for inferring gene flow such as the D statistic (Green *et al.*, 2010; Durand *et al.*, 2011) may not be informative about the direction of gene flow as they do not use all information in the data. The computational strengths and statistical deficiencies of heuristic methods have been reviewed by Hibbins and Hahn (2021), Jiao *et al.* (2021), Huang *et al.* (2022), and Yang and Flouri (2022). We assume that multiple sequences are sampled per species per locus, but will also consider the special case of only one sequence per species. Sampling only one sequence per species leads to unidentifiability or severe reduction in information about gene flow (Degnan, 2018; Yang and Flouri, 2022).

We note that opposite directions of gene flow often create similar features in gene genealogies and thus in the sequence data. In the case of two species (say A and B) with one sequence sampled per species per locus, the coalescent time (t_{ab}) between the two sequences (a, b) has the same distribution under the models with $A \to B$ or $B \to A$ introgression, so that the direction of introgression is unidentifiable using such data (Yang and Flouri, 2022, fig. 10) (see also below). However, the introgression direction is identifiable when multiple sequences are sampled per species per locus. When gene flow occurs between non-sister species, the opposite directions of gene flow affect both gene tree topologies and coalescent times, so that it is easier to infer introgression between non-sister species than between sister species.

We use a combination of mathematical analysis and computer simulation to study the features of the sequence data that are informative about the direction of gene flow. The analysis allows us to compare and quantify the amount of information in the data under different scenarios. A second objective is to examine the impact of misspecified direction of gene flow on Bayesian estimation of parameters in the MSci model (Flouri *et al.*, 2020) and on Bayesian test of introgression (Ji *et al.*, 2022). If introgression occurs from species A to B but is assumed to occur from B to A in analysis of genomic sequence data, is gene flow likely to be inferred? What will be the estimated rate of introgression, relative to the true rate in the opposite direction? In a previous study, we examined the impacts of incorrect assignments of introgression events onto branches on the species tree, of ghost species which exchange migrants but are not included in the data sample, and of the misspecified mode of gene flow (e.g., gene flow occurs continuously over time as in the MSC-M model but assumed to occur at a fixed time point as in the MSci model) (Huang *et al.*, 2022). In this paper we focus on the direction of gene flow. We study the distribution of the coalescent times (t_{aa}, t_{ab}, t_{bb}) under models of introgression between two species A and B. Analysis of the two-species case provides important insights into more complex cases. Next, we explore the amount of information gained when a third species is added to branches of the species tree for two species. Finally, we study the impact of introgression direction when gene flow involves non-sister species. Our results provide practical guidelines for inferring introgression from genomic sequence data. We analyze two datasets (one coding and another noncoding) from three species of *Heliconius* butterflies to demonstrate the feasibility of using genomic sequence data to infer the direction of gene flow, as well as its timing and strength.

Results

The MSci model and its parameters in the case of two species

We use the MSci model of figure 1**a** with introgression from species A to B to introduce the notation and set up the problem. The model assumes that the two species diverged at time τ_R and came into contact and hybridized at time τ_X. The rate of introgrssion or hybridization is measured by the introgression probability φ_Y, which is the proportion of immigrants from A in population B at the time of introgression/hybridization. There are three types of parameters in the model: species divergence times or introgression times (τ_R, τ_X), population sizes for extant and extinct species $(\theta_A, \theta_B, \theta_X, \theta_Y, \theta_R)$, and the introgression probability (φ_Y). The divergence time parameter is defined as $\tau = T\mu$, where T is the divergence time in generations and μ is the mutation rate per site per generation. Each branch on the species tree represents a species or population and is associated with a population size parameter, $\theta = 4N\mu$, where N is the (effective) population size of the species. A branch on the species tree is also referred to by

its daughter node so that branch RX is also branch X, with population size parameter θ_X. Both τ and θ are measured by the expected number of mutations per site; i.e., one time unit is the expected time to accumulate one mutation per site. At this time scale, coalescent occurs between any two sequences in a population of size θ as a Poisson process at the rate $\frac{2}{\theta}$.

We consider estimation of parameters in the MSci model using multilocus sequence data. Each dataset consists of sequence alignments at L loci, with n_A sequences from A and n_B sequences from B at each locus, and with n sites in each sequence. Underlying the sequences at each locus is a gene tree with branch lengths (coalescent times). The probability density for the gene tree with coalescent times is given by Yu *et al.* (2014). The data are analyzed under three introgression models: the I model with $A \to B$ introgression, the O model with $B \to A$ introgression, and the B model with bidirectional introgression ($A \leftrightarrows B$) (fig. 1**a**-**c**). The 'inflow' (I) and 'outflow' (O) labels are used here in anticipation of models involving more than two species to be analyzed later.

We are interested in the following questions. First, what features of sequence data are informative about the direction of introgression (i.e., about distinguishing the I and O models of fig. 1**a**&**b**)? Second, what are the biases in estimated introgression probability and introgression time if the introgression direction is misspecified (i.e., if data are generated under the I model and analyzed under the O model)? We analyze the distributions of the coalescent times as major features of the multilocus sequence alignments, and augment our mathematical analysis by computer simulation using the Bayesian program BPP (Flouri *et al.*, 2018, 2020).

Distribution of coalescent times under the two-species model

While likelihood methods under the MSci model average over the full distribution of the gene tree (G) and coalescent times ($\boldsymbol{t}$) for the sampled sequences at every locus, this distribution depends on the sampling configuration (n_A, n_B) and is too complex to analyze. Instead, we study the marginal distributions of the coalescent times between two sequences sampled from the same population (t_{aa}, t_{bb}) or from different populations (t_{ab}), which are analytically tractable.

Under model I ('inflow') with $A \to B$ introgression (fig. 1**a**), the probability density of the coalescent time between two sequences sampled from species A is

$$f_{\rm I}(t_{aa}) = \begin{cases} \frac{2}{\theta_A} \mathrm{e}^{-\frac{2}{\theta_A} t_{aa}}, & \text{if } 0 < t_{aa} < \tau_X, \\ \mathrm{e}^{-\frac{2}{\theta_A}\tau_X} \frac{2}{\theta_X} \mathrm{e}^{-\frac{2}{\theta_X}(t_{aa}-\tau_X)}, & \text{if } \tau_X < t_{aa} < \tau_R, \\ \mathrm{e}^{-\frac{2}{\theta_A}\tau_X} \mathrm{e}^{-\frac{2}{\theta_X}(\tau_R-\tau_X)} & \\ \quad \times \frac{2}{\theta_R} \mathrm{e}^{-\frac{2}{\theta_R}(t_{aa}-\tau_R)}, & \text{if } t_{aa} > \tau_R. \end{cases} \tag{1}$$

This is a function of $\tau_R, \tau_X, \theta_A, \theta_X, \theta_R$, independent of $\theta_B, \theta_Y, \varphi_Y$. From the viewpoint of the coalescent between two A sequences, there are demographic changes in population size with $\theta_A, \theta_X, \theta_R$, respectively, for the three time segments $(0, \tau_X)$, (τ_X, τ_R), and (τ_R, ∞). The density is plotted for four sets of parameter values in figure 2.

The coalescent time between two sequences sampled from species B has the distribution

$$f_{\rm I}(t_{bb}) = \begin{cases} \frac{2}{\theta_B} \mathrm{e}^{-\frac{2}{\theta_B} t_{bb}}, & \text{if } 0 < t_{bb} < \tau_X, \\ \mathrm{e}^{-\frac{2}{\theta_B}\tau_X} \Big[(1-\varphi_Y)^2 \frac{2}{\theta_Y} \mathrm{e}^{-\frac{2}{\theta_Y}(t_{bb}-\tau_X)} & \\ \quad + \varphi_Y^2 \frac{2}{\theta_X} \mathrm{e}^{-\frac{2}{\theta_X}(t_{bb}-\tau_X)} \Big], & \text{if } \tau_X < t_{bb} < \tau_R, \\ \mathrm{e}^{-\frac{2}{\theta_B}\tau_X} \Big[(1-\varphi_Y)^2 \mathrm{e}^{-\frac{2}{\theta_Y}(\tau_R-\tau_X)} & \\ \quad + \varphi_Y^2 \mathrm{e}^{-\frac{2}{\theta_X}(\tau_R-\tau_X)} + 2\varphi_Y(1-\varphi_Y) \Big] & \\ \quad \times \frac{2}{\theta_R} \mathrm{e}^{-\frac{2}{\theta_R}(t_{bb}-\tau_R)}, & \text{if } t_{bb} > \tau_R. \end{cases} \tag{2}$$

This is a function of $\tau_R, \tau_X, \theta_B, \theta_X, \theta_Y, \theta_R, \varphi_Y$, independent of θ_A. The density is plotted in figure 2. In the time interval $(0, \tau_X)$, coalescent between the two B sequences occurs at the rate $2/\theta_B$, as when there is no gene flow. There is a suppression of coalescent events during the interval (τ_X, τ_R), due to introgression, as no coalescent is possible if one of the two B sequences migrates into X (with time running backwards).

The coalescent time between two sequences sampled from A and B has the distribution

$$f_{\rm I}(t_{ab}) = \begin{cases} \varphi_Y \frac{2}{\theta_X} \mathrm{e}^{-\frac{2}{\theta_X}(t_{ab}-\tau_X)}, & \text{if } \tau_X < t_{ab} < \tau_R, \\ \Big[(1-\varphi_Y) + \varphi_Y \mathrm{e}^{-\frac{2}{\theta_X}(\tau_R-\tau_X)} \Big] & \\ \quad \times \frac{2}{\theta_R} \mathrm{e}^{-\frac{2}{\theta_R}(t_{ab}-\tau_R)}, & \text{if } t_{ab} > \tau_R. \end{cases} \tag{3}$$

This is a function of $\tau_R, \tau_X, \theta_X, \theta_R$, and φ_Y, independent of $\theta_A, \theta_B, \theta_Y$. Note that coalescent between a and b is impossible in the interval $(0, \tau_X)$, while over the interval (τ_X, τ_R) it may occur if the B sequence migrates into A (with time running backwards).

Under model O with $B \to A$ introgression (fig. 1**b**), the densities $f_{\rm O}(t_{aa}), f_{\rm O}(t_{ab})$, and $f_{\rm O}(t_{bb})$ are similarly derived, and are indeed given by $f_{\rm I}(t_{bb})$, $f_{\rm I}(t_{ab})$, and $f_{\rm I}(t_{aa})$ with a change of symbols.

In particular, $f_{\rm I}(t_{ab}) = f_{\rm O}(t_{ab})$ for all $t_{ab} > 0$, with the parameter mapping $\tau_R^{(\rm O)} = \tau_R^{(\rm I)}, \tau_X^{(\rm O)} = \tau_X^{(\rm I)}, \theta_Y^{(\rm O)} = \theta_X^{(\rm I)}$, $\theta_R^{(\rm O)} = \theta_R^{(\rm I)}$ and $\varphi_X = \varphi_Y$ (Yang and Flouri, 2022, fig. 10). Here the superscripts indicate the

assumed model. Thus if only one sequence is sampled per species per locus in the data, models I and O are unidentifiable, and the direction of introgression cannot be inferred. However, if the direction of introgression is specified (given model I, say), the introgression probability (φ_Y, say) is identifiable using data of one sequence per species per locus.

Furthermore, the direction of introgression is identifiable if multiple sequences are sampled from A and B. Information for distinguishing models I and O of figure 1 are mostly from the coalescent times between sequences from the same species (t_{aa}, t_{bb}). If gene flow is from $A \to B$, the coalescent time for sequences from the donor species, t_{aa}, is not affected by the $A \to B$ introgression. If different populations on the species tree have the same size ($\theta_A = \theta_X = \theta_R$), t_{aa} will have a smooth exponential distribution (e.g., fig. 2**a**&**b**). Otherwise the distribution is discontinuous at time points τ_X and τ_R, as in a demographic model of population size change (e.g., fig. 2**c**&**d**). In contrast, the coalescent time for sequences from the recipient species, t_{bb}, will have a mixture distribution, depending on which parental hybridizing species each of the two B sequences is traced back to on the gene genealogy. The $A \to B$ and $B \to A$ models thus make different predictions about the coalescent times t_{aa} and t_{bb}, allowing the models to be identified using genomic sequence data, which contain information about gene trees and coalescent times.

Best-fitting parameters in the two-species model

We consider multilocus datasets generated under the I model with $A \to B$ introgression (fig. 1a) and analyzed under both the true I model and the misspecified O model with $B \to A$ introgression. When the amount of data (the number of loci) $L \to \infty$, the maximum likelihood estimates (MLEs) under the I model will converge to the true parameter values, with $\hat{\Theta}_{\rm I} \to \Theta_{\rm I}$. Under the O model, when $L \to \infty$ the estimates will converge to the infinite-data limit, $\hat{\Theta}_{\rm O} \to \Theta^*_{\rm O}$. Known as the *best-fitting* or *pseudo-true* parameter values, $\Theta^*_{\rm O}$ minimizes the Kullback–Leibler (KL) divergence from the true model to the fitting model (e.g., Yang and Zhu, 2018). It does not seem possible to calculate $\Theta^*_{\rm O}$ analytically with arbitrary data configurations. Instead we use as substitute the averages of posterior means in BPP analysis of simulated large datasets of $L =$ 4000 loci (with $S = 4$ sequences per species per locus and $n = 500$ sites in the sequence), shown in table S1. Note that at this datasize, the average estimates under the true I model are extremely close to the true values (table S1), suggesting that the data size may be large enough for the average estimates to be close to the infinite-data limits. We aim to understand the estimates $\Theta^*_{\rm O}$ by comparing the true distributions of the coalescent times under model I, $f_{\rm I}(t_{aa}), f_{\rm I}(t_{ab})$, and $f_{\rm I}(t_{bb})$ (eqs. 1–3), with the fitted distributions $f_{\rm O}(t_{aa}), f_{\rm O}(t_{ab})$, and $f_{\rm O}(t_{bb})$, calculated using $\Theta^*_{\rm O}$. In effect, the former are data, while the latter are the best fit to data achieved by the misspecified model O.

We used four sets of parameter values in model I (fig. 1**a**) in the numerical calculation, representing four different scenarios: (a) same θ tall tree, (b) same θ short tree, (c) small to large, and (d) large to small (fig. 2, table S1). See Methods section.

First, we consider the case where all populations on the species tree have the same size (table S1**a**&**b**). The true distribution $f_{\rm I}(t_{aa})$ is a continuous exponential density (fig. 2, black curves in cases **a** and **b**). The true distribution $f_{\rm I}(t_{bb})$ is discontinuous at τ_X and τ_R, with a drop in the probability mass during the time interval (τ_X, τ_R); the $A \to B$ introgression makes coalescence between the two B sequences less likely over (τ_X, τ_R). The true distribution $f_{\rm I}(t_{bb})$ is discontinuous at τ_R, and is exponential within each of the two time segments: (τ_X, τ_R) and (τ_R, ∞). In the fitting O model, the assumed $B \to A$ introgression should lead to the opposite expected pattern, i.e., discontinuities in $f_{\rm O}(t_{aa})$ but not in $f_{\rm O}(t_{bb})$. The introgression time $\hat{\tau}^{(\rm O)}_X$ is largely determined by the smallest coalescent time between sequences from the two species (t_{ab}), while the discontinuity in $f_{\rm I}(t_{ab})$ as well as in $f_{\rm I}(t_{bb})$ should be very informative about $\tau^{(\rm O)}_R$. Thus we expect those parameters to be close to the true values under model I: $\hat{\tau}^{(\rm O)}_R \approx \tau^{(\rm I)}_R$ and $\hat{\tau}^{(\rm O)}_X \approx \tau^{(\rm I)}_X$. If the sequence is infinitely long with $n = \infty$ (i.e., if the true coalescent times are given as data), the best fitting parameter values for $\tau^{(\rm O)}_R$ and $\tau^{(\rm O)}_X$ should match the true values in model I (Huang *et al.*, 2022). Population sizes $\hat{\theta}^{(\rm O)}_A, \hat{\theta}^{(\rm O)}_B$, and $\hat{\theta}^{(\rm O)}_R$ should be close to the true values as well, as those are well-estimated from the multiple samples from the same species. Here we focus on $\varphi_X, \theta^{(\rm O)}_X$, and $\theta^{(\rm O)}_Y$.

Consider t_{aa}. In the true model, both A sequences will enter X and may coalesce in (τ_X, τ_R). In the fitting O model, the two A sequences may be separated due to introgression (one in X and one in Y), so they may not coalesce in (τ_X, τ_R) as often. Thus we expect $\hat{\theta}^{(\rm O)}_X \le \theta^{(\rm I)}_X$ as an increased coalescent rate in X may compensate for this deficit of coalescence over (τ_X, τ_R) in model O. Next, consider t_{bb}. In the true model, $A \to B$ introgression reduces the chance of coalescence between sequences from B during (τ_X, τ_R). In the fitting model, both B sequences will enter Y, leading to a higher chance of coalescence during (τ_X, τ_R). Thus we expect $\theta^{(\rm O)}_Y \ge \theta^{(\rm I)}_Y$ to reduce the chance of coalescence in (τ_X, τ_R). Finally, consider t_{ab}. By matching the amount of coalescence between sequences a and b over the time interval (τ_X, τ_R), or by matching the probability densities $f_{\rm I}(t_{ab})$ and $f_{\rm O}(t_{ab})$ for $\tau_X < t_{ab} < \tau_R$, we have approximately

$$\hat{\varphi}_X \left[1 - {\rm e}^{-2\Delta\tau/\hat{\theta}^{(\rm O)}_Y}\right] = \varphi_Y \left[1 - {\rm e}^{-2\Delta\tau/\theta^{(\rm I)}_X}\right], \qquad (4)$$

where $\Delta\tau = \tau_R - \tau_X$ is assumed to be the same under models I and O based on the arguments above. Eq. 4 predicts that more gene flow will be inferred under model O ($\hat{\varphi}_X > \varphi_Y$) when $\hat{\theta}_Y^{(O)} > \theta_X^{(I)}$, while $\hat{\varphi}_X < \varphi_Y$ if $\hat{\theta}_Y^{(O)} < \theta_X^{(I)}$. Note that $\hat{\theta}_Y^{(O)} > \theta_X^{(I)}$ means that the coalescent rate between sequences a and b during (τ_X, τ_Y) is lower in the fitting model than in the true model, so that a higher introgression rate $\hat{\varphi}_X$ than the true rate φ_Y will increase the chance of such coalescence and achieve a better fit to $f_I(t_{ab})$.

In summary, if the population sizes are the same for all species on the species tree in the true model, consideration of t_{aa} predicts $\hat{\theta}_X^{(O)} < \theta_X^{(I)}$, consideration of t_{bb} predicts $\hat{\theta}_Y^{(O)} > \theta_Y^{(I)}$, while consideration of t_{ab} predicts that $\hat{\varphi}_X$ is greater or smaller than φ_Y depending as $\hat{\theta}_Y^{(O)}$ is greater or smaller that $\theta_X^{(I)}$. The actual estimates of $\varphi_X, \theta_X^{(O)}$, and $\theta_Y^{(O)}$ will depend on the parameter values under the true model. Those predictions match the estimates of table S1 (cases **a** and **b**). Using the estimates of $\hat{\theta}_Y^{(O)}$, eq. 4 predicts $\hat{\varphi}_X$ to be 0.31 and 0.35 for cases **a** and **b**, respectively, compared with $\varphi_X^* = 0.27$ and 0.30 in table S1. The approximation is reasonably good.

The case where different species have different population sizes is more complex (table S1**c**&**d**). As in the case of equal population sizes, τ_X is largely determined by the smallest coalescent time t_{ab}, while τ_R is determined by the discontinuities in the distribution of t_{aa}, t_{ab}, t_{bb}. Both should be correctly estimated despite the misspecification of introgression direction in model O. Similarly population sizes for the extant species (θ_A, θ_B) are informed by the multiple samples from those species and thus correctly estimated, as is θ_R for the root population. However, estimates of $\varphi_X, \theta_X^{(O)}$, and $\theta_Y^{(O)}$ depend on all of t_{aa}, t_{bb}, and t_{ab}, which may make conflicting predictions about the parameter estimates. For example, eq. 4 based on t_{ab} predicts $\hat{\varphi}_X$ to be 0.44 and 0.22 for cases **c** and **d**, respectively, compared with $\varphi_X^* = 0.98$ and 0.17 in table S1. The approximations are poor, especially for case **c**, suggesting that φ_X^* is influenced not only by t_{ab} but also by t_{aa} and t_{bb}. We discuss the parameter estimates in model O in those cases later when we describe the simulation results.

The limiting parameter values determine the behavior of the Bayesian test of gene flow (Ji *et al.*, 2022) in large datasets, as they determine the asymptotic behavior of posterior probabilities of the compared models (the null model of no introgression and the alternative model of introgression) (Yang and Zhu, 2018). If data are simulated under model I (with $\varphi_Y > 0$) and analyzed under model I, the posterior probability for the true model I should approach 1, the Bayes factor for comparing model I against model 0 (no gene flow, fig. 1**d**) $B_{I0} \to \infty$, and the power of the test should approach 100%, when the data size $L \to \infty$ (Yang and Zhu, 2018). If the data are simulated under model I and analyzed under model B, the power for testing φ_Y (which has the true value $\varphi_Y > 0$) should approach 100%, and the false positive rate for testing φ_X (which has the true value $\varphi_X = 0$) should approach 0, when the data size $L \to \infty$.

If the data are generated under model I and analyzed under model O, with the introgression direction misspecified, both the null model (0: no gene flow) and the alternative model (O: $B \to A$ introgression) are wrong. Then the asymptotic behavior of the Bayesian test depends on whether or not the pseudo-true parameter value $\varphi_X^* = 0$. As $\varphi_X^* > 0$ according to our analysis, model O is a 'less wrong' model than model 0 (Yang and Zhu, 2018). Thus when $L \to \infty$, $B_{O0} \to \infty$, and the probability of rejecting $H_0 : \varphi_X = 0$ will approach 100%. In this comparison of models O against 0, the biological interpretation of test results is somewhat ambiguous. If one emphasizes the fact that model O allows gene flow while model 0 does not, rejecting the null and detecting gene flow may be considered a correct result. However, if one considers model O as a wrong model with misspecified introgression direction, rejecting the null and accepting model O may be considered a false positive error. In this paper, we use the second interpretation.

It may be noteworthy that the Bayesian test based on Bayes factors may lead to a strong support for the null model, rejecting the more general alternative model with great force. When two models fit the data equally well, Bayesian test favors the simpler model with fewer parameters. This is in contrast to Frequentist hypothesis testing, which may fail to reject the null but never provides strong support for the null. Those expectations are confirmed later in our analyses of the simulated and real data.

Simulation under the two-species models

To verify and extend our theoretical analysis, we simulated datasets under model I of figure 1**a** using four sets of parameters (cases **a**–**d**), with $\varphi_Y = 0.2$ in all cases (table S1). Each dataset consists of $L = 250, 1000$ or 4000 loci, with $S = 4$ sequences sampled per species per locus and $N = 500$ sites in the sequence. The data were then analyzed under models I, O, and B (fig. 1**a**-**c**) using BPP. The posterior means and the 95% highest-probability-density (HPD) credibility intervals (CIs) are plotted in figure 3 (see also table S1 for $L = 4000$). The results of Bayesian test of introgression are in figure S1.

Performance under the true model

Model I is the true model, so that the performance under this model constitutes the best-case scenario. Indeed all parameters were well estimated, with the posterior means approaching the true values and the

CI width approaching 0 when the amount of data $L \to \infty$ (fig. 3 and table S1, cases **a**-**d**, model I). Population sizes for extant species (θ_A, θ_B) were much better estimated than those for ancestral species (θ_X, θ_Y). The divergence times (τ_R, τ_X) were well estimated as well. The introgression probability (φ_Y) had substantial uncertainties with wide CIs but with $L = 4000$ loci in the data, the estimates were fairly precise. Thousands of loci appeared to be necessary to get reliable and precise estimates of the introgression probability. The results are similar to those found in our early simulation examining the impact of various factors (L, S, n) on the information content concerning parameters in the MSci model (Huang *et al.*, 2020).

Model B allows bidirectional introgression and thus is a correct model although it is over-parametrized with an extra introgression probability parameter (φ_X). When the amount of data increases, $\hat{\varphi}_Y$ should converge to the true value while $\hat{\varphi}_X$ to 0. Estimates of other parameters were very similar to those under the I model, and the CI widths under models I and B were very similar. In particular, φ_Y was estimated with very similar precision in the two models. In the large datasets of $L = 4000$ loci, the average CI width was 0.07, 0.12, 0.08, and 0.16 for cases **a**–**d** under model I, compared with 0.07, 0.12, 0.09, 0.17 under model B. Even in the small or intermediate datasets with $L = 250$ or 1000 loci, the CIs for φ_Y were similar between the two models. Over-parametrization did not seem to have incurred any major cost to the statistical performance of the method. This result may appear surprising, as given the difficulty of inferring the direction of introgression, one might expect the assumed nonexistent introgression in the $B \to A$ direction in model B to interfere with the estimation of the rate in the right direction ($A \to B$), so that $\hat{\varphi}_{A\to B}$ may be expected to have much larger variances under model B than under model I. Nevertheless, the information concerning $\varphi_{A\to B}$ is mostly determined by the number of sequences reaching the hybridization node Y and by the difficulty with which one can tell the parental path taken by each B sequence at Y. There is thus little difference in information content about $\varphi_{A\to B}$ between models I and B. Computationally, model B is much more expensive due to sampling an extra parameter in the Markov chain Monte Carlo (MCMC) algorithm and to MCMC mixing issues (Yang and Flouri, 2022).

Information content for estimating φ_Y under model I

Next, we discuss the comparison of model I results among the four cases, to highlight the impact of multiple factors on the amount of information in the data for estimating φ_Y.

Consider tracing the genealogical history of sequences at a locus backwards in time. When sequences from B reach the hybridization node Y in the species tree (fig. 1a), there is in effect a binomial sampling process, with each sequence taking the horizontal (introgression) parental path with probability φ_Y and the vertical parental path with $1 - \varphi_Y$. However the outcome of the sampling process (i.e., the parental path taken by the sequence) is not directly observed and is instead reflected in the gene tree and coalescent times, which are in turn reflected in the observed mutations in the sequences. If a sequence from B coalesces with an A sequence during the time interval (τ_X, τ_R), it will be clear that the B sequence has taken the introgression parental path. Thus the amount of information for estimating φ_Y at any locus is affected by (i) the number of B sequences reaching Y and (ii) how easy it is to tell the parental path taken by each B sequence at Y. The number of B sequences reaching Y is given by the number of B sequences sampled at the locus (n_B) minus the number of coalescent events among them in B before reaching Y, the latter of which is affected by the length of branch B measured in coalescent units: $2\tau_Y/\theta_B$. Each locus contains a greater amount of information for a larger n_B or smaller $2\tau_Y/\theta_B$. Also increasing the number of sampled sequences (n_B) is less effective than increasing the number of sequences reaching node Y, which is in turn less effective than increasing the number of loci (L).

The second factor concerns the probability that two sequences entering X coalesce in X before reaching R; there is more information about φ_Y in the data the longer the internal branch RX is or the smaller the population size θ_X is. This may be illustrated by considering the special case where the data consists of one sequence per species per locus and where the true coalescent time (t_{ab}) is available at each locus. Then the information content for estimating φ_Y may be measured by the Fisher information, given by

$$\begin{aligned} I_{\mathrm{I},t_{ab}}(\varphi_Y) &\approx \mathbf{E}\Big[-\tfrac{\partial^2}{\partial \varphi_Y^2} \log f_{\mathrm{I}}(t_{ab})\Big] \\ &= \tfrac{P_X}{\varphi_Y(1-\varphi_Y P_X)} < \tfrac{1}{\varphi_Y(1-\varphi_Y)}, \end{aligned} \tag{5}$$

where the expectation is with respect to t_{ab} (eq. 3), and where $P_X = 1 - \mathrm{e}^{-\frac{2}{\theta_X}(\tau_R - \tau_X)}$ is the probability that two sequences entering population X (sequences a, b) coalesce in X. Eq. 5 is approximate as it ignores the correlation between parameters and the Fisher information is in this case a 5×5 matrix (see eq. 3). The asymptotic variance of the estimate is

$$\mathbb{V}(\hat{\varphi}_Y) \approx \tfrac{1}{IL} = \tfrac{\varphi_Y(1-\varphi_Y P_X)}{LP_X} \geq \tfrac{\varphi_Y(1-\varphi_Y)}{L}, \tag{6}$$

the binomial variance, with equality held if $P_X = 1$. There is thus more information for estimating φ_Y if P_X is greater, or if the branch length $\frac{2}{\theta_X}(\tau_R - \tau_X)$ is greater. Indeed eq. 6 suggests that increasing P_X is more effective in reducing $\mathbb{V}(\hat{\varphi}_Y)$ than increasing the amount of data (L) by the same factor.

In our simulation (fig. 3), the introgression probability φ_Y was better estimated in case **a** "same θ tall tree" than in case **b** "same θ short tree". For $L = 4000$, the 95% HPD CI width was 0.07 for case **a**, and 0.12 for case **b**. We suggest that two effects may explain the difference. First, in case **a**, branch YB is longer, with length $2\tau_Y/\theta_B$ in coalescent units, so that fewer sequences reach Y, providing less information about φ_Y than in case **b**. Indeed, starting with $n_B = 4$ sequences sampled from B, the probability that 1, 2, 3, 4 sequences remain by time τ_Y is 0.388, 0.515, 0.095, 0.002, respectively in case **a**, with on average 1.71 sequences reaching Y. For the short tree of case **b**, the corresponding probabilities are 0.122, 0.481, 0.347, 0.050, with on average 2.32 B sequences reaching Y. The average number of sequences reaching Y differ by a factor 1.36. Second, in case **a** the species tree is tall, so that any B sequence reaching Y and taking the left parental path is more likely to coalesce with A sequences in X than in case **b**. $P_X = 1 - e^{-1} = 0.632$ in case **a** while $P_X = 1 - e^{-0.5} = 0.393$ in case **b**, different by a factor of 1.61. As mentioned earlier, increasing P_X is more effective than increasing the number of loci L (eq. 6, which is in turn more effective than increasing the average number of sequences reaching Y. Thus φ_Y was far more precisely estimated in case **a** than in **b** (table 3, table S1). If we use the same population size θ_B in cases **a**&**b**, the number of sequences reaching Y will be the same, and the performance differences between the two cases will be even greater.

Introgression probability φ_Y was much better estimated with narrower CIs in case **c** "small to large" than in case **d** "large to small" (fig. 3). At $L = 4000$, the CI width was 0.08 for case **c** compared with 0.16 for case **d** (table S1). In case **c** more B sequences reach Y because of the large θ_B than in (d). Furthermore B sequences that reach Y and migrate into X (with time running backwards) have a high chance of coalescing with other sequences in population X. Both effects make it easier to estimate φ_Y in case **c** than in case **d**.

Biases in parameters under model O with misspecified introgression direction

As discussed above, we expect species divergence time τ_R, the introgression time τ_X, and population sizes for the extant species and for the root $(\theta_A, \theta_B, \theta_R)$ to be well estimated under model O despite the misspecification of the introgresion direction. Here we focus on estimation of parameters φ_X, θ_X, θ_Y under model O (fig. 3, I-O in cases **a**–**d**).

In case **a** (same θ tall tree), all populations have the same size (fig. 3, case **a**). As the true model I predicts a smooth density for t_{aa} while model O may have a deficit of t_{aa} over the interval (τ_X, τ_R) due to the assumed introgression, we expect $\hat{\theta}_X^{(O)} \leq \theta_X^{(I)}$ to compensate by increasing the coalescent rate for sequences from A in X. Similarly, based on the distribution of t_{bb}, we expect $\hat{\theta}_Y^{(O)} \geq \theta_Y^{(I)}$. Given $\hat{\theta}_Y^{(O)} \geq \theta_X^{(I)} = \theta_Y^{(I)}$, we expect from the distribution of t_{ab} that $\hat{\varphi}_X \geq \varphi_Y$. Because the coalescent rate between sequences a and b during (τ_X, τ_Y) is higher in the true model than in the fitting model, a higher introgression rate φ_X than the true rate φ_Y helps with the fit to the distribution of t_{ab}.

Case **b** (same θ short tree) is similar to case **a**, but the divergence times (τ_R, τ_X) were half smaller. As in case **a**, we expect $\hat{\theta}_X^{(O)} < \theta_X^{(I)}$, $\hat{\theta}_Y^{(O)} > \theta_Y^{(I)}$ and $\hat{\varphi}_X > \varphi_Y$. Furthermore, we expect $\hat{\varphi}_X$ to be larger in case **b** than in case **a**. Note that when $\theta_Y^{(O)}$ and $\theta_X^{(I)}$ are fixed with $\theta_Y^{(O)} > \theta_X^{(I)}$ (or when $\hat{\theta}_Y^{(O)}$ is similar in the two cases, table S1), the smaller $\Delta\tau$ of case **b** (than in case **a**) means a larger $\hat{\varphi}_X$ according to eq. 4. We have $\varphi_X^* \approx 0.27$ and 0.30 for cases **a** and **b** respectively (table S1).

Case **c** (small to large) assumes that populations on the left of the species tree (fig. 1**a**) are $1/5$ as small as those on the right, with $\theta_A = \theta_X = \theta_R = 0.01$ and $\theta_B = \theta_Y = 0.002$. Based on t_{ab}, we need a large $\hat{\varphi}_X$ or a small $\hat{\theta}_Y^{(O)}$ to increase the chance of coalescence between sequences a and b over (τ_X, τ_R) to mimic the effect of the smaller source population in the true model. However, consideration of t_{bb} suggests $\hat{\theta}_Y^{(O)} > \theta_Y^{(I)}$ to compensate for the reduced coalescence between B sequences caused by the $A \to B$ introgression in the true model. Thus t_{ab} and t_{bb} make conflicting predictions about $\hat{\theta}_Y^{(O)}$. In the simulation, $\hat{\theta}_Y^{(O)}$ was close to $\theta_Y^{(I)}$. Then $\hat{\varphi}_X$ needs to be large to obtain a good fit to t_{ab}. However, $\hat{\varphi}_X$ was nearly 100%, much larger than $\varphi_Y = 0.2$ in the true model, and consequently, $\hat{\theta}_X^{(O)}$ was poorly estimated (fig. 3). The extreme estimate of φ_X, with $\varphi_X/\theta_Y^{(O)} \approx \varphi_Y/\theta_X^{(I)}$, may be explained by considering t_{ab}. In the true I model (small to large), many B sequences reach Y (as θ_B is large) and when a B sequence takes the introgression parental path, it has a high chance of coalescing with an A sequence in population RX (as θ_X is small), resulting in an excess of coalescence during $\tau_X < t_{ab} < \tau_R$. In the fitting O model, few A sequences reach node X (as θ_A is small), and A sequences reaching X and taking the introgression parental path at node X may coalesce slowly with B sequences in population Y if θ_Y is large. Thus having a very large φ may increase coalescence during (τ_X, τ_R) and may improve the model fit to the data. The extreme estimate ($\hat{\varphi}_X \approx 100\%$) also caused small biases in the divergence time τ_R and the introgression time τ_X.

Case **d** (large to small) is the opposite to case **c** and assumes that population sizes on the left of the species tree, $\theta_A = \theta_X = \theta_R$ (fig. 1**a**) are five times as large as those on the right ($\theta_B = \theta_Y$), and introgression is from a large population to a small one. We expect $\hat{\theta}_X^{(O)} < \theta_X^{(I)}$ based on t_{aa}, and $\hat{\theta}_Y^{(O)} > \theta_Y^{(I)}$ based on t_{bb},

as before. Moreover, the larger source population in the true model means t_{ab} is less common in (τ_X, τ_R), with most coalescence occurring in the common ancestor R. Thus based on t_{ab} we predict a larger $\hat{\theta}_Y^{(O)}$ or a smaller φ_X to reduce the amount of coalescence in (τ_X, τ_R) in the fitting model. Thus considerations of both t_{bb} and t_{ab} suggest $\hat{\theta}_Y^{(O)} > \theta_Y^{(I)}$. Depending on whether $\hat{\theta}_Y^{(O)}$ is smaller or greater than $\theta_X^{(I)}$, the introgression rate $\hat{\varphi}_X$ may be greater or smaller than the true rate φ_Y, according to eq. 4. In our setting, $\hat{\theta}_Y^{(O)} = 0.0107$, slightly greater than $\theta_X^{(I)} = 0.01$, and $\hat{\varphi}_X = 0.17$, slightly smaller than $\varphi_Y = 0.2$ (table S1).

In summary, from the distributions of coalescent times between sequences expected under the true and fitting models (fig. 3), we can predict or explain the patterns of parameter estimates in the fitting model. When introgression is assumed in the wrong direction, the species divergence time and introgression time (τ_R, τ_X) are often correctly estimated, except in case **c** where model O has a extreme estimate with $\varphi_X \approx 100\%$, which affected the estimation of τ_R and τ_X as well. Population sizes for extant species and for the ancestral species unrelated to the introgression event on the species tree $(\theta_A, \theta_B, \theta_R)$ are well estimated as well.

Bayesian test of introgression

We applied the Bayesian test of introgression of Ji *et al.* (2022) to the data simulated under model I of figure 1**a**. The MCMC samples generated in the BPP runs of figures 3 were processed to calculate the Bayes factor B_{10} in favor of the introgression model (H_1) against the null model of no gene flow (H_0) via the Savage-Dickey density ratio (see Methods). The results are summarized in figure S1.

The power of the Bayesian test for introgression when the data were simulated and analyzed under model I was high, reaching $\sim 100\%$ at $L = 1000$ loci. When the data were analyzed under model O, the direction of introgression was misspecified, in which case detection of gene flow was considered a false positive error. The false positive rate was comparable to the power in the analysis under model I. When the data were analyzed under model B, power to detect the $A \to B$ introgression was only very slightly lower than under model I, also reaching $\sim 100\%$ at $L = 1000$, while the false positive rate for detecting the non-existent $B \to A$ introgression was low, below the nominal 1%.

Added information from including a third species

Given two species (A, B) with introgression from $A \to B$ at the rate of φ (fig. 1**a**), we consider the information gain for estimating φ from including a third species (C). There are five branches on the two-species tree onto which C can be attached, creating five scenarios: (**a**) the root population, (**b**, **c**) the source and target populations before gene flow, and (**d**, **e**) the source and target populations after gene flow (fig. 4**a**–**e**). Case **c** is one of 'inflow', with gene flow from the outgroup species (A) into one of the ingroup species (B), while **b** represents 'outflow', with gene flow from an ingroup species (A) into the outgroup (B). Note that in all cases the correct MSci model is used in the analysis, so that the Bayesian estimate (posterior mean) of φ will converge to the true value (which is 0.2). However, the information content may differ among the five cases. We assumed the same population size $\theta_1 = 0.01$ for all populations, but examined the impact of different population sizes in cases **b** and **c**. We simulated 100 replicate datesets in each case. The posterior means, the posterior standard deviation (SD), and the width of the HPD CI for φ are summarized in figure 4**f**-**h**. While the average of the estimate ($\hat{\varphi}$) was close to the true value in all cases, the posterior SD was smaller and the posterior CI was narrower if the data were more informative about φ. The 95% CIs for other parameters are shown in figure S2.

Equal populations sizes on the species tree

If all populations on the species tree have the same size (θ), we expect the amount of information for estimating φ to be in the order **a** $\prec$ **d** $\prec$ (**b**, **e**) $\prec$ **c**, with the order of **b** and **e** undecided (fig. 4**f**–**h**).

First, **a** $\prec$ **d**. Cases **a** and **d** are the least informative. Adding an outgroup species C in case **a** adds very little information about φ. In **d**, the C sequences may reach node X and coalesce with a B sequence in RX, providing information about whether sequences from B take the introgression parental path at node Y. Thus we expect more information in the data in **d** than in **a**.

Next, **d** $\prec$ **b**. The number of B sequences reaching node Y is the same in the two cases, so the only difference is in the difficulty of inferring the parental path taken by B sequences at Y. In case **b**, coalescence of a sequence b with one from A causes a change to gene tree topology. In case **d**, introgression does not cause such topological change to the gene tree. The information content may thus be higher in **b** than in **d**.

Next, **d** $\prec$ **e**. In case **e**, sequences from both B and C may reach the hybridization node Y while in **d** only sequences from B may reach Y. The sample size for binomial sampling at Y is thus larger in **e** than in **d**. In **d**, more sequences enter node X, increasing the probability of coalescence for any B sequence that take the introgression parental path at Y, but this effect may be less important than that of increased sample size in **e**.

Next, **b** $\prec$ **c** or it is easier to infer inflow than outflow. In both cases, the same number of B sequences reach node Y so that the sample size for the binomial sampling at Y is the same. However, the two cases

differ in the difficulty with which one can tell the parental path taken by each B sequence at Y. In **c**, the coalescence of a sequence b with a sequence a over the time period (τ_X, τ_R) causes a change to the gene tree topology. In **b**, the coalescence of a sequence b with an a over (τ_X, τ_S) causes a gene tree topology change but the resulting gene tree has a shorter internal branch and is harder to infer than in case **c**. In addition, if sequence b does not coalesce with an a over (τ_X, τ_S) and enters population S, it may coalesce with both sequences a and c and may not lead to a topology change. Both effects suggest that it is easier to resolve the parental path taken by each B sequence at Y in case **c** than in **b**, and the data are more informative about φ_Y in case **c** than in **b**. It is easy to infer inflow than outflow.

Finally, **e** $\prec$ **c**. In case **c**, introgression leads to changes in gene tree topology which may be very informative about φ whereas in **e**, more sequences reach Y so that in effect the sample size for the binomial sampling is increased. In the simulation here, the increased sample size was less informative than the gene tree topology change (fig. 4**g**&**h**, case **c** same-θ vs. case **e**). Nevertheless, we note that in both **c** and **e**, the data are more informative about φ the smaller τ_X is, and that in **e**, the data are more informative the closer τ_S is to τ_Y.

Different populations sizes on the species tree

For cases **b** and **c**, we also examined the impact of different population sizes, with gene flow from a small population to a large one (small→large) or in the opposite direction (large→small) (fig. 4**f**–**h**). The patterns are more complex than in the case of equal population sizes.

First, in case **b** (outflow), φ is most poorly estimated in the large→small setting, much better estimated in the same-θ setting, and best in the small→large setting. Note that the same-θ setting is a case of large→large. In particular, case **b** large→small is much worse than case **b** same-θ. This seems to be mainly due to the different numbers of B sequences reaching Y or the sample size for the binomial sampling in the two cases. In the same-θ case, we used $\theta_1 = 0.01$ for all branches on the species tree, with $\tau_Y = \theta_1/2$. From Huang *et al.* (2022, fig. S4), the probability that all $S = 4$ sequences from B have coalesced with only one sequence reaching time τ_Y is $\mathbb{P}(t_{\text{MRCA}} < \tau_Y / \frac{\theta_B}{2}) = \mathbb{P}(t_{\text{MRCA}} < 1) = 0.387$, so there is a 61.3% chance that two or more B sequences reach Y. Note here that $\tau_Y / \frac{\theta_B}{2} = 1$ is the length of branch YB in coalescent units, or in other words the introgression time τ_Y is equal to the expected coalescent time for a pair of sequences from a population of size θ_B. In the large→small case, we have $\theta_B = \theta_0 = 0.002$, and $\mathbb{P}(t_{\text{MRCA}} < 5) = 0.988$, so at 98.8% of loci the four sequences from B should have coalesced before reaching Y with only one sequence reaching Y. This large difference in the sample size should explain why φ is far more poorly estimated in case **b** than in case **b** same-θ. Next, data are far more informative about φ in case **b** small→large than in **b** same-θ. In both cases, the number of B sequences reaching node Y is the same, but in **b** small→large, B sequences taking the introgression path at node Y coalesce at much higher rate with sequences from A in population SX and with sequences from A and C in population RS, making it much easier to tell the parental path taken by B sequences at node Y.

Similarly in case **c** (inflow), φ was more poorly estimated in the large→small and same-θ settings, and was best in the small→large setting. The differences among the three settings are much smaller than in case **b**. Case **c** large→small and large→large (same-θ) had similar performance, which may be explained by two factors having opposite effects. The first factor is that in **c** large→small, fewer B sequences reach Y on average leading to a smaller sample size for the binomial sampling. This favors **c** large→large (same-θ). The second factor is that in **c** large→small, the smaller sizes of populations SY and RS means that B sequences reaching Y and taking the vertical parental path coalesce faster. This favors **c** large→small. The two effects may have similar magnitude leading to similar performance between **c** large→small and **c** large→large (same-θ). Next, case **c** small→large is more informative than **c** large→large (same-θ), because in the small→large setting, B sequences taking the introgression path at node Y coalesce at much higher rate with sequences from A in population RX.

While in the case of same-θ, **b** outflow is less informative than **c** inflow, the order is reversed in the small→large setting, with **b** inflow small→large being more informative than **c** outflow small→large. The same number of B sequences reach node Y in the two cases, so the difference must be due to the different levels of difficulty by which one can tell the parental paths taken by B sequences at node Y. In **b**, B sequences taking the introgression parental path go through the small population SX and may coalesce at the high rate with sequences from A (which lead to changes to the gene tree topology informative about introgression), and with sequences from both A and C in population RS. Note that because species divergence and introgression times (τ_R, τ_X) are well estimated, between-species coalescent times younger than species divergence ($t_{bc} < \tau_R$) are informative about introgression as well. In **c**, B sequences taking the vertical parental path may coalesce in population RS with C sequences, but given that both populations SY and RS are large, this effect may be expected to be minor. While multiple factors have inconsistent effects on the relative information content concerning φ in cases **b** versus **c** small→large, on balance, the data are more informative in case **b** than in **c**.

Simulation in the case of four species

We simulated datasets under the three MSci models of figure 5 for four species on the species tree $((A,(B,C)),D)$, with introgression between non-sister species A and B in different directions: inflow (I), outflow (O), and bidirectional introgression (B). The data were analyzed under the same three models (I, O, B), resulting in nine combinations. We first consider the case where all populations on the species tree have the same size ($\theta_0 = \theta_1 = 0.01$) in the simulation model. The average posterior means and 95% HPD CIs are shown in figure 6 and the results for the large datasets of $L = 4000$ are summarized in table S2. We then assumed different population sizes, with $\theta_0 = 0.002$ for the thin branches and $\theta_1 = 0.01$ for the thick branches on the species trees of figure 5. The results are summarized in figure S3 and table S3.

Equal population sizes

First we consider the case where all species on the tree have the same population size (θ) (fig. 6, table S2). Some parameters that are shared among all three models were well estimated, with no discernible impact from the model misspecification. These include the population sizes for the extant species ($\theta_A, \theta_B, \theta_C, \theta_D$), population sizes for ancestral species R and S (θ_R, θ_S), and the species divergence and introgression times (τ_R, τ_S, τ_T, and $\tau_X = \tau_Y$). The correct estimation of the introgression time when the introgression direction is misspecified (e.g., τ_X in the I-O and O-I settings) is noteworthy. As discussed earlier for the case of two species, the estimate of introgression time ($\hat{\tau}_X$) is dominated by the minimum sequence divergence between the species involved in introgression (t_{ab}). Ancestral population sizes (θ_R, θ_S) were slightly less well estimated but appeared to converge to the correct values in all settings when the number of loci $L \to \infty$ (fig. 6). Below we focus on introgression probabilities (φ_X, φ_Y) and population sizes θ_X, θ_Y, and θ_T.

In the I-I, O-O, and B-B settings, the true model was used in the analysis, and the results provided a reference for comparison. The introgression probability φ_Y in the I-I setting was more precisely estimated than φ_X in the O-O setting, with narrower CIs. Inflow was easier to infer than outflow, as found earlier in our simulations for the three-species case (fig. 4**b**&**c**, same θ). Similarly in the B-B setting, the inflow rate φ_Y was better estimated than the outflow rate φ_X. The B-B setting had wider CIs for φ_X and φ_Y than in the I-I and O-O settings, because of more parameters in model B; however, the differences were very small (fig. 6, table S2). Overall the introgression probabilities were well estimated under all three settings, although thousands of loci appeared necessary to obtain precise estimates. Population size θ_X was better estimated in the I-I setting than in the O-O and B-B settings, as the estimation was affected by uncertainties in φ_X in model O and B. Similarly θ_Y was better estimated in the O-O setting than in the I-I and B-B settings.

In the I-B and O-B settings, gene flow occurred in one direction but the model assumed bidirectional gene flow. The model was over-parametrized but not misspecified. As Bayesian estimation under the correct model is consistent, the introgression probability for the nonexistent introgression (φ_X in I-B, φ_Y in O-B) should converge to 0 when the data size approaches ∞. Results in both the I-B and O-B settings are consistent with this expectation (fig. 6). Other parameters were well-estimated, with CI widths indistinguishable from those in the I-I and O-O settings. The results suggest that the over-parametrization in the bidirectional model had incurred little cost to the statistical performance of the method, as found for the case of two species (fig. 3).

In the I-O and O-I settings, introgression is assumed to occur in the wrong (opposite) direction. Our analysis of the two-species case predicts that this misspecification should only affect the estimation of the introgression probability as well as the population sizes θ_X, θ_Y, and θ_T, while other parameters should be largely unaffected (see fig. 3). In particular, the time of introgression should be correctly estimated. This was indeed the case (fig. 6). In the I-O setting, we expect θ_X to be underestimated, θ_Y and θ_T to be overestimated, and the introgression probability φ_X may be larger or smaller than φ_Y depending on how the estimates of θ_Y and θ_T are compared with the true value of θ_X. Simulation results confirmed these predictions (fig. 6). In the O-I setting, the effects were the opposite: θ_X was overestimated while θ_Y and θ_T were underestimated. The introgression probability φ_Y was estimated to be smaller than φ_X.

Finally, in the B-O and B-I settings, introgression occurs in both directions but is assume to occur in only one direction. The estimates of θ_X, θ_Y, and θ_T followed the same pattern of the I-O and O-I, respectively. The introgression probability (φ_X in B-O, φ_Y in B-I) was larger and less well-estimated than the case when the true model had unidirectional gene flow (φ_X in O-I, φ_Y in I-O). This positive bias may be explained by the fact that gene flow in the two directions in the true model B have an additive effects on the distribution of the coalescent time between species (t_{ab}). For instance, in the B-I setting, $B \to A$ introgressions in the true model B are expected to increase the chance of coalescence during $\tau_X < t_{ab} < \tau_S$, and such introgression events may be recognized and misinterpreted as $A \to B$ introgression in the fitting model I, leading to $\hat{\varphi}_Y^{(\mathrm{I})} > \varphi_Y^{(\mathrm{B})}$.

Different population sizes

Next, we consider the case where the source and target populations differ in size (fig. 5). Data were simulated under models I, O, and B, and analyzed under I, O, and B, resulting in nine settings (fig. S3, table S3). In the simulation, model I assumes inflow from a small population to a large one, model O assumes outflow from a large population to a small one, while model B assumes both inflow from a small population to a large one and outflow from a large population to a small one (fig. 5).

As in the case of equal population sizes, species divergence and introgression times (τ_R, τ_S, τ_T, and $\tau_X = \tau_Y$) and population sizes for extant species ($\theta_A, \theta_B, \theta_C, \theta_D$) and common ancestors R and S (θ_R, θ_S) are all well estimated, irrespective of possible model misspecification. Thus we focus on introgression probabilities (φ_X, φ_Y) and population sizes θ_X, θ_Y, and θ_T.

In the I-I, O-O, and B-B settings, the correct model was assumed in data analysis. The introgression rate φ_Y in model I was far more precisely estimated than φ_X in model O. At $L = 4000$ loci, the average 95% CI was 0.19-0.22 for I-I and 0.16-0.24 for O-O (table S3). The difference was far greater than in the case of equal population sizes where the inflow rate φ_Y in model I was slightly better estimated than the outflow rateφ_X in model O (table S2). The large performance difference is mainly because it is easier to estimate the rate of small $\rightarrow$ large introgression (in model I) than in the opposite direction. Similarly in the B-B setting, the inflow rate of small $\rightarrow$ large introgression (φ_Y) is much better estimated than the outflow rate of large $\rightarrow$ small introgression (φ_X): for $L = 4000$, the average 95% CIs were 0.18-0.23 for φ_Y and 0.15-0.24 for φ_X (table S3).

In the I-B and O-B settings, the B model was over-parametrized. Performance was very similar to the I-I and O-O settings, respectively, with the rate for the nonexistent migration approaching 0 when the increase in data size (L) (fig. S3, table S3).

In the I-O and O-I settings, the direction of introgression was misspecified. In the I-O setting, $\hat{\varphi}_X$ was much greater than in the case of equal population sizes. The extremely large $\hat{\varphi}_X$ was noted in our analysis of the two-species case: when gene flow occurs from a small population to a large one but is assumed to occur in the opposite direction, a large introgression rate is preferred to improve the fit to the between-species coalescent time (t_{ab}) (fig. 3**c** small$\rightarrow$large, model O).

In the O-I setting, gene flow was from a large population into a small one, the donor population size θ_Y was grossly underestimated while the recipient population size θ_X was overestimated when the introgression direction was misspecified. The patterns were the same as in the two-species analysis (fig. 3**d** large$\rightarrow$small, model O). The large $\hat{\theta}_X$ helped improve the fit to t_{ab} by reducing the coalescent rate in X in the fitting model to match the reduced coalescence in Y and T because of gene flow in the true model. The estimate $\hat{\varphi}_Y$ was much lower than the true introgression probability $\varphi_X = 0.2$ in the opposite direction.

The B-I and B-O settings showed a cumulative effect in the estimates of the migration rate: $\hat{\varphi}_Y$ was greater in the B-I setting than in the I-I setting, and $\hat{\varphi}_X$ was greater in the B-O setting than in the O-O setting. This is the same pattern as found in the case of equal population sizes (fig. 6).

Bayesian test of introgression

We also applied the Bayesian test of introgression of Ji *et al.* (2022) to the data simulated under the I, O, and B models of figure 5 and analyzed in figures 6&S3. The results are summarized in figures S4&S5. When the correct model is assumed to analyze data (as in the I-I, O-O, and B-O settings), the power of the test was high, reaching $\sim 100\%$ at $L = 1000$ loci (figs. S4&S5). In the I-B and O-B settings, the power detecting introgression that exists in the true B model was high, while the false positive rate for detecting the non-existent introgression was low, below the nominal 1%. In the I-O and O-I settings, the direction of introgression was misspecified, in which case we consider detection of gene flow as a false positive error. The false positive rate was very high, comparable to the power in the analysis under the correct model. Overall, the results are very similar to those for the two-species simulation (fig. S1).

Analysis of Heliconius genomic datasets to infer the direction of introgression

To demonstrate the feasibility of inferring the direction of gene flow using genomic sequence data, we analyzed two datasets of noncoding and coding loci from three species of *Heliconius* butterflies: *H. hecale* (H), *H. cydno* (C), and *H. melpomene* (M) (fig. 7). One dataset consisted of 5341 noncoding loci, the other of 4942 coding loci. We fitted four models: (0) MSC with no gene flow, (I) MSci with $C \rightarrow M$ introgression, (O) MSci with $M \rightarrow C$ introgression, and (B) MSci with $C \leftrightarrows M$ bidirectional introgression. We ran the MCMC algorithm in BPP to generate the posterior estimates of parameters in each model (Flouri *et al.*, 2020) (table 1) and conducted Bayesian test of introgression (Ji *et al.*, 2022) (table 2). To compare the four different models, we calculated Bayes factors using two approaches: thermodynamic integration with Gaussian quadrature (Lartillot and Philippe, 2006; Rannala and Yang, 2017) and Savage-Dickey density ratio (Ji *et al.*, 2022); see Materials and Methods section.

First, we discuss the results of the Bayesian test (table 2). The precise value of the Bayes factor for the same test showed differences depending on the quadrature points in the thermodynamic integration

approach, and was often ∞ by the Savage-Dickey density ratio, reflecting the challenges of calculating reliably the marginal likelihood or Bayes factors in very large datasets (Rannala and Yang, 2017). For example the logarithm of B_{I0} for comparison of the I model with $C \rightarrow M$ introgression against the 0 model of no gene flow was 1087.1 and 1082.5, respectively, when $K = 32$ and 64 quadrature points were used. This difference is mainly due to the difficulty of calculating the power posterior reliably rather than the use of too few quadrature points (Rannala and Yang, 2017). Nevertheless, both values were far greater than the cutoff of 100. Similarly the Savage-Dickey density ratio approach estimated B_{I0} to be ∞ at all three threshold values ($\varepsilon = 1\%, 0.1\%, 0.01\%$). Both approaches thus strongly support model I with $C \rightarrow M$ introgression and reject model 0 with no gene flow.

For both datasets, the two approaches to Bayes factor calculation led to the same conclusions, as did the three threshold values for the Savage-Dickey density ratio ($\varepsilon = 1\%, 0.1\%, 0.01\%$). The null hypothesis $\varphi_{C\rightarrow M} = 0$ was rejected in the I-0 and B-O comparisons, with strong support for the $C \rightarrow M$ introgression, whether or not the $M \rightarrow C$ introgression was accommodated in the null and alternative hypotheses.

The B-I comparison tests the null hypothesis $\varphi_{M\rightarrow C} = 0$ when both the null and alternative models accommodate the $C \rightarrow M$ introgression. This test was not significant at rejecting the null model I. In fact, the test led to strong support for the null model I, with $B_{\mathrm{BI}} < 0.01$. Unlike Frequentist hypothesis testing, which never supports the null strongly, here the Bayesian test strongly favored the null model I, rejecting the more general alternative model B. In other words, under the assumption that there is $C \rightarrow M$ introgression, the test strongly supported the absence of $M \rightarrow C$ introgression.

However, test of $\varphi_{M\rightarrow C} = 0$ was significant when the $C \rightarrow M$ introgression was not accommodated in the null and alternative models (the O-0 comparison). The result mimics our mathematical analysis and computer simulation, which showed that the test of gene flow was often significant if the assumed gene flow was in the wrong direction.

Finally, models I and O are not nested, but the Bayes factor can be used to compared them. B_{IO} suggested strong preference for model I ($C \rightarrow M$ gene flow) over model O ($M \rightarrow C$ gene flow).

In summary, all the tests led to the same conclusions. Both the noncoding and coding datasets strongly supported the presence of *H. cydno* $\rightarrow$ *H. melpomene* introgression, and both strongly supported the absence of the *H. melpomene* $\rightarrow$ *H. cydno* introgression.

Next, we consider the estimates of introgression probabilities. Consistent with the results of Bayesian testing above, estimates of φ under model B with bidirectional introgression suggested that gene flow was unidirectional. The estimates for the noncoding data were $\hat{\varphi}_{C\rightarrow M} = 0.28$ (95% HPD CI: 0.25–0.31) and $\hat{\varphi}_{M\rightarrow C} < 1\%$ in the opposite direction, while for the coding data, they were $\hat{\varphi}_{C\rightarrow M} = 0.51$ (95% HPD CI: 0.47–0.54) and $\hat{\varphi}_{M\rightarrow C} < 1\%$ (table 1). The reasons for the higher rate ($\hat{\varphi}_{C\rightarrow M}$) for the coding than the noncoding data are unknown. One possible factor may be that introgression was mostly adaptive, driven by natural selection and coding loci are under stronger selection. The time of introgression was nearly zero, suggesting that gene flow may be on-going. Estimates under model I were nearly identical to those under model B. In model O where only $M \rightarrow C$ gene flow was allowed, the introgression probability was estimated to be $\hat{\varphi}_{M\rightarrow C} = 0.17$ (0.15,0.20) for the noncoding data, and 0.14 (0.08, 0.20) for the coding data. Those rates were substantial, consistent with the significant test results when model O was compared against model 0. Even though gene flow appeared to be unidirectional from C to M, assuming introgression in the opposite (and presumably wrong) direction led to high estimates of the rates and significance test results. Those results mimics our findings in the simulations (figs. S1, S4 & S5). The misspecified introgression direction in model O caused large estimates of θ_s and reduced τ_s. Those results are consistent with the behavior of the misspecified model in the large$\rightarrow$small case in our theoretical analysis and simulation of the two-species case (fig. 3, table S1**d** large$\rightarrow$small).

Finally we note that the divergence time between *H. cydno* and *H. melpomene* (τ_s) was estimated to be much smaller, and θ_S was much larger under model 0 (no gene flow) than under models I or B. This is because ignoring gene flow when it occurs leads to underestimation of the divergence time between species.

Discussion

Asymmetry of gene flow in nature

No systematic studies have examined the frequency of unidirectional versus bidirectional gene flow given that two species are involved in introgression or hybridization. Both scenarios appear to be common. Sometimes gene flow occurs in one direction even though opportunities exist for gene flow in the opposite direction. A well-documented example is gene flow in the *Anopheles gambiae* group of mosquitoes in sub-Saharan Africa. Analysis of genomic data provides strong evidence for gene flow from *A. arabiensis* to *A. gambiae* or its sister species *A. coluzzii*, while the rate of gene flow in the opposite direction was estimated to be 0 (Thawornwattana *et al.*, 2018; Flouri *et al.*, 2020). This result from comparisons of genomic sequences is consistent with crossing experiments which supported introgression of autosomal regions from *A. arabiensis* into *A. coluzzii* but not in the

opposite direction (della Torre *et al.*, 1997; Slotman *et al.*, 2005b). One possible explanation is genetic incompatibilities of the X chromosome from one species in an autosomal background of another species (Slotman *et al.*, 2004, 2005a). The introgression is thought to facilitate the range expansion of *A. gambiae* and *A. coluzzii* into more arid savanna habitats of *A. arabiensis* (Coluzzi *et al.*, 1979; Ayala and Coluzzi, 2005).

Note that the rate of gene flow in the MSci model estimated from the genomic sequence data is an 'effective' rate, and reflects the combined effects of gene flow, natural selection, and genetic drift. Most introgressed alleles may be expected to be purged in the recipient population because they were not compatible with the host genomic background. It appears highly unlikely that the chance of acceptance of introgressed alleles from species *A* in the genomic background of species *B* should be the same as that of introgressed alleles from *B* in the background of *A*. This reasoning appears to suggest that by norm one should expect gene flow to be asymmetrical, with different rates in the two directions.

Gene flow in Heliconius butterflies

Heliconius cydno and *H. melpomene* are broadly sympatric across Central America and northwestern South America, and are known to hybridize regularly in the wild (Mallet *et al.*, 2007). Our analysis supports recent unidirectional gene flow from *H. cydno* into *H. melpomene* (fig. 7, tables 1–2). In Panama, *H. cydno chioneus* and *H. melpomene rosina* are broadly sympatric. When they hybridize, male F_1 hybrids are fertile while female F_1 hybrids are sterile; male hybrids backcross readily in both directions (Naisbit *et al.*, 2001, 2002).

Previous studies used several different approaches to estimate gene flow between these two species. Early phylogenetic analyses of multilocus data attributed recent gene flow between *H. cydno chioneus* and *H. melpomene rosina* as a cause of gene tree variation among loci (Beltrán *et al.*, 2002). An isolation-with-migration (IM) analysis (Hey and Nielsen, 2004) using a small number of loci yielded an estimated symmetric bidirectional migration rate m between the two species of 1.7×10^{-6} (95% CI 1.0–45 $\times 10^{-6}$) per generation, with *H. cydno chioneus* having a larger population size (Bull *et al.*, 2006). An IM model allowing for different migration rates in each direction found evidence for unidirectional gene flow from *H. cydno* and *H. melpomene*, with $2N_M m_{C\to M} = 0.294$ (90% HPD interval: 0.116–0.737) whereas $2N_C m_{M\to C} = 0.000$ (0.000, 0.454) (Kronforst *et al.*, 2006), consistent with our results. Similar patterns were obtained in a subsequent IMa2 analysis (Hey, 2010) of a larger dataset (Kronforst *et al.*, 2013). In a more recent analysis of genome-scale data, Martin *et al.* (2015) estimated a symmetric bidirectional migration rate between *H. c. chioneus* and *H. m. rosina* to be $M = Nm = 0.20$ (90% HPD interval: 0.09–0.40) per generation. Lohse *et al.* (2016) compared three models: complete isolation after divergence, and two IM models with unidirectional gene flow. Consistent with our results, the preferred model was the IM with gene flow from *H. cydno* into *H. melpomene rosina*, with estimated migration rate $4Nm = 1.5$, and with *H. cydno* having a larger population size than *H. melpomene* and the ancestral population (which were assumed to have the same size). Other genomic analyses relied on summary statistics of sequence data such as D and f_d statistics or gene tree topologies across the genome to detect gene flow (Martin *et al.*, 2013; Martin and Van Belleghem, 2017; Martin *et al.*, 2019). In Martin *et al.* (2019), frequencies of gene trees across the genome were used to suggest that there has been extensive gene flow from *H. cydno* into *H. melpomene* in Panama.

In summary, our results are consistent with previous studies, although our analyses provided estimates of a number of important population parameters including the species divergence times and population sizes, as well as the rate of introgression (table 1).

Inferring the direction of gene flow using genomic data

If introgression is assumed to occur in the wrong (opposite) direction, the estimated introgression rate will typically be nonzero, and may even be higher than the true introgression rate. If one tests for introgression and assumes the wrong direction, for example, by using the Bayesian test of Ji *et al.* (2022), the test may have high power and be often significant. Thus neither a high estimate of the introgression rate nor a significant test of introgression is reliable evidence that introgression occurred in the specified direction. If one considers the presence or absence of gene flow but ignore the direction, this result may be considered good power for the test. However, if one insists on correct inference of the direction of introgression, both the large estimates of the introgression rate and the significant test results for the non-existing introgression event may be surprising and disturbing.

If there is uncertainty concerning the direction of gene flow, application of the bidirectional model may be a feasible option, although this adds computational cost. Our results in this paper suggest that if gene flow is truly unidirectional, the over-parametrization of the bidirectional model appears to incur very little cost in statistical performance, as the posterior CIs and the power to detect gene flow under the bidirectional model are very similar to those under the true unidirectional model.

Materials and Methods

Asymptotic analysis and simulation in the case of two species

We examined the distributions of coalescent times and conducted computer simulations in the case of two species (A, B) with introgression from A to B. This is model I of figure 1**a**. We used four sets of parameter values.

(a) same θ tall tree: all populations have the same size with $\theta = 0.01$. The other parameters are $\tau_R = \theta, \tau_X = 0.5\theta$, and $\varphi_Y = 0.2$.
(b) same θ short tree: $\theta = 0.01$ for all populations, $\tau_R = 0.5\theta, \tau_X = 0.25\theta$, and $\varphi_Y = 0.2$.
(c) small to large: different species on the species tree have different population sizes, with $\theta_A = \theta_X = \theta_R = \theta_0 = 0.002$ on the left of the tree and $\theta_B = \theta_Y = \theta_1 = 0.01$ for other branches (fig. 1). Other parameters are $\tau_R = 3\theta_0, \tau_X = 1.5\theta_0$ and $\varphi_Y = 0.2$.
(d) large to small: This is the same to case (**c**) except that $\theta_B = \theta_Y = \theta_0 = 0.002$ while all other populations have $\theta_1 = 0.01$.

We simulated multilocus sequence datasets under model I (fig. 1**a**) and analyzed them under models I, O, and B (fig. 1**a**–**c**). Each replicate dataset consists of $L = 250, 1000$ or 4000 loci, with $S = 4$ sequences sampled per species per locus. The sequence length is $N = 500$ sites. The `simulate` option of BPP (Flouri *et al.*, 2018) was used to simulate gene trees with coalescent times and to 'evolve' sequences along the gene tree under the JC mutation model (Jukes and Cantor, 1969). Sequences at the tips of the gene tree constitute the data. The number of replicates is 100.

Each replicate dataset is then analyzed using BPP (Flouri *et al.*, 2018, 2020) under models I, O, and B of figure 1**a**–**c**. This is the so-called A00 analysis, with the model fixed (Yang, 2015). The JC model was assumed in the analysis. Gamma priors are assigned to the age of the root of the species tree (τ_R) and to population size parameters (θ), with the shape parameter $\alpha = 2$ so that the prior is diffuse and with the rate parameter β chosen so that the prior mean is close to the true values. We used $\tau_R \sim$ G(2, 200) and $\theta \sim$ G(2, 200) for case **a** "same θ tall tree"; $\tau_R \sim$ G(2, 400) and $\theta \sim$ G(2, 200) for case **b** "same θ short tree"; $\tau_R \sim$ G(2, 400) and $\theta \sim$ G(2, 400) for case **c** "small to large" and **d** "large to small". Introgression probability φ was assigned the beta prior beta$(1, 1)$, which is $\mathbb{U}(0, 1)$.

MCMC settings were chosen by performing pilot runs, with MCMC convergence assessed by verifying consistency between replicate runs for the same analysis. The same setting is then used to analyze all replicate datasets. We used 16,000 MCMC iterations as burnin, and then took 10^5 samples, sampling every 2 iterations. Running time for analyzing one replicate dataset is $\sim$ 45mins for $L = 250$ loci or $\sim$ 3hrs for $L = 1000$ using one thread, and $\sim$ 12hrs for $L = 4000$ using two threads.

Simulation to evaluate the gain in information for estimating the introgression rate by adding a third species

Given the introgression model for two species (A, B) of figure 1**a**, with $A \to B$ introgression, we add a third species (C) and assess the gain in information for estimating φ. There are five branches on the two-species tree, resulting in five cases: (**a**) the root population, (**b**, **c**) the source and target populations before gene flow, and (**d**, **e**) the source and target populations after gene flow (fig. 4**a**–**e**). Note that case **c** represents 'inflow', with gene flow from the outgroup species (A) into one of the ingroup species (B), while **b** represents 'outflow'. The true introgression rate was $\varphi = 0.2$. The original two-species tree had the divergence time $\tau_R = \theta_1$ and the introgression time $\tau_X = \theta_1/2$. In cases **b**–**e**, species C was attached to the midpoint of the target branch, while in **a**, the new root was 1.25× as old as the old root. For models **a**, **d** & **e**, all populations on the species tree had the same size, with $\theta_1 = 0.01$. For cases **b** and **c**, three scenarios are considered: equal population size, with $\theta_1 = 0.01$ for all populations; from small to large, with $\theta_A = \theta_X = \theta_S = \theta_0 = 0.002$ for the thin branches in case **b** and $\theta_A = \theta_X = \theta_0 = 0.002$ in case **c** and with $\theta_1 = 0.01$ for all other branches; and from large to small, with $\theta_B = \theta_Y = \theta_0 = 0.002$ in case **b** and $\theta_B = \theta_Y = \theta_S = \theta_0 = 0.002$ in case **c** and with $\theta_1 = 0.01$ for all other branches.

For each parameter setting, we simulated 100 replicate datesets. Each dataset consisted of $L = 1000$ loci, with $S = 4$ sequences per species per locus and $N = 500$ sites in the sequence. Each dataset was analyzed using BPP to estimate the parameters in the MSci model (fig. 4**a**–**e**). Gamma priors are assigned to τ_R and θ: $\tau_R \sim$ G(2, 200) and $\theta \sim$ G(2, 200), while $\varphi_{A\to B} \sim \mathbb{U}(0, 1)$. We used 32,000 MCMC iterations as burnin, and then took 10^6 samples, sampling every 10 iterations. Running time for analyzing one dataset using one thread is $\sim$30 hrs.

Simulation in the case of four species: inflow versus outflow

We simulated data under the three MSci models (I, O, B) of figure 5**a**-**c**, with introgression between non-sister species A and B on a four-specie tree $((A, (B, C)), D)$. The three models differ in the assumed direction of gene flow, with I for inflow from A to B, O for outflow from B to A, and B for bidirectional introgression between A and B. We used two sets of parameter values. In the first set (same-θ), all species on the tree had the same population size, with $\theta_0 = \theta_1 = 0.01$. In the second set (different-θ), the thin branches had $\theta_0 = 0.002$ while the thick branches

had $\theta_1 = 0.01$ (fig. 5**a**-**c**). Other parameters were the same in the two settings, with $\tau_R = 4\theta_0$, $\tau_S = 3\theta_0$, $\tau_T = 2\theta_0$, and $\tau_X = \tau_Y = 1.5\theta_0$, and the introgression probabilities are $\varphi_X = \varphi_Y = 0.2$.

Each dataset consisted of $L = 250, 1000$, or 4000 loci, with $S = 4$ sequences per species per locus and with $N = 500$ sites in the sequence. The number of replicates was 100. With three MSci models (I, O, B), two population-size settings (same-θ vs. different-θ), and three data sizes (L), a total of $3 \times 2 \times 3 \times 100 = 1800$ datasets were generated. Each dataset was analyzed under the three models (I, O, B), with the direction of introgression either correctly or incorrectly specified. Gamma priors were assigned to τ_R and θ: $\tau_R \sim$ G(2, 200) and $\theta \sim$ G(2, 400), while $\varphi \sim \mathbb{U}(0,1)$. We used 32,000 MCMC iterations as burnin, and took 2×10^5 samples, sampling every 5 iterations. Running time for analyzing one dataset was ∼12hrs for small datasets of $L = 250$ loci and 60hrs for $L = 1000$ using one thread, and ∼120hrs for $L = 4000$ using two threads.

Analysis of the Heliconius butterfly dataset

We fitted MSci models with different introgression directions to two genomic datasets (one noncoding and another coding) from three species of *Heliconius* butterflies: *H. hecale* (*H*), *H. cydno* (*C*), and *H. melpomene* (*M*). Raw genomic sequencing data from Edelman *et al.* (2019) were processed to compile one dataset of 5341 noncoding loci and another of 4942 coding loci from chromosome 1, following the procedure of Thawornwattana *et al.* (2022). Each locus consisted of three unphased diploid sequences, with one sequence from each species. The heterozygote phase in the diploid sequence was resolved using an analytical integration algorithm in the likelihood calculation in BPP (Gronau *et al.*, 2011; Flouri *et al.*, 2018; Huang *et al.*, 2022). We fitted four models: (0) MSC with no gene flow, (I) $C \to M$ introgression, (O) $M \to C$ introgression, and (B) $C \leftrightarrows M$ bidirectional introgression.

We assigned priors $\tau_r \sim$ G(4, 200), $\theta \sim$ G(2, 200), and $\varphi \sim \mathbb{U}(0,1)$. We used 10^5 MCMC iterations for burnin, and recorded 10^4 samples, sampling every 100 iterations. For each model, we performed ten independent runs to confirm consistency between runs. The resulting MCMC samples were combined to produce final posterior estimates. Each run took ∼ 80hrs.

Bayesian test of introgression

We applied the Bayesian test of introgression to test whether there is significant evidence for introgression (Ji *et al.*, 2022). The test was applied to data for two species simulated under the models of figure 1**a**-**c**, the data for four species simulated under models I, O, and B of figure 5, and the *Heliconius* dataset for three species (fig. 7).

The Bayes factor for comparing two models H_0 and H_1 is defined as the ratio of the marginal likelihood values M_0 and M_1 for the two models: $B_{10} = \frac{M_1}{M_0}$. If the prior probabilities for the models are π_0 and π_1, the Bayes factor can be converted into posterior model probabilities, $\frac{\mathbb{P}(H_1|X)}{\mathbb{P}(H_0|X)} = \frac{\pi_1}{\pi_0} \cdot B_{10}$. If the prior probabilities are equal ($\pi_0 = \pi_1$), then $B_{10} = 100$ translates to the posterior probability $\mathbb{P}(H_1|X) \approx 1\%$. Thus $B_{10} > 100$ may be considered strong evidence in support of H_1 over H_0, while $B_{10} < 0.01$ is strong evidence in favor of H_0 over H_1.

In the test of introgression, the two models are nested, with $H_0 : \varphi = 0$ to be the null model of no gene flow and $H_1 : \varphi > 0$ to be the alternative model of introgression. Then B_{10} can be calculated using the Savage-Dickey density ratio (Dickey, 1971), by using an MCMC sample under H_1 (Ji *et al.*, 2022). Define an interval of null effects, $\emptyset : \varphi < \varepsilon$, inside which the introgression probability is so small that introgression may be considered nonexistent. The Bayes factor in favor of H_1 over H_0 is then

$$B_{10,\varepsilon} = \frac{\mathbb{P}(\emptyset)}{\mathbb{P}(\emptyset|X)}, \tag{7}$$

where $\mathbb{P}(\emptyset)$ is the prior probability of the null interval, while $\mathbb{P}(\emptyset|X)$ is the posterior probability, both calculated under H_1 (Ji *et al.*, 2022). Note that $\mathbb{P}(\emptyset) = \mathbb{P}(\varphi < \varepsilon) = \varepsilon$ if the prior is $\varphi \sim \mathbb{U}(0,1)$. When $\varepsilon \to 0$, $B_{10,\varepsilon} \to B_{10}$ (Ji *et al.*, 2022). We used a few values for ε in the range 0.01%–1% to assess its effect. This approach has a computational advantage as it requires running the MCMC under H_1 only and avoids trans-model MCMC algorithms and calculation of marginal likelihood values.

For the *Heliconius* datasets, we in addition used thermodynamic integration (TI) combined with Gaussian quadrature to calculate the marginal likelihood under each model, using 32 or 64 quadrature points (Lartillot and Philippe, 2006; Rannala and Yang, 2017). This approach applies even if the compared models are nonnested, and was used to conduct pairwise comparisons of all four models fitted to the *Heliconius*.

Acknowledgments

This study has been supported by Biotechnology and Biological Sciences Research Council grant (BB/T003502/1) and a BBSRC equipment grant (BB/R01356X/1) to Z.Y., and a Natural Science Foundation of China grant (32200490) to J.H.

References

Arnold, M. L. and Kunte, K. 2017. Adaptive genetic exchange: A tangled history of admixture and evolutionary innovation. *Trends Ecol. Evol.*, 32(8): 601–611.

Ayala, F. J. and Coluzzi, M. 2005. Chromosome speciation: Humans, *Drosophila*, and mosquitoes. *Proc. Natl. Acad. Sci. USA*, 102(suppl 1): 6535–6542.

Beltrán, M., Jiggins, C. D., Bull, V., Linares, M., Mallet, J., McMillan, W. O., and Bermingham, E. 2002. Phylogenetic discordance at the species boundary: Comparative gene genealogies among rapidly radiating *Heliconius* butterflies. *Mol. Biol. Evol.*, 19(12): 2176–2190.

Bull, V., Beltrán, M., Jiggins, C. D., McMillan, W. O., Bermingham, E., and Mallet, J. 2006. Polyphyly and gene flow between non-sibling *Heliconius* species. *BMC Biol.*, 4(1): 11.

Campbell, C. R., Poelstra, J. W., and Yoder, A. D. 2018. What is speciation genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. *Biol. J. Linn. Soc.*, 124(4): 561–583.

Coluzzi, M., Sabatini, A., Petrarca, V., and Di Deco, M. A. 1979. Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex. *Trans. R. Soc. Trop. Med. Hyg.*, 73(5): 483–497.

Coyne, J. A. and Orr, H. A. 2004. *Speciation*. Sinauer Assoc., Sunderland, Massachusetts.

Degnan, J. H. 2018. Modeling hybridization under the network multispecies coalescent. *Syst. Biol.*, 67(5): 786–799.

della Torre, A., Merzagora, L., Powell, J., and Coluzzi, M. 1997. Selective introgression of paracentric inversions between two sibling species of the *Anopheles gambiae* complex. *Genetics*, 146(1): 239–244.

Dickey, J. M. 1971. The weighted likelihood ratio, linear hypotheses on normal location parameters. *Ann. Math. Statist.*, 42(1): 204–223.

Durand, E. Y., Patterson, N., Reich, D., and Slatkin, M. 2011. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.*, 28: 2239–2252.

Edelman, N. B. and Mallet, J. 2021. Prevalence and adaptive impact of introgression. *Annu. Rev. Genet.*, 55(1): 265–283.

Edelman, N. B., Frandsen, P. B., Miyagi, M., Clavijo, B., and Davey, J. e. a. 2019. Genomic architecture and introgression shape a butterfly radiation. *Science*, 366(6465): 594–599.

Feurtey, A. and Stukenbrock, E. H. 2018. Interspecific gene exchange as a driver of adaptive evolution in fungi. *Annu. Rev. Microbiol.*, 72: 377–398.

Flouri, T., Jiao, X., Rannala, B., and Yang, Z. 2018. Species tree inference with BPP using genomic sequences and the multispecies coalescent. *Mol. Biol. Evol.*, 35(10): 2585–2593.

Flouri, T., Jiao, X., Rannala, B., and Yang, Z. 2020. A Bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis. *Mol. Biol. Evol.*, 37(4): 1211–1223.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., and Stenzel, U. e. a. 2010. A draft sequence of the Neandertal genome. *Science*, 328: 710–722.

Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G., and Siepel, A. 2011. Bayesian inference of ancient human demography from individual genome sequences. *Nature Genet.*, 43: 1031–1034.

Hey, J. 2010. Isolation with migration models for more than two populations. *Mol. Biol. Evol.*, 27: 905–920.

Hey, J. and Nielsen, R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, 167: 747–760.

Hey, J., Chung, Y., Sethuraman, A., Lachance, J., Tishkoff, S., Sousa, V. C., and Wang, Y. 2018. Phylogeny estimation by integration over isolation with migration models. *Mol. Biol. Evol.*, 35(11): 2805–2818.

Hibbins, M. S. and Hahn, M. W. 2021. Phylogenomic approaches to detecting and characterizing introgression. *Genetics*, page 10.1093/genetics/iyab173.

Huang, J., Flouri, T., and Yang, Z. 2020. A simulation study to examine the information content in phylogenomic datasets under the multispecies coalescent model. *Mol. Biol. Evol.*, 37(11): 3211–3224.

Huang, J., Thawornwattana, Y., Flouri, T., Mallet, J., and Yang, Z. 2022. Inference of cross-species gene flow under misspecified models. *Mol. Biol. Evol.*, page submitted.

Ji, J., Jackson, D. J., Leache, A. D., and Yang, Z. 2022. Power of Bayesian and heuristic tests to detect cross-species introgression with reference to gene flow in the *Tamias quadrivittatus* group of North American chipmunks. *BioRxiv*, page DOI: 10.1101/2021.12.07.471567.

Jiao, X., Flouri, T., and Yang, Z. 2021. Multispecies coalescent and its applications to infer species phylogenies and cross-species gene flow. *Nat. Sci. Rev.*, 8: nwab127 (DOI: 10.1093/nsr/nwab127).

Jukes, T. and Cantor, C. 1969. Evolution of protein molecules. In H. Munro, editor, *Mammalian Protein Metabolism*, pages 21–123. Academic Press, New York.

Kronforst, M. R., Young, L. G., Blume, L. M., and Gilbert, L. E. 2006. Multilocus analyses of admixture and introgression among hybridizing *Heliconius* butterflies. *Evolution*, 60(6): 1254–1268.

Kronforst, M. R., Hansen, M. E., Crawford, N. G., Gallant, J. R., Zhang, W., Kulathinal, R. J., Kapan, D. D., and Mullen, S. P. 2013. Hybridization reveals the evolving genomic architecture of speciation. *Cell Reports*, 5(3): 666 – 677.

Lartillot, N. and Philippe, H. 2006. Computing Bayes factors using thermodynamic integration. *Syst. Biol.*, 55: 195–207.

Lohse, K., Chmelik, M., Martin, S. H., and Barton, N. H. 2016. Efficient strategies for calculating blockwise likelihoods under the coalescent. *Genetics*, 202(2): 775–786.

Mallet, J., Beltrán, M., Neukirchen, W., and Linares, M. 2007. Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC Evol. Biol.*, 7(1): 28.

Marques, D. A., Meier, J. I., and Seehausen, O. 2019. A combinatorial view on speciation and adaptive radiation. *Trends Ecol. Evol.*, 34(6): 531–544.

Martin, S. H. and Jiggins, C. D. 2017. Interpreting the genomic landscape of introgression. *Curr. Opin. Genet. Dev.*, 47: 69–74.

Martin, S. H. and Van Belleghem, S. M. 2017. Exploring evolutionary relationships across the genome using topology weighting. *Genetics*, 206(1): 429–438.

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., and Jiggins, C. D. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res.*, 23(11): 1817–1828.

Martin, S. H., Eriksson, A., Kozak, K. M., Manica, A., and Jiggins, C. D. 2015. Speciation in *Heliconius* butterflies: Minimal contact followed by millions of generations of hybridisation. *bioRxiv*.

Martin, S. H., Davey, J. W., Salazar, C., and Jiggins, C. D. 2019. Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biol.*, 17(2): e2006288.

Moran, B. M., Payne, C., Langdon, Q., Powell, D. L., Brandvain, Y., and Schumer, M. 2021. The genomic consequences of hybridization. *eLife*, 10: e69016.

Naisbit, R. E., Jiggins, C. D., and Mallet, J. 2001. Disruptive sexual selection against hybrids contributes to speciation between

Heliconius cydno and *Heliconius melpomene*. *Proc. R. Soc. Lond. B*, 268(1478): 1849–1854.

Naisbit, R. E., Jiggins, C. D., Linares, M., Salazar, C., and Mallet, J. 2002. Hybrid sterility, Haldane's rule and speciation in *Heliconius cydno* and *H. melpomene*. *Genetics*, 161(4): 1517–1526.

Nielsen, R. and Wakeley, J. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, 158: 885–896.

Peters, K. J., Myers, S. A., Dudaniec, R. Y., O'Connor, J. A., and Kleindorfer, S. 2017. Females drive asymmetrical introgression from rare to common species in Darwin's tree finches. *J. Evol. Biol.*, 30(11): 1940–1952.

Rannala, B. and Yang, Z. 2017. Efficient Bayesian species tree inference under the multispecies coalescent. *Syst. Biol.*, 66: 823–842.

Slotman, M., Torre, A. d., and Powell, J. R. 2004. The genetics of inviability and male sterility in hybrids between *Anopheles gambiae* and *An. arabiensis*. *Genetics*, 167(1): 275–287.

Slotman, M., della Torre, A., and Powell, J. R. 2005a. Female sterility in hybrids between *Anopheles gambiae* and *A. arabiensis*, and the causes of Haldane's rule. *Evolution*, 59(5): 1016–1026.

Slotman, M. A., della Torre, A., Calzetta, M., and Powell, J. R. 2005b. Differential introgression of chromosomal regions between *Anopheles gambiae* and *An. arabiensis*. *Am. J. Trop. Med. Hyg.*, 73(2): 326–335.

Thawornwattana, Y., Dalquen, D., and Yang, Z. 2018. Coalescent analysis of phylogenomic data confidently resolves the species relationships in the *Anopheles gambiae* species complex. *Mol. Biol. Evol.*, 35(10): 2512–2527.

Thawornwattana, Y., Seixas, F. A., Mallet, J., and Yang, Z. 2022. Full-likelihood genomic analysis clarifies a complex history of species divergence and introgression: the example of the *erato-sara* group of *Heliconius* butterflies. *Syst. Biol.*, 71(5): 1159–1177.

Wen, D. and Nakhleh, L. 2018. Coestimating reticulate phylogenies and gene trees from multilocus sequence data. *Syst. Biol.*, 67(3): 439–457.

Yang, Z. 2014. *Molecular Evolution: A Statistical Approach*. Oxford University Press, Oxford, England.

Yang, Z. 2015. The BPP program for species tree estimation and species delimitation. *Curr. Zool.*, 61: 854–865.

Yang, Z. and Flouri, T. 2022. Estimation of cross-species introgression rates using genomic data despite model unidentifiability. *Mol. Biol. Evol.*, 39(5). msac083.

Yang, Z. and Zhu, T. 2018. Bayesian selection of misspecified models is overconfident and may cause spurious posterior probabilities for phylogenetic trees. *Proc. Natl. Acad. Sci. U.S.A.*, 115(8): 1854–1859.

Yu, Y., Degnan, J. H., and Nakhleh, L. 2012. The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. *PLoS Genet.*, 8(4): e1002660.

Yu, Y., Dong, J., Liu, K. J., and Nakhleh, L. 2014. Maximum likelihood inference of reticulate evolutionary histories. *Proc. Natl. Acad. Sci. U.S.A.*, 111(46): 16448–16453.

Zhang, C., Ogilvie, H. A., Drummond, A. J., and Stadler, T. 2018. Bayesian inference of species networks from multilocus sequence data. *Mol. Biol. Evol.*, 35: 504–517.

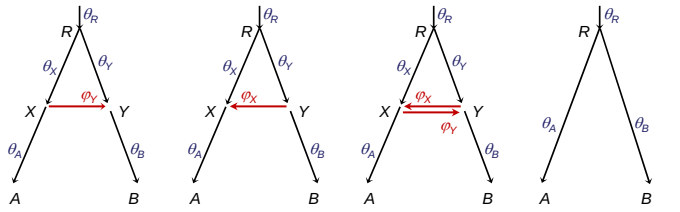

Figure 1: (**a**–**c**) MSci models for two species with (**a**) $A \to B$ introgression (I for 'inflow'), (**b**) $B \to A$ introgression (O for 'outflow'), or (**c**) bidirectional introgression (B). The parameters under each model are $\Theta_{\rm I} = (\tau_R, \tau_X, \theta_A, \theta_B, \theta_X, \theta_Y, \theta_R, \varphi_Y)$, $\Theta_{\rm O} = (\tau_R, \tau_X, \theta_A, \theta_B, \theta_X, \theta_Y, \theta_R, \varphi_X)$, $\Theta_{\rm B} = (\tau_R, \tau_X, \theta_A, \theta_B, \theta_X, \theta_Y, \theta_R, \varphi_X, \varphi_Y)$. The rate of introgression is represented by the introgression probability: $\varphi_Y \equiv \varphi_{A\to B}$ in **a** and **c**, and $\varphi_X \equiv \varphi_{B\to A}$ in **b** and **c**. (**d**) MSC model with no gene flow, with parameters $\Theta_0 = (\tau_R, \theta_A, \theta_B, \theta_R)$.

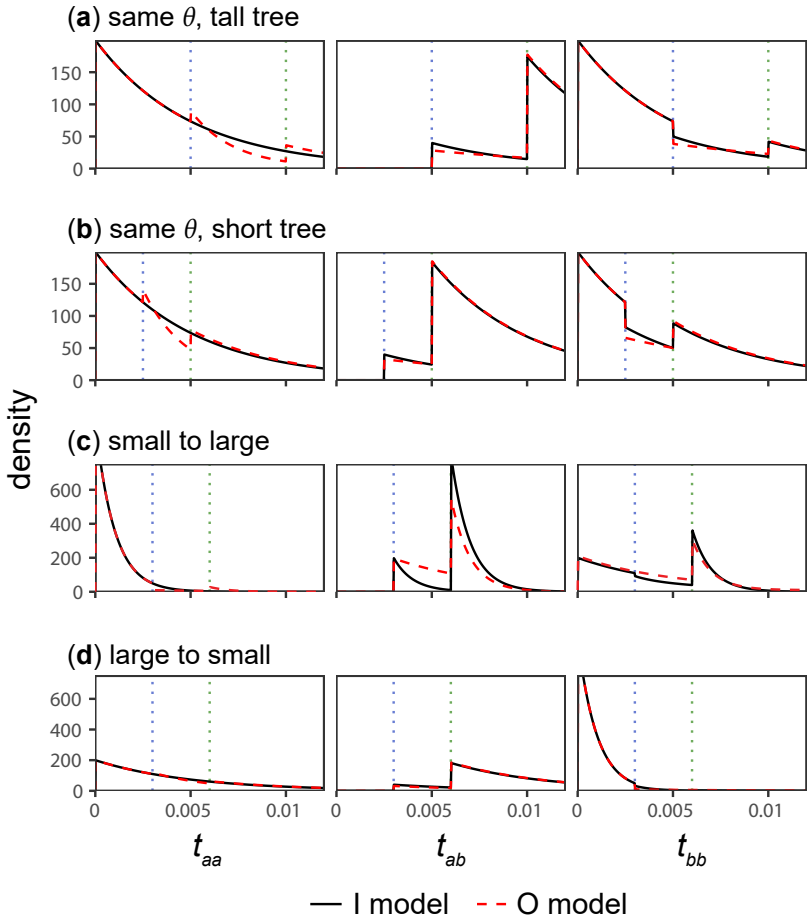

Figure 2: The true (black lines for I) and fitted (red lines for O) distributions of coalescent times (t_{aa}, t_{ab}, t_{bb}) for four sets of parameter values (cases **a**–**d**), when data are generated under the inflow (I) model of figure 1**a** and analyzed under the I, O, and B models of figure 1**a**–**c**. The true parameter values in the I model ($\Theta_{\rm I}$) are shown in table S1. The best-fitting parameter values under the O model ($\Theta^*_{\rm O}$) are approximated using the average estimates in BPP analysis of simulated large datasets of $L = 4000$ loci (with $S = 4$ sequences per species per locus and $n = 500$ sites in the sequence) shown in table S1.

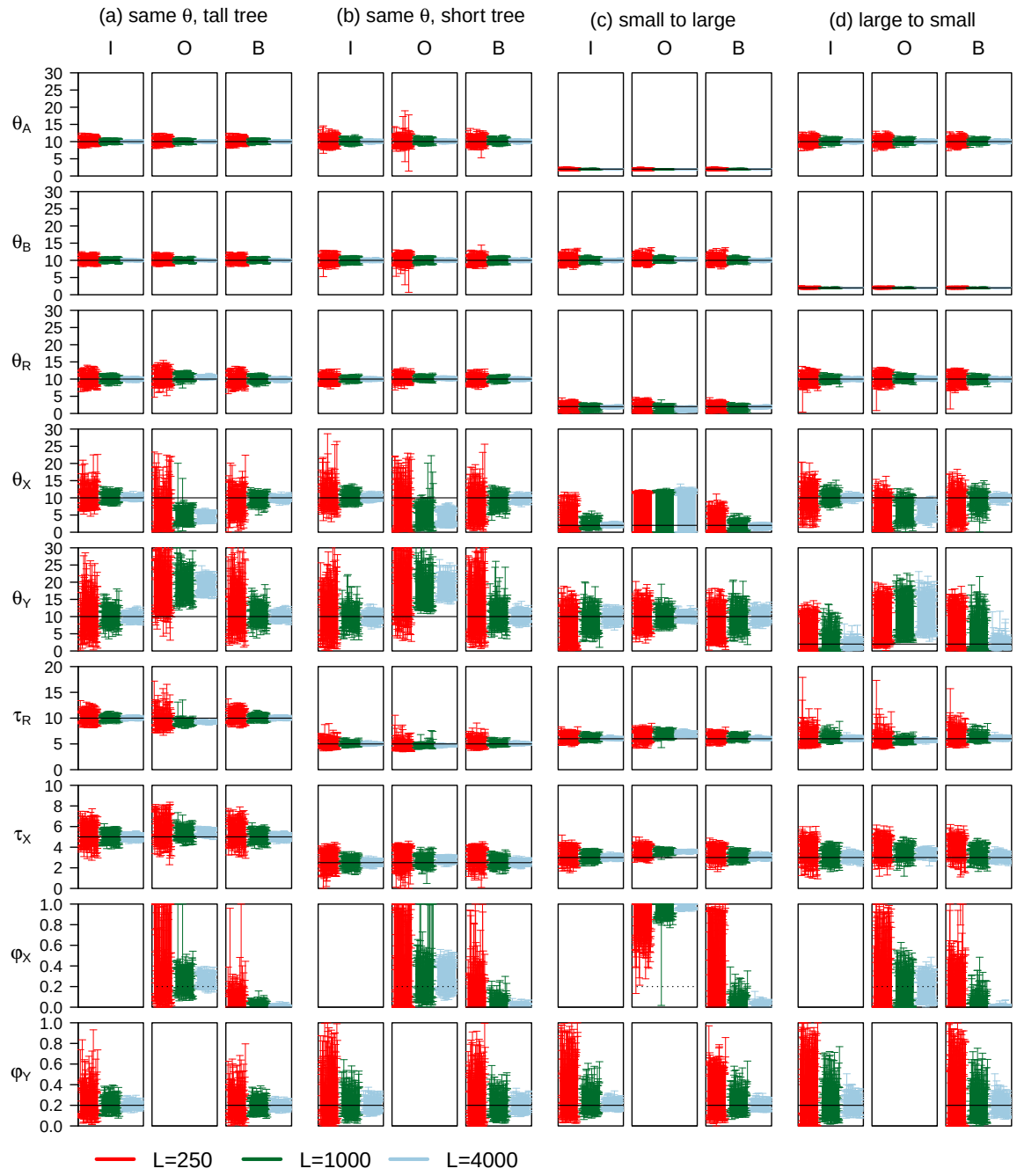

Figure 3: The 95% HPD CIs for parameters in 100 replicate datasets (each of L loci) simulated under model I and analyzed under models I, O, and B of figure 1**a**–**c**. Four sets of parameter values are used (cases **a**–**d**) (table S1). Parameters θs and τs are multiplied by 10^3. Black solid lines indicate the true values. Dotted lines for φ_X in model O indicate the true value of φ_Y in model I.

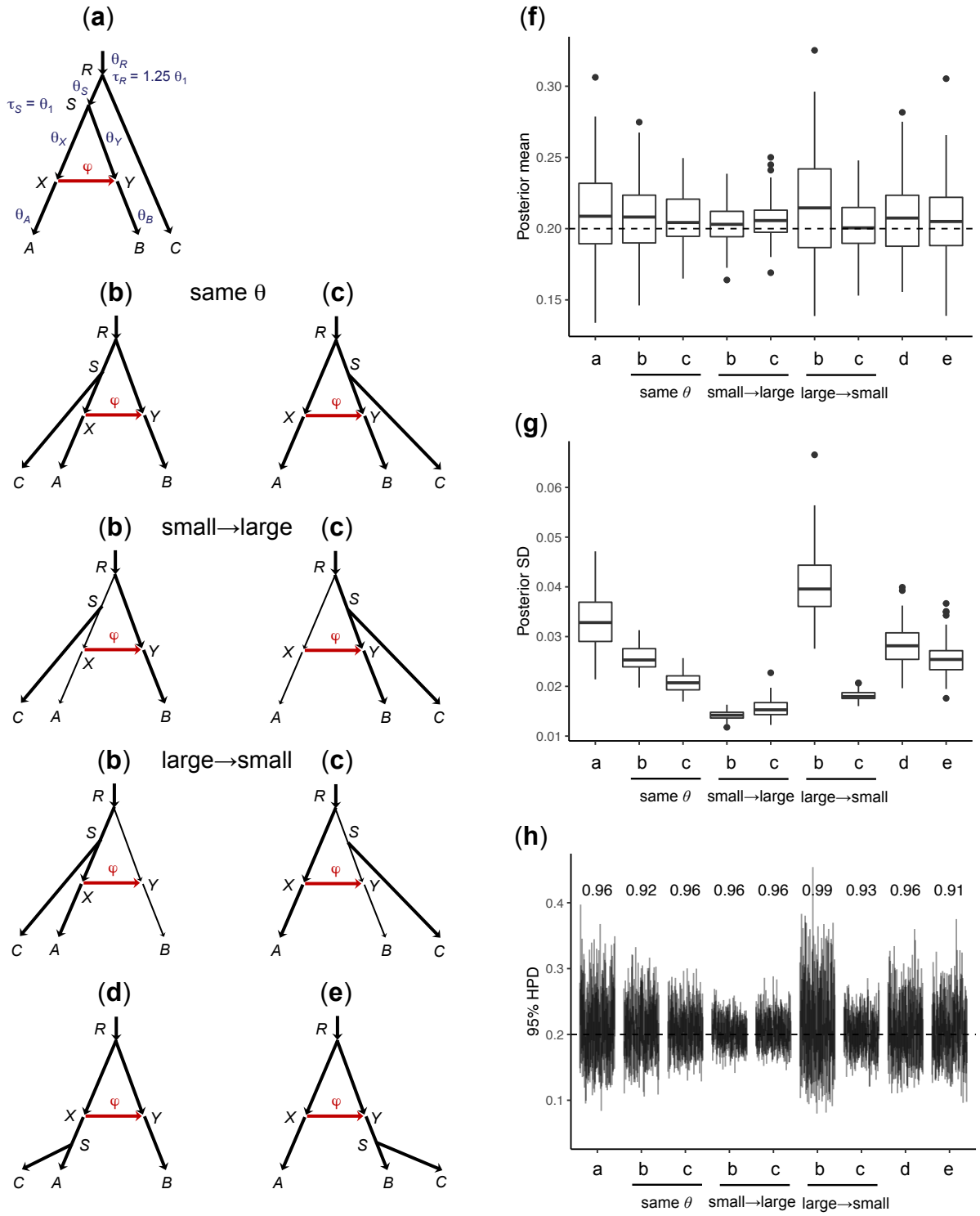

Figure 4: (**a**–**e**) MSci models for three species (A, B, C), with introgression from A to B, resulting from adding a third species C onto the two-species tree of figure 1**a** at five possible locations: (**a**) root population, (**b**, **c**) source and target populations before gene flow, and (**d**, **e**) source and target populations after gene flow. (**f**) Box plots of the posterior means for φ among 100 replicate datasets simulated under each of the five cases (**a**–**e**). The dashed line indicates the true value ($\varphi = 0.2$). (**g**) Box plots of the posterior SD for φ. (**h**) 95% HPD CIs for φ, with the CI coverage above the CI bars. See figure S2 for CIs for other parameters.

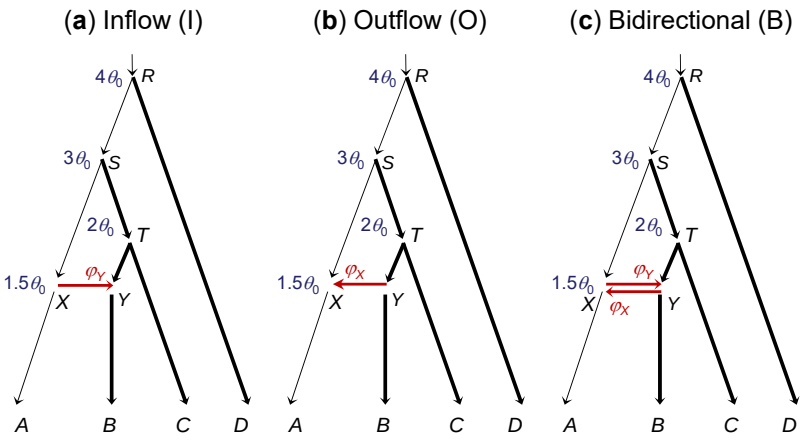

Figure 5: **(4s-trees)** Three MSci models for four species differing in the direction of introgression assumed to simulate and analyze data: (**a**) inflow from A to B (I); (**b**) outflow from B to A (O); and (**c**) bidirectional introgression between A and B (B). Divergence times used are shown next to the nodes: $\tau_R = 4\theta_0$, $\tau_S = 3\theta_0$, $\tau_T = 2\theta_0$, and $\tau_X = \tau_Y = 1.5\theta_0$, with population sizes $\theta_0 = 0.002$ for the thin branches and $\theta_1 = 0.01$ for the thick branches. We also used a setting in which all populations on the species tree have the same size, with $\theta_0 = \theta_1 = 0.01$. The introgression probabilities are $\varphi_X = \varphi_Y = 0.2$.

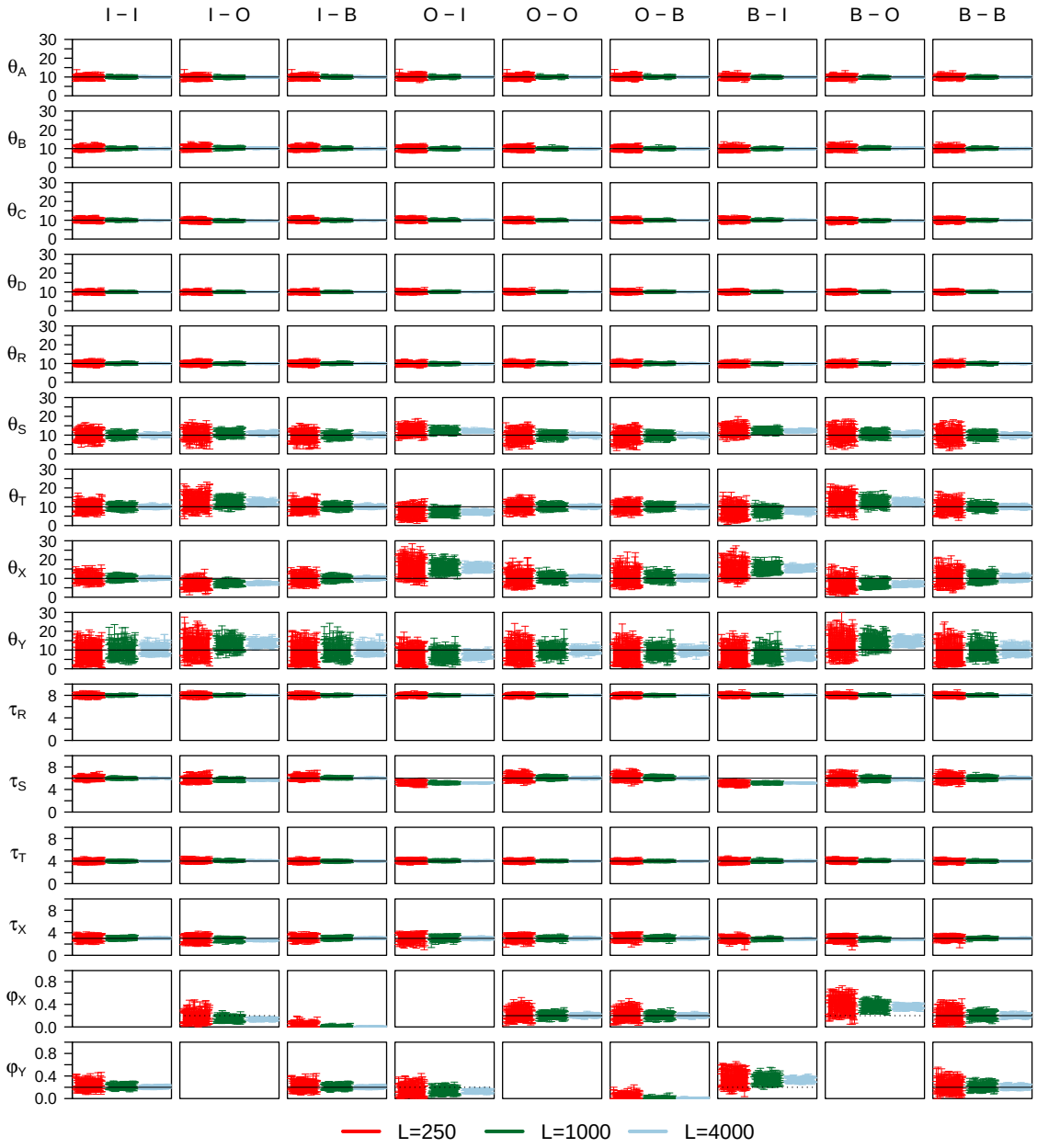

Figure 6: (**4s-same-θ**) The 95% HPD CIs for parameters in 100 replicate datasets simulated and analyzed under models I, O, and B of figure 5, assuming the same population size for all species ($\theta_0 = \theta_1 = 0.01$) when the data are generated. The nine settings are specified in the simulation-analysis format; i.e., 'I-O' means that data were simulated under model I and analyzed under model O. Parameters θ and τ are multiplied by 10^3. Horizontal solid lines indicate the true values. In the I-O setting, the dotted line for φ_X in model O indicates the true value of φ_Y in model I assumed in the simulation, while in the O-I setting, the dotted line for φ_Y indicates the true value of φ_X in the assumed O model. In the B-I and B-O settings, two introgression probabilities exist in the simulation model (φ_X, φ_Y) but only one is assumed in the analysis model, and the dotted line indicates its true value.

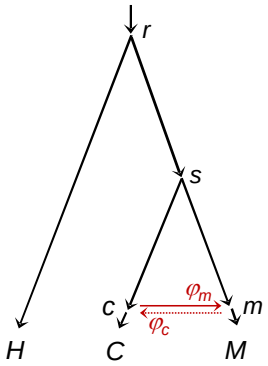

Figure 7: Species tree for *Heliconius hecale* (H), *H. cydno* (C), and *H. melpomene* (M), with introgression between *H. cydno* and *H. melpomene*, used to analyze genomic datasets. The introgresion (MSci) model involves the species divergence and introgression times ($\tau_r, \tau_s, \tau_c = \tau_m$) and population sizes for branches on the tree (e.g., θ_C for branch C and θ_c for branch sc), as well as the introgression probability (e.g., φ_m for the $c \to m$ introgression). The data support the $C \to M$ introgression but not the $M \to C$ introgression, with $\varphi_m > 0$ and $\varphi_c \approx 0$.

Table 1. Posterior means and 95% HPD CIs for parameters in BPP analyses of two datasets of noncoding and coding loci from *Heliconius* butterflies (fig. 7) under four MSci models with different introgression directions

	Model 0 (no gene flow)	Model I ($C \rightarrow M$)	Model O ($M \rightarrow C$)	Model B ($C \rightleftarrows M$)
Noncoding loci ($L = 5{,}341$ loci)				
θ_H	0.0131 (0.0127, 0.0136)	0.0134 (0.0129, 0.0139)	0.0134 (0.0129, 0.0138)	0.0134 (0.0129, 0.0139)
θ_C	0.0407 (0.0329, 0.0496)	0.0500 (0.0274, 0.0759)	0.0231 (0.0070, 0.0415)	0.0499 (0.0267, 0.0759)
θ_M	0.0026 (0.0021, 0.0031)	0.0003 (0.0002, 0.0005)	0.0001 (0.0000, 0.0002)	0.0003 (0.0002, 0.0005)
θ_r	0.0124 (0.0119, 0.0128)	0.0123 (0.0118, 0.0127)	0.0122 (0.0118, 0.0127)	0.0123 (0.0118, 0.0127)
θ_s	0.0343 (0.0328, 0.0358)	0.0152 (0.0141, 0.0162)	0.0185 (0.0175, 0.0194)	0.0152 (0.0141, 0.0162)
θ_c	n/a	0.0256 (0.0241, 0.0271)	0.0230 (0.0206, 0.0254)	0.0255 (0.0240, 0.0270)
θ_m	n/a	0.0188 (0.0162, 0.0214)	0.0294 (0.0262, 0.0327)	0.0189 (0.0164, 0.0215)
τ_r	0.0116 (0.0114, 0.0117)	0.0118 (0.0116, 0.0120)	0.0118 (0.0116, 0.0120)	0.0118 (0.0116, 0.0120)
τ_s	0.0010 (0.0008, 0.0012)	0.0068 (0.0064, 0.0072)	0.0051 (0.0048, 0.0053)	0.0068 (0.0064, 0.0071)
$\tau_c = \tau_m$	n/a	0.0001 (0.0001, 0.0002)	0.0000 (0.0000, 0.0001)	0.0001 (0.0001, 0.0002)
φ_c	n/a	n/a	0.1744 (0.1458, 0.2038)	0.0019 (0.0000, 0.0057)
φ_m	n/a	0.2830 (0.2565, 0.3090)	n/a	0.2802 (0.2530, 0.3067)
Coding loci ($L = 4{,}942$ loci)				
θ_H	0.0055 (0.0053, 0.0058)	0.0055 (0.0053, 0.0058)	0.0055 (0.0052, 0.0057)	0.0055 (0.0053, 0.0058)
θ_C	0.0054 (0.0048, 0.0060)	0.0361 (0.0203, 0.0545)	0.0307 (0.0133, 0.0513)	0.0363 (0.0204, 0.0553)
θ_M	0.0016 (0.0015, 0.0018)	0.0010 (0.0008, 0.0011)	0.0005 (0.0003, 0.0008)	0.0010 (0.0008, 0.0011)
θ_r	0.0092 (0.0088, 0.0096)	0.0092 (0.0088, 0.0096)	0.0094 (0.0090, 0.0098)	0.0092 (0.0088, 0.0096)
θ_s	0.0117 (0.0111, 0.0124)	0.0027 (0.0004, 0.0054)	0.0092 (0.0084, 0.0100)	0.0027 (0.0004, 0.0053)
θ_c	n/a	0.0059 (0.0055, 0.0063)	0.0044 (0.0032, 0.0055)	0.0058 (0.0053, 0.0062)
θ_m	n/a	0.0119 (0.0076, 0.0168)	0.0105 (0.0072, 0.0144)	0.0129 (0.0077, 0.0189)
τ_r	0.0049 (0.0047, 0.0050)	0.0049 (0.0047, 0.0050)	0.0048 (0.0047, 0.0050)	0.0049 (0.0047, 0.0050)
τ_s	0.0009 (0.0008, 0.0010)	0.0047 (0.0045, 0.0049)	0.0017 (0.0015, 0.0019)	0.0047 (0.0045, 0.0049)
$\tau_c = \tau_m$	n/a	0.0005 (0.0004, 0.0006)	0.0002 (0.0001, 0.0003)	0.0005 (0.0004, 0.0006)
φ_c	n/a	n/a	0.1360 (0.0783, 0.1959)	0.0073 (0.0000, 0.0194)
φ_m	n/a	0.5119 (0.4780, 0.5451)	n/a	0.5064 (0.4722, 0.5412)

Table 2. Bayes factors for comparing four models of introgression for the *Heliconius* datasets (fig. 7, table 1), calculated using thermodynamic integration with 32 or 64 Gaussian quadrature points and Savage-Dickey density ratio with threshold $\varepsilon = 1\%$, 0.1%, or 0.01%

	Thermodynamic integration		Savage-Dickey density ratio		
	32 points	64 points	$\varepsilon = 1\%$	$\varepsilon = 0.1\%$	$\varepsilon = 0.01\%$
Noncoding loci ($L = 5{,}341$ loci)					
B_{I0} ($\varphi_{C\to M}$)	$e^{1087.1}$	$e^{1082.5}$	∞	∞	∞
B_{O0} ($\varphi_{M\to C}$)	$e^{946.9}$	$e^{904.9}$	∞	∞	∞
B_{BI} ($\varphi_{M\to C}$)	$e^{-5.6}$	$e^{-9.9}$	0.0101	0.0025	0.0020
B_{BO} ($\varphi_{C\to M}$)	$e^{134.6}$	$e^{167.8}$	∞	∞	∞
B_{IO} ($\varphi_{M\to C}$ vs. $\varphi_{C\to M}$)	$e^{140.2}$	$e^{177.6}$	n/a	n/a	n/a
B_{B0} ($\varphi_{M\to C}$ and $\varphi_{C\to M}$)	$e^{1081.6}$	$e^{1072.6}$	n/a	n/a	n/a
Coding loci ($L = 4{,}942$ loci)					
B_{I0} ($\varphi_{C\to M}$)	$e^{359.9}$	$e^{358.5}$	∞	∞	∞
B_{O0} ($\varphi_{M\to C}$)	$e^{128.0}$	$e^{147.6}$	∞	∞	∞
B_{BI} ($\varphi_{M\to C}$)	$e^{-13.0}$	$e^{-8.6}$	0.0073	0.0090	0.0136
B_{BO} ($\varphi_{C\to M}$)	$e^{218.9}$	$e^{202.3}$	∞	∞	∞
B_{IO} ($\varphi_{M\to C}$ vs. $\varphi_{C\to M}$)	$e^{231.9}$	$e^{210.9}$	n/a	n/a	n/a
B_{B0} ($\varphi_{M\to C}$ and $\varphi_{C\to M}$)	$e^{346.8}$	$e^{349.9}$	n/a	n/a	n/a

Note.— The four models (table 1) are (0) MSC with no gene flow, (I) C $\to$ M introgression, (O) M $\to$ C introgression, and (B) bidirectional introgression (C $\leftrightarrows$ M). The approach based on Savage-Dickey density ratio produces $B = \infty$ if all values of φ in the MCMC sample are $> \varepsilon$.

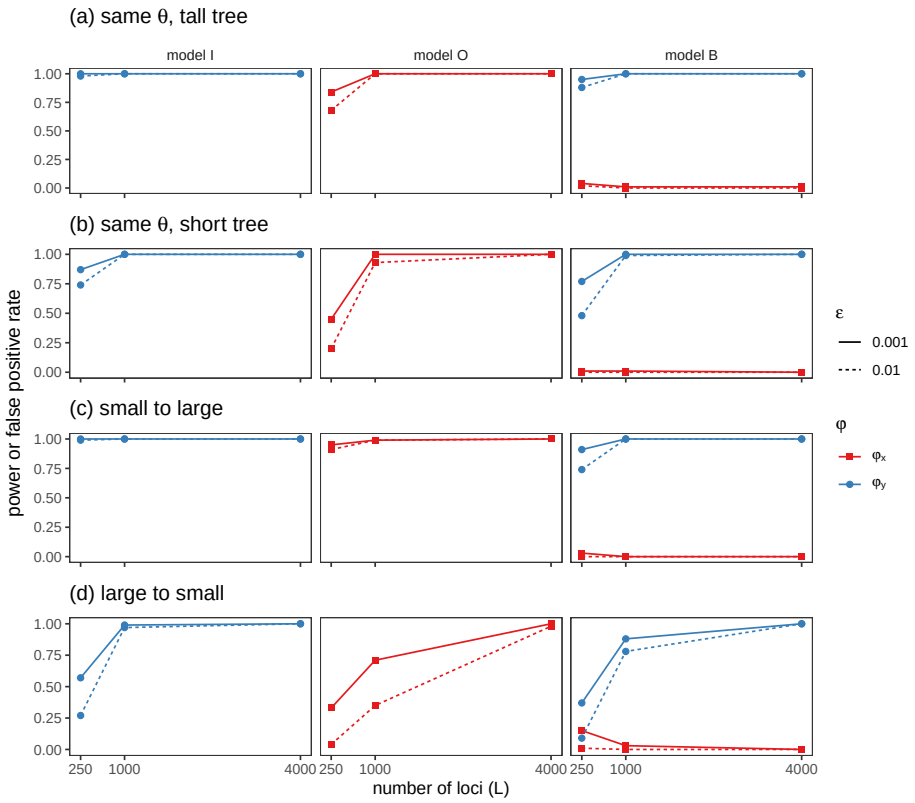

Figure S1: (**2s-power**) Power (blue) and false positive rate (red) of Bayesian test for introgression applied to the simulated data with two species under the I model of figure 1**a** using four sets of parameter values (cases **a**–**d**). Bayesian test is conducted using a cutoff of 100 for the Bayes factor, calculated using the Savage-Dickey density ratio with the small value for null effect (ε). Note that the I model is the true model with $A \rightarrow B$ introgression at the rate φ_Y. A significant result for testing φ_Y under the I or B models is considered a true positive, whereas a significant result for testing φ_X under the O or B models is considered a false positive. Parameter estimates from those data are summarized in figure 3.

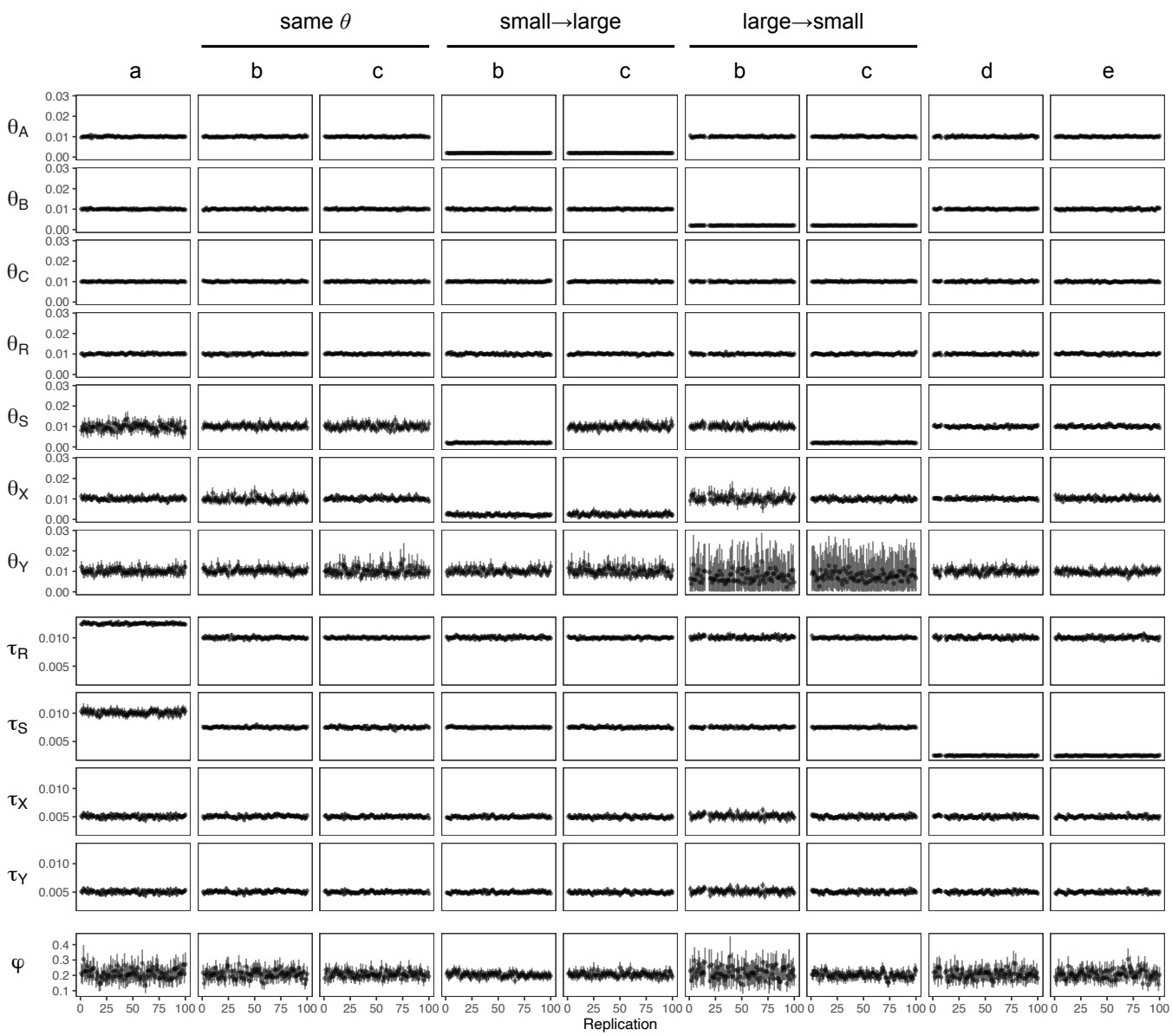

Figure S2: The 95% HPD CIs for parameters in 100 replicate datasets simulated and analyzed under the models of figure 4**a**–**e**. Results for φ are summarized in figure 4**h**.

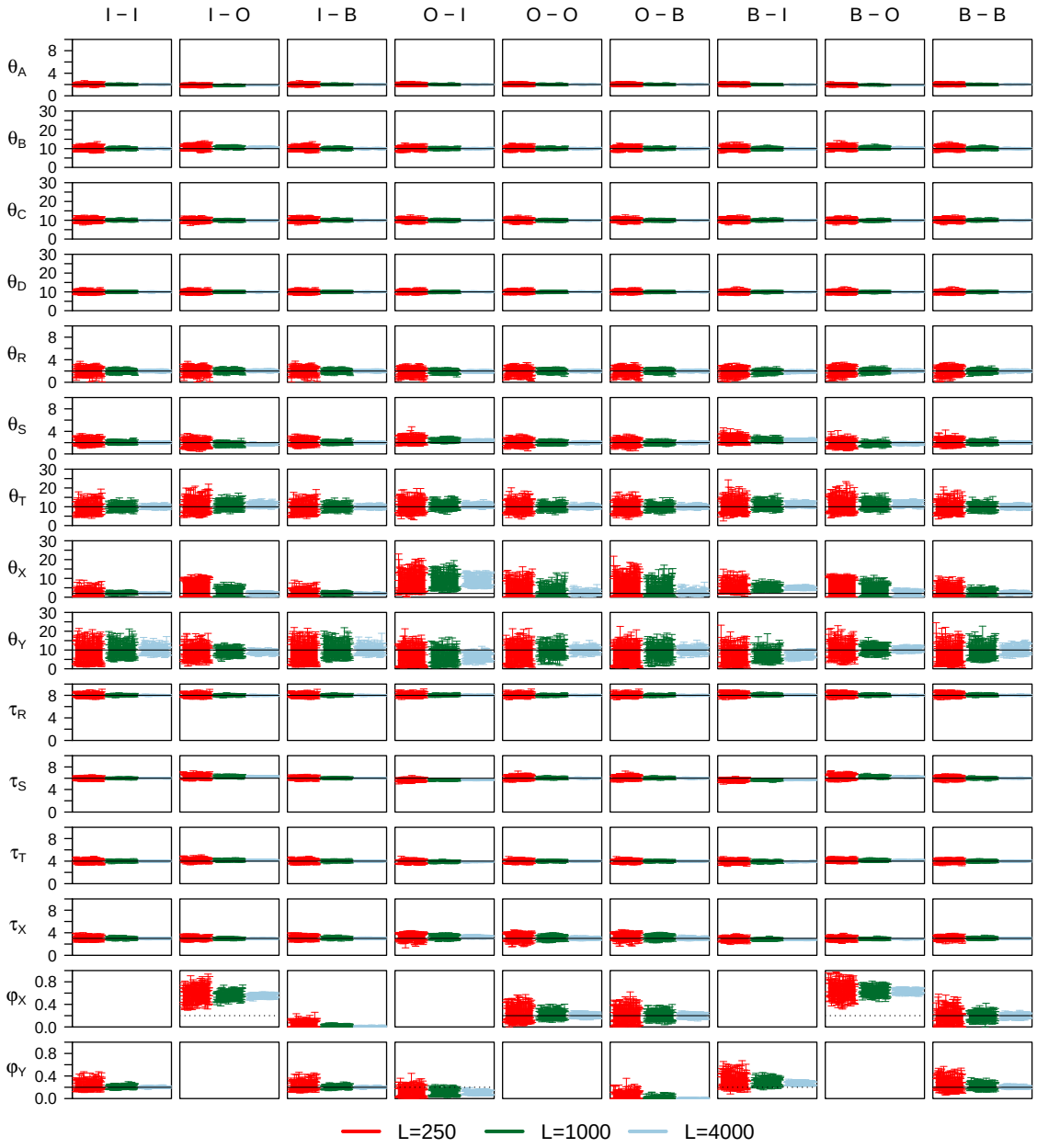

Figure S3: **(4s-diff-θ)** The 95% HPD CIs for parameters in 100 replicate datasets simulated and analyzed under models I, O, and B of figure 5, with $\theta_0 = 0.002$ for the thin branches and $\theta_1 = 0.01$ for the thick branches in the species tree. See legend to figure 6.

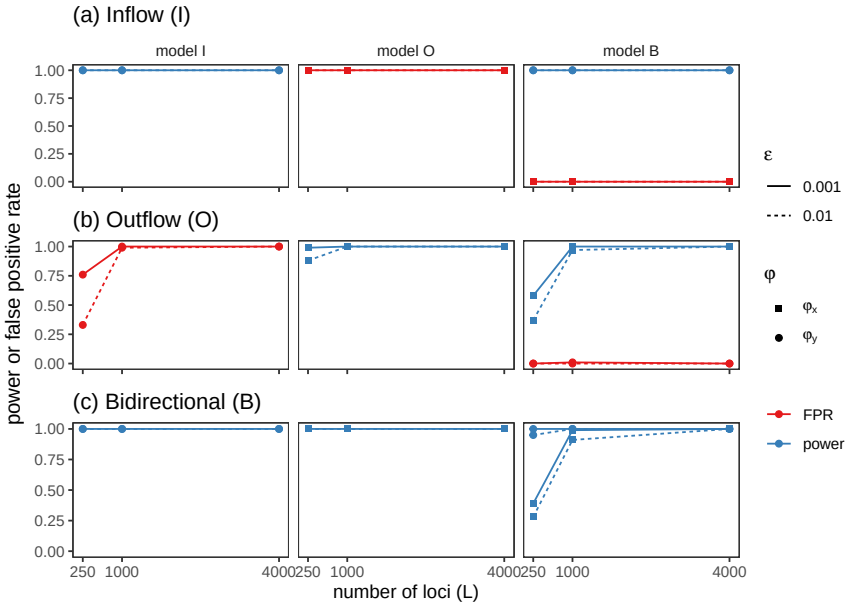

Figure S4: (**4s-same-θ-power**) Power (blue) and false positive rate (red) of Bayesian test for introgression applied to data of four species simulated under the (**a**) inflow (I), (**b**) outflow (O), and (**c**) the bidirectional (B) models of figure 5**a**–**c**, assuming the same θ for all populations. The data were analyzed under the same I, O, and B models, resulting in nine combinations. Parameter estimates from those data are summarized in figure 6.

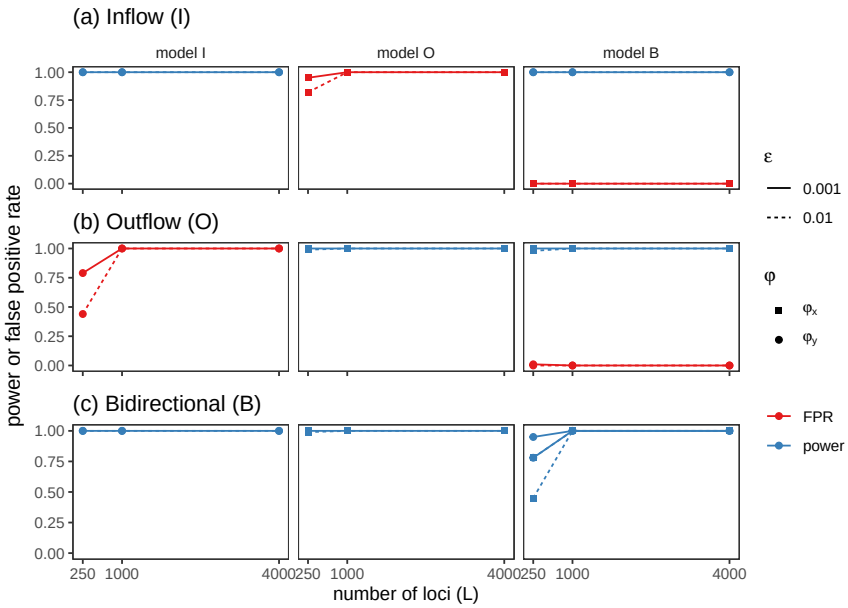

Figure S5: (**4s-diff-θ-power**) Power (blue) and false positive rate (red) of Bayesian test for introgression applied to data of four species simulated under the I, O, and B models of figure 5**a**–**c**, assuming different θ for populations on the species tree. Parameter estimates from those data are summarized in figure S3. See legend to figure S4.

Table S1. Average posterior means and 95% HPD CIs for parameters over 100 replicate datasets of $L = 4000$ loci simulated under the inflow (I) model and analyzed under the inflow (I), outflow (O) and bidirectional (B) models of figure 1 using four sets of parameter values (cases a–d)

	(**a**) Same θ tall tree				(**b**) Same θ short tree				(**c**) Small to large				(**d**) Large to small			
	Θ_I	$\hat{\Theta}_I$	$\hat{\Theta}_O$	$\hat{\Theta}_B$	Θ_I	$\hat{\Theta}_I$	$\hat{\Theta}_O$	$\hat{\Theta}_B$	Θ_I	$\hat{\Theta}_I$	$\hat{\Theta}_O$	$\hat{\Theta}_B$	Θ_I	$\hat{\Theta}_I$	$\hat{\Theta}_O$	$\hat{\Theta}_B$
θ_A	1.0	1.00 (0.97, 1.03)	1.00 (0.97, 1.03)	1.00 (0.97, 1.03)	1.00	1.01 (0.97, 1.04)	1.01 (0.97, 1.04)	1.01 (0.97, 1.04)	0.2	0.20 (0.19, 0.21)	0.19 (0.19, 0.20)	0.20 (0.19, 0.21)	1.0	1.00 (0.97, 1.03)	1.00 (0.97, 1.03)	1.00 (0.97, 1.03)
θ_B	1.0	1.00 (0.97, 1.03)	1.00 (0.97, 1.03)	1.00 (0.97, 1.03)	1.00	1.00 (0.97, 1.04)	1.00 (0.97, 1.04)	1.00 (0.97, 1.04)	1.0	1.00 (0.97, 1.03)	1.02 (0.99, 1.05)	1.00 (0.97, 1.03)	0.2	0.20 (0.19, 0.21)	0.20 (0.19, 0.21)	0.20 (0.19, 0.21)
θ_R	1.0	1.00 (0.96, 1.04)	1.06 (1.02, 1.10)	1.00 (0.95, 1.04)	1.00	1.00 (0.97, 1.03)	1.02 (0.99, 1.05)	1.00 (0.97, 1.03)	0.2	0.20 (0.16, 0.23)	0.14 (0.09, 0.19)	0.19 (0.16, 0.23)	1.0	1.00 (0.96, 1.04)	1.02 (0.98, 1.05)	1.00 (0.96, 1.04)
θ_X	1.0	1.01 (0.94, 1.08)	0.45 (0.33, 0.58)	0.98 (0.90, 1.06)	1.00	1.00 (0.93, 1.08)	0.44 (0.24, 0.65)	0.97 (0.88, 1.05)	0.2	0.20 (0.16, 0.25)	0.50 (0.01, 1.19)	0.18 (0.11, 0.24)	1.0	1.00 (0.93, 1.08)	0.65 (0.40, 0.88)	0.99 (0.91, 1.07)
θ_Y	1.0	1.00 (0.88, 1.12)	1.91 (1.69, 2.14)	1.02 (0.90, 1.15)	1.00	1.00 (0.85, 1.14)	1.84 (1.56, 2.14)	1.02 (0.87, 1.17)	1.0	0.98 (0.81, 1.15)	1.00 (0.88, 1.12)	1.01 (0.83, 1.19)	0.2	0.21 (0.09, 0.34)	1.07 (0.56, 1.65)	0.24 (0.10, 0.39)
τ_R	1.0	1.00 (0.97, 1.03)	0.93 (0.90, 0.95)	1.00 (0.98, 1.03)	0.50	0.50 (0.48, 0.52)	0.47 (0.45, 0.49)	0.50 (0.48, 0.52)	0.6	0.60 (0.58, 0.63)	0.69 (0.65, 0.73)	0.60 (0.58, 0.63)	0.6	0.60 (0.57, 0.63)	0.57 (0.54, 0.59)	0.61 (0.57, 0.64)
τ_X	0.5	0.50 (0.47, 0.53)	0.55 (0.52, 0.57)	0.50 (0.47, 0.53)	0.25	0.25 (0.23, 0.28)	0.28 (0.25, 0.31)	0.25 (0.23, 0.28)	0.3	0.30 (0.28, 0.32)	0.36 (0.34, 0.37)	0.31 (0.29, 0.33)	0.3	0.30 (0.27, 0.34)	0.35 (0.31, 0.38)	0.31 (0.27, 0.34)
φ_X	n/a	n/a	0.27 (0.20, 0.33)	0.01 (0.00, 0.02)	n/a	n/a	0.30 (0.18, 0.44)	0.01 (0.00, 0.04)	n/a	n/a	0.98 (0.96, 1.00)	0.02 (0.00, 0.05)	n/a	n/a	0.17 (0.06, 0.31)	0.01 (0.00, 0.02)
φ_Y	0.2	0.20 (0.17, 0.24)	n/a	0.19 (0.16, 0.23)	0.20	0.21 (0.15, 0.27)	n/a	0.20 (0.14, 0.26)	0.2	0.21 (0.17, 0.25)	n/a	0.20 (0.16, 0.25)	0.2	0.21 (0.13, 0.29)	n/a	0.20 (0.12, 0.29)

Note.— Θ_I denotes the true parameter values, while $\hat{\Theta}_I$, $\hat{\Theta}_O$, and $\hat{\Theta}_B$ are estimates under models I, O and B, respectively (fig. 1). There are $S = 4$ sequences per species per locus and $N = 500$ sites in the sequence. Values of τ and θ are multiplied by 100. Estimates for individual datasets and for all datasizes ($L = 250, 1000, 4000$) are plotted in figure 3.

Table S2. (4s-same-θ) Average posterior means and 95% HPD CIs for parameters over 100 replicate datasets of $L = 4000$ loci simulated and analyzed under the inflow (I), outflow (O) and bidirectional (B) models of figure 5 with the same population size (θ) for all species

	True model I				True model O				True model B			
	Θ_I	$\hat{\Theta}_I$	$\hat{\Theta}_O$	$\hat{\Theta}_B$	Θ_O	$\hat{\Theta}_I$	$\hat{\Theta}_O$	$\hat{\Theta}_B$	Θ_B	$\hat{\Theta}_I$	$\hat{\Theta}_O$	$\hat{\Theta}_B$
θ_A	1.0	1.00 (0.97, 1.03)	0.98 (0.95, 1.02)	1.00 (0.97, 1.03)	1.0	1.00 (0.97, 1.03)	1.00 (0.97, 1.03)	1.00 (0.97, 1.03)	1.0	1.00 (0.96, 1.03)	0.98 (0.95, 1.01)	1.00 (0.97, 1.03)
θ_B	1.0	1.00 (0.97, 1.03)	1.02 (0.99, 1.06)	1.00 (0.97, 1.03)	1.0	0.99 (0.96, 1.03)	1.00 (0.97, 1.03)	1.00 (0.97, 1.03)	1.0	0.99 (0.96, 1.03)	1.02 (0.98, 1.05)	1.00 (0.97, 1.03)
θ_C	1.0	1.00 (0.97, 1.03)	0.97 (0.95, 1.00)	1.00 (0.97, 1.03)	1.0	1.02 (0.99, 1.05)	1.00 (0.97, 1.03)	1.00 (0.97, 1.03)	1.0	1.01 (0.98, 1.04)	0.98 (0.95, 1.00)	1.00 (0.97, 1.03)
θ_D	1.0	1.00 (0.98, 1.02)	1.00 (0.98, 1.02)	1.00 (0.98, 1.02)	1.0	1.00 (0.98, 1.02)	1.00 (0.98, 1.02)	1.00 (0.98, 1.02)	1.0	1.00 (0.98, 1.02)	1.00 (0.98, 1.02)	1.00 (0.98, 1.02)
θ_R	1.0	1.00 (0.97, 1.03)	0.99 (0.96, 1.02)	1.00 (0.97, 1.03)	1.0	0.98 (0.96, 1.01)	1.00 (0.97, 1.03)	1.00 (0.97, 1.03)	1.0	0.98 (0.95, 1.01)	0.99 (0.96, 1.02)	1.00 (0.97, 1.02)
θ_S	1.0	1.00 (0.93, 1.08)	1.11 (1.03, 1.19)	0.99 (0.92, 1.07)	1.0	1.22 (1.16, 1.29)	0.99 (0.91, 1.07)	0.99 (0.91, 1.07)	1.0	1.23 (1.17, 1.30)	1.07 (0.99, 1.16)	1.00 (0.91, 1.08)
θ_T	1.0	1.00 (0.92, 1.08)	1.24 (1.12, 1.36)	1.00 (0.92, 1.08)	1.0	0.71 (0.62, 0.81)	1.00 (0.93, 1.08)	1.00 (0.92, 1.07)	1.0	0.77 (0.66, 0.88)	1.26 (1.15, 1.38)	1.00 (0.91, 1.08)
θ_X	1.0	1.00 (0.94, 1.07)	0.74 (0.68, 0.80)	0.99 (0.92, 1.06)	1.0	1.55 (1.39, 1.72)	1.00 (0.90, 1.11)	1.01 (0.91, 1.12)	1.0	1.52 (1.39, 1.66)	0.69 (0.59, 0.79)	1.01 (0.88, 1.13)
θ_Y	1.0	1.00 (0.77, 1.26)	1.36 (1.19, 1.54)	1.01 (0.77, 1.28)	1.0	0.74 (0.58, 0.91)	0.98 (0.82, 1.14)	0.96 (0.80, 1.12)	1.0	0.70 (0.52, 0.90)	1.44 (1.26, 1.63)	0.98 (0.76, 1.21)
τ_R	8.0	0.80 (0.79, 0.81)	0.80 (0.79, 0.81)	0.80 (0.79, 0.81)	8.0	0.80 (0.79, 0.81)	0.80 (0.79, 0.81)	0.80 (0.79, 0.81)	8.0	0.80 (0.80, 0.81)	0.80 (0.79, 0.81)	0.80 (0.79, 0.81)
τ_S	6.0	0.60 (0.59, 0.61)	0.57 (0.56, 0.58)	0.60 (0.59, 0.61)	6.0	0.52 (0.51, 0.53)	0.60 (0.59, 0.62)	0.60 (0.59, 0.62)	6.0	0.52 (0.51, 0.52)	0.58 (0.56, 0.60)	0.60 (0.58, 0.62)
τ_T	4.0	0.40 (0.39, 0.41)	0.41 (0.40, 0.42)	0.40 (0.39, 0.41)	4.0	0.41 (0.40, 0.42)	0.40 (0.39, 0.41)	0.40 (0.39, 0.41)	4.0	0.40 (0.39, 0.41)	0.41 (0.40, 0.42)	0.40 (0.39, 0.41)
τ_X	3.0	0.30 (0.29, 0.32)	0.27 (0.26, 0.29)	0.30 (0.29, 0.32)	3.0	0.30 (0.28, 0.32)	0.30 (0.28, 0.32)	0.30 (0.29, 0.32)	3.0	0.29 (0.28, 0.30)	0.28 (0.27, 0.30)	0.30 (0.29, 0.31)
φ_X	n/a	n/a	0.13 (0.11, 0.16)	0.01 (0.00, 0.01)	0.2	n/a	0.20 (0.17, 0.23)	0.20 (0.17, 0.23)	0.2	n/a	0.36 (0.32, 0.40)	0.20 (0.17, 0.23)
φ_Y	0.2	0.20 (0.18, 0.23)	n/a	0.20 (0.18, 0.22)	n/a	0.13 (0.10, 0.16)	n/a	0.01 (0.00, 0.01)	0.2	0.33 (0.29, 0.37)	n/a	0.21 (0.17, 0.24)

Note.— Θ_I, Θ_O, and Θ_B denote the true parameter values in the true model, while $\hat{\Theta}_I$, $\hat{\Theta}_O$ and $\hat{\Theta}_B$ are estimates (fig. 5). Each dataset consists of $L = 4000$ loci, with $S = 4$ sequences per species per locus and $N = 500$ sites in the sequence. Values of τ and θ are multiplied by 100. Results for all data sizes with $L = 250, 1000$ or 4000 loci are shown in figure 6.

Table S3. (4s-diff-θ) Average posterior means and 95% HPD CIs for parameters over 100 replicate datasets of $L = 4000$ loci simulated and analyzed under the I, O, and B models of figure 5 with different population sizes

	Model I				Model O				Model B			
	Θ_I	$\hat{\Theta}_I$	$\hat{\Theta}_O$	$\hat{\Theta}_B$	Θ_O	$\hat{\Theta}_I$	$\hat{\Theta}_O$	$\hat{\Theta}_B$	Θ_B	$\hat{\Theta}_I$	$\hat{\Theta}_O$	$\hat{\Theta}_B$
θ_A	0.2	0.20 (0.19, 0.21)	0.19 (0.18, 0.19)	0.20 (0.19, 0.21)	0.2	0.20 (0.19, 0.21)	0.20 (0.19, 0.21)	0.20 (0.19, 0.21)	0.2	0.20 (0.19, 0.20)	0.19 (0.18, 0.20)	0.20 (0.19, 0.21)
θ_B	1.0	1.00 (0.97, 1.03)	1.06 (1.03, 1.10)	1.00 (0.97, 1.03)	1.0	1.00 (0.97, 1.03)	1.00 (0.97, 1.03)	1.00 (0.97, 1.03)	1.0	1.00 (0.96, 1.03)	1.04 (1.01, 1.07)	1.00 (0.97, 1.03)
θ_C	1.0	1.00 (0.97, 1.03)	0.99 (0.96, 1.01)	1.00 (0.97, 1.03)	1.0	1.00 (0.97, 1.03)	1.00 (0.97, 1.03)	1.00 (0.97, 1.03)	1.0	1.00 (0.97, 1.03)	0.99 (0.96, 1.01)	1.00 (0.97, 1.03)
θ_D	1.0	1.00 (0.97, 1.02)	1.00 (0.97, 1.02)	1.00 (0.97, 1.02)	1.0	1.00 (0.97, 1.02)	1.00 (0.98, 1.02)	1.00 (0.98, 1.02)	1.0	1.00 (0.97, 1.02)	1.00 (0.97, 1.02)	1.00 (0.97, 1.02)
θ_R	0.2	0.20 (0.18, 0.22)	0.20 (0.19, 0.22)	0.20 (0.18, 0.22)	0.2	0.19 (0.18, 0.21)	0.20 (0.18, 0.22)	0.20 (0.18, 0.22)	0.2	0.19 (0.17, 0.21)	0.20 (0.18, 0.22)	0.20 (0.18, 0.22)
θ_S	0.2	0.20 (0.19, 0.21)	0.17 (0.16, 0.19)	0.20 (0.18, 0.21)	0.2	0.24 (0.22, 0.25)	0.20 (0.18, 0.22)	0.20 (0.18, 0.22)	0.2	0.25 (0.23, 0.26)	0.18 (0.16, 0.20)	0.20 (0.18, 0.22)
θ_T	1.0	1.00 (0.91, 1.09)	1.11 (0.99, 1.23)	0.99 (0.90, 1.08)	1.0	1.08 (0.97, 1.19)	1.00 (0.91, 1.10)	1.00 (0.90, 1.09)	1.0	1.12 (1.00, 1.25)	1.14 (1.01, 1.27)	1.00 (0.90, 1.11)
θ_X	0.2	0.20 (0.17, 0.24)	0.17 (0.07, 0.28)	0.19 (0.16, 0.23)	0.2	0.85 (0.59, 1.14)	0.22 (0.11, 0.33)	0.24 (0.12, 0.37)	0.2	0.48 (0.41, 0.56)	0.20 (0.05, 0.36)	0.22 (0.15, 0.29)
θ_Y	1.0	0.99 (0.78, 1.23)	0.89 (0.78, 1.00)	1.01 (0.78, 1.25)	1.0	0.64 (0.41, 0.86)	0.99 (0.82, 1.17)	0.98 (0.80, 1.16)	1.0	0.71 (0.56, 0.86)	1.03 (0.91, 1.15)	0.96 (0.75, 1.19)
τ_R	8.0	0.80 (0.79, 0.81)	0.80 (0.79, 0.81)	0.80 (0.79, 0.81)	8.0	0.80 (0.79, 0.81)	0.80 (0.79, 0.81)	0.80 (0.79, 0.81)	8.0	0.81 (0.80, 0.82)	0.80 (0.79, 0.81)	0.80 (0.79, 0.81)
τ_S	6.0	0.60 (0.59, 0.61)	0.63 (0.62, 0.64)	0.60 (0.59, 0.61)	6.0	0.57 (0.56, 0.58)	0.60 (0.59, 0.61)	0.60 (0.59, 0.61)	6.0	0.57 (0.56, 0.58)	0.62 (0.61, 0.63)	0.60 (0.59, 0.61)
τ_T	4.0	0.40 (0.39, 0.41)	0.41 (0.40, 0.42)	0.40 (0.39, 0.41)	4.0	0.39 (0.38, 0.40)	0.40 (0.39, 0.41)	0.40 (0.39, 0.41)	4.0	0.39 (0.38, 0.40)	0.41 (0.40, 0.42)	0.40 (0.39, 0.41)
τ_X	3.0	0.30 (0.29, 0.31)	0.30 (0.29, 0.30)	0.30 (0.29, 0.31)	3.0	0.33 (0.31, 0.35)	0.30 (0.28, 0.32)	0.30 (0.28, 0.33)	3.0	0.28 (0.27, 0.29)	0.29 (0.28, 0.30)	0.30 (0.29, 0.31)
φ_X	n/a	n/a	0.55 (0.51, 0.59)	0.01 (0.00, 0.02)	0.2	n/a	0.20 (0.16, 0.24)	0.20 (0.15, 0.24)	0.2	n/a	0.63 (0.59, 0.67)	0.19 (0.15, 0.24)
φ_Y	0.2	0.20 (0.19, 0.22)	n/a	0.20 (0.18, 0.22)	n/a	0.10 (0.07, 0.14)	n/a	0.00 (0.00, 0.01)	0.2	0.27 (0.24, 0.30)	n/a	0.20 (0.18, 0.23)

Note.— Results for all data sizes with $L = 250, 1000$ or 4000 loci are in figure S3. See legend to table S2.