Bayesian test of gene flow between sister lineages using genomic data

Ziheng Yang (orcid: 0000-0003-3351-7981)[1*], Xiyun Jiao (orcid: 0009-0006-3924-985X)[2], Sirui Cheng[1,3,4], Tianqi Zhu (orcid: 0009-0007-0360-487X)[3]

[1*]Department of Genetics, Evolution, and Environment, University College London, Gower Street, London WC1E 6BT, UK.

[2]Department of Statistics and Data Science, China Southern University of Science and Technology, Shenzhen, Guangdong 518055, China.

[3]State Key Laboratory of Mathematical Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China.

[4]University of Chinese Academy of Sciences, Beijing 100049, China.

*Corresponding author: z.yang@ucl.ac.uk;

Inference of interspecific gene flow using genomic data is important to reliable reconstruction of species phylogenies and to our understanding of the speciation process. Gene flow is harder to detect if it involves sister lineages than nonsisters; for example, most heuristic methods based on data summaries are unable to infer gene flow between sisters. Likelihood-based methods can identify introgression between sisters but the test exhibits several nonstandard features, including boundary problems, indeterminate parameters, and multiple routes from the alternative to the null hypotheses. In the Bayesian test, those irregularities pose challenges to the use of the Savage-Dickey (S-D) density ratio to calculate the Bayes factor. Here we develop a theory for applying the S-D approach under nonstandard conditions. We show that the Bayesian test of introgression between sister lineages has low false-positive rates and high power. We discuss issues surrounding the estimation of the rate of gene flow between sister lineages, especially at very low or very high rates, and suggest that evidence for gene flow between sisters be assessed via a Bayesian test. We find that the species split time has a major impact on the information content in the data, with more information at deeper divergence. We use a genomic dataset from *Sceloporus* lizards to illustrate the test of gene flow between sister lineages.

Keywords: Bayes factor, BPP, introgression, multispecies coalescent, Savage-Dickey density ratio

Introduction

Genomic data can be used to test introgression between sister species

Detecting gene flow between species and estimating the rate of gene flow is important to reconstructing species phylogenies and testing theories of speciation (Sousa and Hey, 2013; Payseur and Rieseberg, 2016; Roux *et al.*, 2016). A number of methods have been developed to infer gene flow using genomic data, including heuristic methods based on data summaries and likelihood methods based on parametric models of gene flow (Jiao *et al.*, 2021; Hibbins and Hahn, 2022). The former class of methods includes the *D*-statistic (or ABBA-BABA test) (Green *et al.*, 2010), HyDe (Blischak *et al.*, 2018) and PhyNest (Kong *et al.*, 2025), which use genome-wide site-pattern counts in a species quartet; SNaQ (Solis-Lemus and Ane, 2016; Solis-Lemus *et al.*, 2017; see also Yu *et al.*, 2012) which uses reconstructed gene tree topologies; and ∂a∂i (Gutenkunst *et al.*, 2009) and fastsimcoal2 (Excoffier *et al.*, 2013), which use the joint site frequency spectrum at single nucleotide polymorphism (SNP) sites. The latter class of methods, including both maximum likelihood (ML) and Bayesian inference, has the major feature that they use the multispecies coalescent (MSC) model with gene flow to average over genealogical histories underlying sequences sampled at each locus (Jiao *et al.*, 2021). In ML, this is achieved through numerical integration (Zhu and Yang, 2012; Dalquen *et al.*, 2017), while in Bayesian inference, it is through Markov chain Monte Carlo (MCMC) algorithms. Gene flow has been modelled as either a discrete introgression/hybridization event which occurred at a particular time point in the past (the MSC-I model; e.g., PhyloNet, Wen and Nakhleh, 2018; BPP, Flouri *et al.*, 2020) or as a continuous process with a constant migration rate per generation (the MSC-M model; IMa3, Hey *et al.*, 2018; BPP, Flouri *et al.*, 2023). Our main focus in this paper is inference under the discrete introgression model (MSC-I) in the Bayesian framework.

Introgression may be harder to detect if it involves sister lineages than nonsisters. For example, gene flow between nonsisters may cause asymmetries in the gene tree distribution and in site-pattern counts in genomic data for a species quartet, while gene flow between sisters does not. As a result, quartet summary methods

such as *D*, HyDe and SNaQ, which rely on counts of estimated gene trees or site patterns for species quartets, can detect gene flow between nonsisters but not between sisters. Likelihood methods use information in the coalescent times or branch lengths on gene trees, and in the genealogical variation across the genome, and can thus infer gene flow between sister lineages (Jiao *et al.*, 2021; Hibbins and Hahn, 2022). They also accommodate multiple samples from the same species (in particular from the recipient lineage), which may boost the information content in the data about gene flow considerably (Jiao *et al.*, 2021; Thawornwattana *et al.*, 2026). In simulations, Bayesian inference under both MSC-I and MSC-M models showed similar power in inferring gene flow between sisters to that between nonsisters (Huang *et al.*, 2020; Thawornwattana *et al.*, 2023, 2025).

We note that in this context "sister species" simply means that the lineages involved in introgressive hybridization are sister branches on the phylogeny, and does not even imply that the two species are closely related. For example, given the phylogeny $(A, (B, C))$, the $A \to B$ introgression is between nonsisters, but is between sisters if species C is not included in the data (e.g., if C is extinct or unsampled). Thus the distinction between gene flow involving sisters and that involving nonsisters has no biological significance. Also in this paper we use the terms species, populations, and lineages interchangeably.

The Bayes factor for testing nested hypotheses can be calculated via the Savage-Dickey (S-D) density ratio

The standard device for model comparison in the Bayesian framework is the Bayes factor, defined as the ratio of the marginal likelihood values between the two compared models (Jeffreys, 1935). The marginal likelihood under a model is an average (often a high-dimensional integral) over the parameters under the model. This typically requires expensive computation involving many MCMC runs (Fourment *et al.*, 2020, see also Discussion). In the special case of comparing nested hypotheses, the Bayes factor may be calculated via the Savage-Dickey (S-D) density ratio (Dickey, 1971; Ji *et al.*, 2023), which involves far less computation. In phylogenetics Suchard *et al.* (2003) used the S-D density ratio to conduct Bayesian tests of the molecular clock; however, in that study the phylogeny is allowed to differ between the clock and no-clock models which appears to invalidate the S-D approach.

Let H_0 be the null hypothesis and H_1 the alternative hypothesis, with H_0 being a special case of H_1. In other words, H_1 reduces to H_0 when certain parameters of interest in H_1 take special values (called the *null value*): $\omega = \omega_0$. Suppose 'nuisance parameters' λ exist in both models, so that the prior under H_0 is $\pi_0(\lambda)$ and that under H_1 is $\pi(\lambda, \omega) = \pi(\omega)\pi(\lambda|\omega)$. On the condition that the priors on nuisance parameters match between the two models at the null value,

$$\pi_0(\lambda) = \pi(\lambda|\omega_0), \tag{1}$$

the Bayes factor B_{10} in support of H_1 against H_0 is given by

$$B_{10} = \frac{\pi(\omega_0)}{\pi(\omega_0 \mid \boldsymbol{x})}, \tag{2}$$

where $\boldsymbol{x}$ represent the data, and $\pi(\omega_0)$ and $\pi(\omega_0|\boldsymbol{x})$ are the prior and posterior densities of ω, respectively, under H_1 evaluated at $\omega = \omega_0$. Dickey (1971) provided a proof of eq. 2, which accounts for presence of nuisance parameters (λ) but does not accommodate the possible presence of unidentifiable parameters. This is dealt with in the extended proof of Ji *et al.* (2023).

Irregularities of the test of introgression between sister lineages and objectives of this paper

Consider two species that diverged time τ ago with the phylogeny (A, B) (Fig. 1). In analysis of genomic sequence data, it is convenient to measure time by mutations, and τ is thus measured in the expected number of mutations per site. One time unit is then the expected time to accumulate one mutation per site. At this time scale, the coalescence rate for two sequences sampled from a diploid species with population size N is $\frac{2}{\theta}$ (i.e., $\frac{1}{2N}$ per generation), where $\theta = 4N\mu$ is known as the population size parameter, with μ to be the mutation rate per site per generation. Also the 'coalescent time unit' or the expected waiting time for two sequences to coalesce is $2N$ generations or $\frac{\theta}{2}$ mutations per site. The MSC model with no gene flow is the null hypothesis (H_0) with parameters τ and θ (Fig. 1A). In the alternative hypothesis (H_1), introgression occurs from species A to B (i.e., from X to Y) at time $\tau_X \equiv \tau_Y$, with the introgression probability $\varphi \equiv \varphi_{XY}$ defined as the proportion of immigrants in the recipient population Y from the donor population X (Fig. 1B). Here we let time run forward when we define parameters. Both H_0 and H_1 involve a population size parameter θ, which is dropped in our notation. The parameter vector is thus $\boldsymbol{\theta}_0 = (\tau)$ for H_0 and $\boldsymbol{\theta} = (\varphi, \tau_Y, \tau_R)$ for H_1 (Fig. 1A&B).

While H_0 is clearly nested within H_1, the test comparing the two models involves several non-standard features.

1. There exist three different routes via which H_1 is reduced to H_0 (by fixing the parameter of interest to its null value): (a) $\varphi = 0$, (b) $\tau_Y = \tau_R$, and (c) $\varphi = 1$ (Fig. 1C).
2. For each route, certain parameters become unidentifiable or indeterminate when the parameter of interest takes the null value. For example, for route (a), H_1 becomes H_0 when $\varphi = 0$, but when $\varphi = 0$, the introgression time (τ_Y) becomes unidentifiable; if there is no introgression, the time of introgression is irrelevant (Fig. 1C, table 1A).

Whereas in regular cases H_0 corresponds to one point in the parameter space for H_1, here H_0 corresponds to three lines in H_1 (Fig. 1D&E). Under such nonstandard conditions neither of the proofs of Dickey (1971) and Ji *et al.* (2023) works. Our first objective in this paper is thus to extend the theory for computing the Bayes factor via the S-D density ratio to nonstandard conditions.

Note that the irregularities discussed here affect the use of the S-D density ratio to calculate the Bayes factor, not the Bayes factor itself. However methods for calculating the Bayes factor via marginal likelihood (which are not affected by the irregularities discussed here) involve many MCMC runs and >100 times more computation than the S-D approach (Fourment *et al.*, 2020, and see Discussion). We study simpler problems of Gaussian mixtures to develop the theory.

Our second objective is to study the behavior of the MCMC algorithm, influenced by the irregularities just mentioned. Figure 1F&G shows posterior summaries and trace plots for parameters in H_1 from a typical BPP analysis of a large dataset simulated under the null model of no gene flow (H_0, Fig. 1A). As the data are informative, the MCMC spends most of the time visiting the neighborhoods of the three lines of Figure 1D which represent the true model H_0. However, do there exist multiple modes on the posterior surface (as the marginal summary of φ in Fig. 1F seems to suggest)? How often does the chain visit the neighborhoods of the three lines? How should one summarize the MCMC sample and what biological conclusions should one draw from such posterior samples? Our effort in this direction has not been completely successful, but we characterize the parameter space and posterior surface under H_1, and the probabilities with which the MCMC visits different regions of the parameter space for H_1 when large datasets are generated under H_0 and analyzed under H_1. We provide guidelines on Bayesian inference when gene flow between sister lineages is absent or occurs at very low rates. A recent simulation examined the robustness of the Bayesian test of gene flow between sister lineages to recombination and selection (Thawornwattana *et al.*, 2026).

Our third objective is to evaluate the power of the Bayesian test and the precision of Bayesian estimation of the introgression probability, assessing the information content in multilocus sequence data concerning introgression between sister lineages.

Our paper is structured as follows. First we develop a theory for applying the S-D approach to computing the Bayes factor for testing introgression between sister lineages. We develop and illustrate the theory through simple examples, for which numerical calculation is efficient. We prove that when multiple routes exist for reducing H_1 to H_0, the local behavior of each route can be used to calculate the Bayes factor via the S-D approach. We study the posterior of parameters in the introgression model when large datasets are generated under the null model of no gene flow and suggest that the posterior probabilities for different regions of the parameter space may not converge to a point but rather to a non-degenerate probability distribution, when the data-size approaches infinity. We also highlight the extreme sensitivity of the posterior to the prior when there is no or little gene flow in the data. Finally we conduct a simulation study and an asymptotic analysis (treating data of infinite size) to examine the power of the Bayesian test of introgression and the precision of parameter estimation. We find that information content is greatly influenced by the species split time, with more information at deeper divergence. We discuss the impact of the irregularities of the testing problem on frequentist tests of introgression, such as the likelihood ratio test (LRT) based on the composite likelihood used in ∂a∂i (Gutenkunst *et al.*, 2009) and FASTSIMCOAL2 (Excoffier *et al.*, 2013).

Theory

Savage-Dickey (S-D) density ratio under nonstandard conditions

We consider the general case first and illustrate the theory using two simple examples involving Gaussian mixtures, before dealing with the Bayesian test of introgression between sister species.

Let the vector of parameters be $\boldsymbol{\theta}_0$ with the prior $\pi_0(\boldsymbol{\theta}_0)$ under H_0, and $\boldsymbol{\theta}$ with the prior $\pi(\boldsymbol{\theta})$ under H_1. Given data $\boldsymbol{x}$, let the likelihood be $L_0(\theta_0) = p_0(\boldsymbol{x}|\theta_0)$ under H_0 and $L(\boldsymbol{\theta}) = p(\boldsymbol{x}|\boldsymbol{\theta})$ under H_1. Here we use the subscript 0 for densities under H_0 and no subscript for those under H_1. The Bayes factor in support of H_1 over H_0 is defined as

$$B_{10} = \frac{m}{m_0}, \tag{3}$$

where m and m_0 are the marginal likelihood values under the two models:

$$\begin{aligned} m &= p(\boldsymbol{x}) = \int \pi(\boldsymbol{\theta}) L(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta}, \\ m_0 &= p_0(\boldsymbol{x}) = \int \pi_0(\boldsymbol{\theta}_0) L_0(\boldsymbol{\theta}_0)\, \mathrm{d}\boldsymbol{\theta}_0. \end{aligned} \tag{4}$$

We recognize three types of parameters in H_1, with $\boldsymbol{\theta} = (\omega, \lambda, \xi)$, where ω are parameters of interest, which are constrained to formulate the null hypothesis (H_0); λ are nuisance parameters, which are shared between H_0 and H_1 and usually have the same interpretation in the two models; and ξ are unidentifiable/indeterminate parameters, which become unidentifiable when H_1 is reduced to H_0.

In simple or regular cases, H_1 reduces to H_0 when the parameters of interest take special values that represent no effect (called *null value* or *null point*), that is, $\omega = \omega_0$. For example, in the two-sample test (see SI text 1), samples are taken from two populations $N(\mu_1, 1)$ and $N(\mu_2, 1)$ to test the null hypotehesis $H_0 : \mu_1 = \mu_2 = \mu$. The parameter vector is $\boldsymbol{\theta}_0 = (\mu)$ in H_0, and $\boldsymbol{\theta} = (\mu_1, \mu_2)$ in H_1, and H_1 is reduced to H_0 when $\mu_1 = \mu_2 = \mu$. The parameter space for H_1 is the 2-D μ_1-μ_2 plane, in which the diagonal line $\mu_1 = \mu_2$ represents the parameter space for H_0 (which we refer to as the null space), while a point on the diagonal line (e.g., $\mu_1 = \mu_2 = 0.5$) is the null point or null value.

In contrast, in the test of introgression, there exist three routes from H_1 to H_0 (a, b, c in Fig. 1C), and the null point is not one point in the parameter space of H_1 but rather corresponds to three lines or line segments, referred to later simply as segments (a, b, c in Fig. 1D&E). The null space (the parameter space for H_0) corresponds to the three planes in Figure 1D&E.

Furthermore, the definition of the parameters of interest differ among the different segments, and as a result, the partitioning of the parameter space, $\boldsymbol{\theta} = (\omega, \lambda, \xi)$, differs depending on the route (table 1(i)).

We show that with multiple routes for reducing H_1 to H_0, each can be used to construct an S-D density ratio to compute B_{10}.

Theorem 1. *Suppose the null point (a point in the parameter space of H_0) corresponds to multiple segments in the parameter space of H_1 and at segment j, we partition the parameters as $\boldsymbol{\theta} = (\omega_j, \lambda_j, \xi_j)$, where ω_j are parameters of interest (with $\omega_j = \omega_{j0}$ being the null value), λ_j the nuisance parameters and ξ_j the unidentifiable parameters. On the condition that*

$$\pi_0(\lambda_j) = \pi(\lambda_j|\omega_{j0}), \tag{5}$$

where $\pi(\lambda_j|\omega_{j0}) = \int \pi(\lambda_j, \xi_j|\omega_{j0})\,\mathrm{d}\xi_j$, the Bayes factor B_{10} of eq. 3 is given as

$$B_j = \frac{\pi(\omega_{j0})}{\pi(\omega_{j0}|\boldsymbol{x})} = B_{10}. \tag{6}$$

Proof. Note that H_0 and H_1 are nested, so that $L(\omega_{j0}, \lambda_j, \xi_j) = L_0(\lambda_j)$. Thus

$$\begin{aligned} B_j &= \frac{\pi(\omega_{j0})}{\iint \pi(\omega_{j0}, \lambda_j, \xi_j|\boldsymbol{x})\,\mathrm{d}\lambda_j\,\mathrm{d}\xi_j} \\ &= \frac{\pi(\omega_{j0})}{\iint \frac{1}{p(\boldsymbol{x})}\pi(\omega_{j0})\pi(\lambda_j, \xi_j|\omega_{j0})L(\omega_{j0}, \lambda_j, \xi_j)\,\mathrm{d}\lambda_j\,\mathrm{d}\xi_j} \\ &= \frac{m\pi(\omega_{j0})}{\iint \pi(\omega_{j0})\pi(\lambda_j, \xi_j|\omega_{j0})L(\omega_{j0}, \lambda_j, \xi_j)\,\mathrm{d}\lambda_j\,\mathrm{d}\xi_j} \\ &= \frac{m}{\iint \pi(\lambda_j, \xi_j|\omega_{j0})L_0(\lambda_j)\,\mathrm{d}\lambda_j\,\mathrm{d}\xi_j} \\ &= \frac{m}{\int \pi_0(\lambda_j)L_0(\lambda_j)\,\mathrm{d}\lambda_j} \\ &= \frac{m}{m_0}. \end{aligned} \tag{7}$$

□

This proof resembles the reverse of the proof of Ji *et al.* (2023, eq. A1). However, here the argument concerns the local behavior of the densities around segment j. While the marginal likelihood values (m_0, m in eq. 4) are defined by averaging over the whole parameter space in each model, the Bayes factor is given here by the local behavior of the posterior under H_1 around the null point (or each of the segments that constitute the null point).

More generally, H_0 may be specified by setting up equality constraints on parameters in H_1, that is $c(\boldsymbol{\theta}) = 0$, rather than letting parameters of interest take the null value ($\omega = \omega_0$). For example in the two-sample test we have $\boldsymbol{\theta} = (\mu_1, \mu_2)$, so the constraint is $c(\boldsymbol{\theta}) = \mu_1 - \mu_2 = 0$. As the parameters of interest depend on the segment, we use $\boldsymbol{\theta}$ instead of ω_j to specify constraints. Suppose there are J routes of going from H_1 to H_0. For route j, let there be K_j equality constraints, $c_j = (c_{j1}, \cdots, c_{jK_j})$, so that segment j is represented by

$$\Omega_{j0} = \left\{\boldsymbol{\theta} : \bigcap_k c_{jk}(\boldsymbol{\theta}) = 0\right\}. \tag{8}$$

For simplicity, we use K instead of K_j but note that K may differ among segments (one such case is the test of bidirectional introgression; see table 1(ii)). We use infinitesimals $\epsilon_{jk} > 0$, $j = 1, \cdots, J; k = 1, \cdots, K$, to define a *null region*, a subspace of the parameter space of H_1 in the neighborhood of segment j,

$$\Omega_{j,\epsilon} = \left\{\boldsymbol{\theta} : \bigcap_k |c_{jk}(\boldsymbol{\theta})| < \epsilon_{jk}\right\}. \tag{9}$$

Let $\epsilon_j = (\epsilon_{j1}, \cdots, \epsilon_{jK})$. Define

$$\begin{aligned} \pi(\Omega_{j0}) &= \lim_{\epsilon_j \to 0} \frac{\mathbb{P}(\Omega_{j,\epsilon})}{\epsilon_{j1}\cdots\epsilon_{jK}}, \\ \pi(\Omega_{j0}|\boldsymbol{x}) &= \lim_{\epsilon_j \to 0} \frac{\mathbb{P}(\Omega_{j,\epsilon}|\boldsymbol{x})}{\epsilon_{j1}\cdots\epsilon_{jK}}, \end{aligned} \tag{10}$$

to be the prior and posterior densities for Ω_{j0} under H_1. Then

$$B_j = \frac{\pi(\Omega_{j0})}{\pi(\Omega_{j0}|\boldsymbol{x})} \approx \frac{\mathbb{P}(\Omega_{j,\epsilon})}{\mathbb{P}(\Omega_{j,\epsilon}|\boldsymbol{x})} \tag{11}$$

is an approximation of B_{10}.

One may reparametrize H_1 and treat $c_j \equiv \omega_j$ as parameters, and derive the induced prior on the new parameters. Then the Bayes factor for segment j of eq. 11 takes the simpler form of eq. 6. Thus we do not see any conceptual difference between eqs. 6 and 11, and will henceforth use the notation of eq. 11, possibly with reparametrization used to verify the condition on priors (eq. 5) (see Remark 2 below). Note that the number of constraints K in eqs. 10&11, which may be considered the degree of freedom for the test, equals the dimension of ω_j in eq. 6.

We will use the differentials ϵ_j to calculate $\mathbb{P}(\Omega_{j,\epsilon}|\boldsymbol{x})$ for segment j by processing an MCMC sample under H_1 (eq. 11) (Ji *et al.*, 2023). One may also use kernel density smoothing to calculate $\pi(\Omega_{j0}|\boldsymbol{x})$. The prior density $\pi(\Omega_{j0})$ is typically tractable analytically.

If the condition on priors holds for each segment, so that all segments give the same S-D density ratio, with $B_j = B_{10}$ (eqs. 6 or 11), we can use all segments to calculate a combined density ratio. Let

$$\Omega_0 = \bigcup_j \Omega_{j0}, \quad \Omega_\epsilon = \bigcup_j \Omega_{j,\epsilon} \tag{12}$$

be the *null point* (or constraints) and *null region*, respectively. Then

$$B = \frac{\pi(\Omega_0)}{\pi(\Omega_0|\boldsymbol{x})} \approx \frac{\mathbb{P}(\Omega_\epsilon)}{\mathbb{P}(\Omega_\epsilon|\boldsymbol{x})}. \tag{13}$$

We make two observations regarding the condition on priors of eq. 5.

First, if the condition on priors (eq. 5) does not hold, a correction factor may be applied

$$B_j = \frac{\pi(\omega_{j0})}{\pi(\omega_{j0}|\boldsymbol{x})} \Big/ \mathbb{E}\left\{\frac{\pi_0(\lambda_j)}{\pi(\lambda_j|\omega_{j0})}\right\}, \tag{14}$$

with the expectation taken over $\pi(\lambda_j|\boldsymbol{x}, \omega_{j0})$ (Verdinelli and Wasserman, 1995). This does not appear simpler to compute than the marginal likelihood values (m_0, m), and is thus not pursued here. Instead later we discuss ideas for approximate calculation of B_{10} when the condition on priors does not hold.

Second, we note that the condition on priors (eq. 5) may hold for one parametrization but not another. Consider for example the one-sample test, in which $H_0 : N(0, \sigma^2)$ is tested against $H_1 : N(\mu, \sigma^2)$ with unknown variance σ^2. The parameter of interest is μ with the null value $\mu_0 = 0$ while σ^2 is the nuisance parameter. Conditioning on $\mu = 0$ or on $\zeta = \frac{\mu}{\sigma} = 0$ leads to different distributions: $p(\sigma^2|\mu = 0) \neq p(\sigma^2|\zeta = 0)$ (Fig. S1 and SI text 2). This phenomenon is known as the Borel-Kolmogorov paradox, in which conditioning on seemingly equivalent events leads to different conditional distributions. It is prudent to verify the condition on priors when calculating the S-D density ratio.

Note that the issue concerns the S-D approach to calculating the Bayes factor, not the Bayes factor itself. The marginal likelihood values are not affected by reparametrization, so that B_{10} is well-defined and is the same for different parametrizations.

Example GM1. Gaussian mixture model for testing contamination

Problem setup. Here we consider a test of contamination under a Gaussian mixture model. This is not tractable analytically but yields to numerical calculation. It has most of the irregularities discussed above for the test of introgression, including multiple routes from H_1 to H_0, so we use it to illustrate the application of theorem 1 to calculate B_{10}. We also use this example to study the posterior surface under H_1 under nonstandard conditions.

The data consist of an i.i.d. sample, $\boldsymbol{x} = (x_i)$, $i = 1, \cdots, n$. The likelihood model specifies $x_i \sim N(0, 1)$ under H_0, or $x_i \sim \alpha N(\mu, 1) + (1 - \alpha)N(0, 1)$ under H_1, with probability α for contamination from $N(\mu, 1)$. Thus H_1 is a mixture of two Gaussian distributions. We assign the prior $p(\alpha, \mu) = \phi(\mu; 0, 1)$ under H_1, where $\phi(z; \mu, \sigma^2)$ is the probability density function (PDF) for the Gaussian distribution $N(\mu, \sigma^2)$. There are no nuisance parameters, so the condition on priors (eq. 5) holds.

The marginal likelihoods under the two models are

$$m_0 = \mathrm{e}^{-\frac{1}{2}\sum x_i^2},$$
$$m = \int_0^1 \int_{-\infty}^{\infty} \frac{\mathrm{e}^{-\frac{1}{2}\mu^2}}{\sqrt{2\pi}} \prod_i \left[\alpha\, \mathrm{e}^{-\frac{1}{2}(x_i-\mu)^2} + (1-\alpha)\, \mathrm{e}^{-\frac{1}{2}x_i^2}\right] \mathrm{d}\mu\, \mathrm{d}\alpha, \tag{15}$$

so that

$$B_{10} = \int_0^1 \int_{-\infty}^{\infty} \frac{\mathrm{e}^{-\frac{1}{2}\mu^2}}{\sqrt{2\pi}} \prod_i \left[1 - \alpha + \alpha\, \mathrm{e}^{-\frac{1}{2}(\mu^2 - 2\mu x_i)}\right] \mathrm{d}\mu\, \mathrm{d}\alpha. \tag{16}$$

We use m_0 to scale the likelihood under H_1, and recognise the integrand in eq. 16,

$$p(\alpha, \mu, \boldsymbol{x}) = \frac{\mathrm{e}^{-\frac{1}{2}\mu^2}}{\sqrt{2\pi}} \prod_i \left[1 - \alpha + \alpha\, \mathrm{e}^{-\frac{1}{2}(\mu^2 - 2\mu x_i)}\right], \tag{17}$$

as the unnormalized posterior. Then the marginal likelihood for H_1,

$$M = \int_0^1 \int_{-\infty}^{\infty} p(\alpha, \mu, \boldsymbol{x})\, \mathrm{d}\mu\, \mathrm{d}\alpha, \tag{18}$$

is also the Bayes factor: $M = \frac{m}{m_0} = B_{10} \equiv B$.

The S-D approach to calculating B_{10}. The constraint that reduces H_1 to H_0 is

$$\Omega_0 = (\alpha = 0) \cup (\mu = 0). \tag{19}$$

The null point comprises the two lines or axes of Figure 2A: (a) $\alpha = 0$ (with μ unidentifiable) and (b) $\mu = 0$ (with α unidentifiable). Define the null region

$$\Omega_\epsilon = (\alpha < \epsilon_\alpha) \cup (|\mu| < \epsilon_\mu), \tag{20}$$

which are two narrow strips along the axes, of widths ϵ_α and $2\epsilon_\mu$, respectively (Fig. 2A). Each strip can be used to calculate the Bayes factor, giving

$$\begin{aligned} B_\alpha(\epsilon_\alpha) &= \frac{\mathbb{P}(\alpha < \epsilon_\alpha)}{\mathbb{P}(\alpha < \epsilon_\alpha \mid \boldsymbol{x})} = \frac{\epsilon_\alpha}{\mathbb{P}(\alpha < \epsilon_\alpha \mid \boldsymbol{x})}, \\ B_\mu(\epsilon_\mu) &= \frac{\mathbb{P}(|\mu| < \epsilon_\mu)}{\mathbb{P}(|\mu| < \epsilon_\mu \mid \boldsymbol{x})} = \frac{1/\sqrt{2\pi} \times 2\epsilon_\mu}{\mathbb{P}(|\mu| < \epsilon_\mu \mid \boldsymbol{x})}, \end{aligned} \tag{21}$$

where the prior probabilities in the numerator are given by the uniform and Gaussian priors, while the posterior probabilities in the denominator can be estimated using an MCMC sample or calculated numerically using Gaussian quadrature.

We also use both segments to produce a combined estimate, $B_{10} \approx \mathbb{P}(\Omega_\epsilon)/\mathbb{P}(\Omega_\epsilon|\boldsymbol{x})$.

The shape of the posterior surface. Next we study the posterior for parameters $\boldsymbol{\theta}$ in H_1 when large datasets generated under H_0 are analyzed under H_1. In a regular case, the posterior is expected to be concentrated around $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ (the null value or the truth), with $\boldsymbol{\theta}|\boldsymbol{x} \sim N(\boldsymbol{\theta}_0, (nI)^{-1})$, where $\mathcal{I}$ is the Fisher information matrix, when $n \to \infty$ (O'Hagan and Forster, 2004, p.72-74). Here we expect the posterior to be concentrated around the two segments (the α- and μ-axes) which constitute the null value. We are interested in the posterior probabilities for the two subregions around the two axes (Fig. 2A). They represent the proportional contributions of the two subregions to the marginal likelihood (m), and are also the proportions of times that the MCMC algorithm visits those subregions.

Let $\epsilon_\alpha > 0$ and $\epsilon_\mu > 0$ be fixed small values. We calculate the posterior probabilities for the two subregions $P_\alpha = \mathbb{P}(\alpha < \epsilon_\alpha|\boldsymbol{x})$ and $P_\mu = \mathbb{P}(|\mu| < \epsilon_\mu|\boldsymbol{x})$ using Gaussian quadrature. When $n \to \infty$, both P_α and P_μ will increase, with $P_\alpha + P_\mu \to 1$. We suggest that as $n \to \infty$, P_α may not converge to a point value but instead vary among datasets (all generated under H_0) according to a non-degenerate distribution. Here we provide without proof an intuitive argument through analogy and generate the distribution of P_α numerically.

The problem appears to be similar to the fair-coin or fair-balance paradox, the star-tree paradox (Lewis *et al.*,

2005; Yang and Rannala, 2005; Yang, 2007) and in general comparison of equally correct and nondistinct models (Yang and Zhu, 2018, Fig. B2). In the fair-balance paradox, a balance is fair but we test whether it has a negative or positive bias (Yang and Rannala, 2005). We compare two models, $H_- : \mu < 0$ and $H_+ : \mu \geq 0$ when the truth is $\mu = \mu_0 = 0$. The data are an i.i.d. sample of size n, generated from $x_i \sim N(0,1)$ and summarized as the sample mean $\bar{x}$. With the prior $\mu \sim N(0,1)$, the posterior model probability is

$$P_- = \mathbb{P}(H_-|\bar{x}) \approx \Phi\big(-\sqrt{n}\bar{x}\big), \tag{22}$$

where Φ is the cumulative distribution function (CDF) for $N(0,1)$ (Yang and Rannala, 2005, eq. 6). As $\sqrt{n}\bar{x}$ is an $N(0,1)$ variable, P_- has a $U(0,1)$ distribution.

As n is large, $\bar{x}$ is very close to $\mu_0 = 0$ in essentially every dataset. However a small deviation of $\bar{x}$ from 0 can translate to a large departure of P_- from $\frac{1}{2}$. Let $n = 10^6$, say. Then in 2.5% of datasets $\bar{x} < -0.00196$ so that $P_- > 0.975$, while in another 2.5% of datasets $\bar{x} > 0.00196$ so that $P_- < 0.025$. Indeed as $\bar{x}$ varies among datasets according to $\bar{x} \sim N\big(0, \frac{1}{n}\big)$, P_- varies like a random number from $U(0,1)$.

Note that the Bayesian test of $H_0 : N(0,1)$ against $H_1 : N(\mu,1)$, with the prior $\mu \sim N(0,1)$, produces

$$B_{10} = \tfrac{1}{\sqrt{1+n}} \exp\Big\{\tfrac{1}{2} n\bar{x}^2 \cdot \tfrac{n}{1+n}\Big\}. \tag{23}$$

As $n\bar{x}^2 \sim \chi^2_1$, $B_{10} \to 0$ when $n \to \infty$. For example, at $n = 10^6$ and $\sqrt{n}|\bar{x}| = 1.96$, we have $B_{10} = 0.00682$, strongly supporting H_0. However, $P_- = \mathbb{P}(\mu < 0|\bar{x}, H_1)$ fluctuates among datasets like a random number.

In the Gaussian-mixture model of Figure 2, when n is large enough the posterior probability outside the two strips (represented by Ω_ϵ) will be ~ 0 with $P_\alpha + P_\mu \approx 1$. The two strips are like the parameter space for the two models H_- and H_+ in the fair-balance example, with MLEs under H_1 close to the α and μ axes corresponding to $\bar{x} < 0$ or $\bar{x} > 0$. Thus the posterior surface under H_1 may be expected to vary among replicate datasets (all simulated under H_0) in terms of the probability mass above the two strips. Figure 2D&E shows the unnormalized posterior for two datasets, each of size $n = 10^6$ simulated under H_0. In the first, the MLE is close to the α-axis, while in the second, the MLE is close to the μ-axis. Note that in all large datasets, the MLEs are close to the two axes, supporting the null model of no contamination (Fig. 2B).

We simulated 1000 datasets for a few large n values (from 10^6 to 5×10^7) under $H_0 : N(0,1)$, and calculated P_α, P_μ, with the histogram for the rescaled probability, $P'_\alpha = P_\alpha/(P_\alpha + P_\mu)$, shown in Figure 2C. Even with very large n, P'_α varies among datasets according to a distribution rather than converging to a point value.

Example GM2. Another Gaussian-mixture model

Problem setup. We modify the previous example slightly, so that H_0 specifies one population with mean μ while H_1 specifies two populations with means μ_1 and μ_2 in proportions α and $1-\alpha$ (Fig. 3A). This example has all the irregularities encountered in the test of unidirectional introgression of Figure 1A&B and indeed the two problems have the same geometrical structure (compare Fig. 3A with Fig. 1D). This example is not analytically tractable but the MCMC algorithm involves much less computation than the test of introgression (compare Fig. 3B–D with Fig. 1F&G).

The parameter vector is $\boldsymbol{\theta}_0 = (\mu)$ for H_0 and $\boldsymbol{\theta} = (\alpha, \mu_1, \mu_2)$ for H_1 (Fig. 3A). This is again a test for mixture. The data are an i.i.d. sample, $\boldsymbol{x} = (x_i)$, with $x_i \sim N(\mu,1)$ under H_0, or $x_i \sim \alpha N(\mu_1,1) + (1-\alpha)N(\mu_2,1)$ under H_1. The prior under H_0 is $p_0(\mu) = \phi(\mu; 0, \frac{1}{\tau})$, with $\tau = 1$. Under H_1 we assign the prior $U(0,1)$ for α and $N(0, \frac{1}{\tau})$ for μ_1 and μ_2, so that

$$p(\alpha, \mu_1, \mu_2) = \frac{1}{2\pi/\tau} \exp\Big\{-\frac{\tau}{2}\big(\mu_1^2 + \mu_2^2\big)\Big\}. \tag{24}$$

If we re-parametrize H_1 and use $\boldsymbol{\theta}' = (\alpha, \mu_1, \delta)$ instead, with $\delta = \mu_2 - \mu_1$, the prior becomes $p(\alpha, \mu_1, \delta) = \phi(\mu_1; -\frac{\delta}{2}, \frac{1}{2\tau})\phi(\delta; 0, \frac{2}{\tau})$.

Under H_1, the density for x_i is

$$p(x_i; \boldsymbol{\theta}) = \tfrac{1}{\sqrt{2\pi}}\Big[\alpha\, \mathrm{e}^{-\frac{1}{2}(x_i-\mu_1)^2} + (1-\alpha)\, \mathrm{e}^{-\frac{1}{2}(x_i-\mu_2)^2}\Big], \tag{25}$$

so that the posterior is

$$p(\boldsymbol{\theta} \mid \boldsymbol{x}) \propto \mathrm{e}^{-\frac{1}{2}(\mu_1^2+\mu_2^2)} \prod_i p(x_i; \boldsymbol{\theta}). \tag{26}$$

There are three routes from H_1 to H_0, and the constraints that reduce H_1 to H_0 are

$$\Omega_0 = (\alpha = 0) \cup (\mu_1 = \mu_2) \cup (\alpha = 1). \tag{27}$$

The null point corresponds to three lines or line segments in the parameter space of H_1: (a) $\alpha = 0$ (with μ_1 unidentifiable), (b) $\mu_1 = \mu_2$ (with α unidentifiable), and (c) $\alpha = 1$ (with μ_2 unidentifiable). The parameter space for H_0 (the null space) corresponds to three planes in the parameter space of H_1 (Fig. 3A).

We note that the condition on priors (eq. 5) holds for segments (a) $\alpha = 0$ and (c) $\alpha = 1$, but not for segment (b) $\mu_1 = \mu_2$ or $\delta = \mu_2 - \mu_1 = 0$. For segment (b), we have $p_0(\mu) = \phi(\mu; 0, \frac{1}{\tau})$ for H_0, and $p(\mu_1 \mid \delta_0) = \phi(\mu_1; 0, \frac{1}{2\tau})$ for H_1, so there is a mis-match.

Bayes factor via the S-D density ratio. We use the differentials ϵ_α and $\epsilon_{\mu_1} = \epsilon_{\mu_2} = \epsilon_\mu$ to define a null region in the parameter space of H_1,

$$\begin{aligned}\Omega_\epsilon &= \Omega_{\epsilon,a} \cup \Omega_{\epsilon,b} \cup \Omega_{\epsilon,c} \\ &= (\alpha < \epsilon_\alpha) \cup (|\mu_2 - \mu_1| < \epsilon_\mu) \cup (1 - \alpha < \epsilon_\alpha).\end{aligned} \tag{28}$$

If we use each segment to calculate B_{10}, we have

$$B_a = \frac{\mathbb{P}(\Omega_{\epsilon,a})}{\mathbb{P}(\Omega_{\epsilon,a} \mid \boldsymbol{x})} = \frac{\epsilon_\alpha}{\mathbb{P}(\alpha < \epsilon_\alpha \mid \boldsymbol{x})},$$
$$B_b = \frac{\mathbb{P}(\Omega_{\epsilon,b})}{\mathbb{P}(\Omega_{\epsilon,b}|\boldsymbol{x})} = \frac{1/\sqrt{\pi/\tau} \times 2\epsilon_\mu}{\mathbb{P}(|\mu_1 - \mu_2| < \epsilon_\mu \mid \boldsymbol{x})}, \quad (29)$$
$$B_c = \frac{\mathbb{P}(\Omega_{\epsilon,c})}{\mathbb{P}(\Omega_{\epsilon,c} \mid \boldsymbol{x})} = \frac{\epsilon_\alpha}{\mathbb{P}(1 - \alpha < \epsilon_\alpha \mid \boldsymbol{x})}.$$

We use the MCMC sample of Figure 3D to calculate the S-D density ratio with different differentials ϵ_α and ϵ_μ (Fig. S2B). We expect B_a and B_c to approximate B_{10} but B_b may be off by a factor, since the condition on priors holds for segments a and c but not for b. However, for this dataset the three segments produced similar B_{10} values, around 0.25, with the data preferring the null model of no mixture (i.e., the true model).

If we use all segments to calculate a combined S-D density ratio, the prior probability for the null region under H_1 will be

$$\mathbb{P}(\Omega_\epsilon) = \mathbb{P}\{\alpha < \epsilon_\alpha \cup |\mu_2 - \mu_1| < \epsilon_\mu \cup 1 - \alpha < \epsilon_\alpha) \approx p(\alpha = 0)\epsilon_\alpha + p(\delta = 0) \cdot 2\epsilon_\mu + p(\alpha = 1)\epsilon_\alpha, \quad (30)$$

while the posterior probability $\mathbb{P}(\Omega_\epsilon|\boldsymbol{x})$ can be calculated by processing the MCMC sample under H_1. As the different segments produced similar B_{10} estimates in the example of Figure 3D (Fig. S2B), the combined estimate is similar as well.

The shape of the posterior surface. Under H_1, $\boldsymbol{\theta} = (\alpha, \mu_1, \mu_2)$ and $\boldsymbol{\theta}' = (1 - \alpha, \mu_2, \mu_1)$ are unidentifiable and this is an unidentifiability of the label-switching type. As we use a symmetrical prior $\alpha \sim U(0, 1)$, the posterior $p(\alpha|\boldsymbol{x})$ around segments a and c are mirror images of each other. We thus consider the relative posterior probabilities around segments a and c versus segment b (Fig. 3A). Rousseau and Mengersen (2011) analyzed the posterior under an over-fitting finite-mixture model with too many components when large datasets are generated under the null model and when the mixing probabilities are assigned a Dirichlet prior. Example GM2 is case 1 analyzed in Rousseau and Mengersen (2011, Fig. 1). The dynamics is found to depend on the parameters of the Dirichlet prior in relation to the number of parameters in each component (d). If $a < d/2$, the posterior probability that the extra component has non-negligible probability (averaged over datasets) is small, or in other words the posterior will be dominated by 'emptying' a component (with either α or $1 - \alpha$ being ≈ 0, i.e., segments a or c in Fig. 3A). If $a > d/2$ the posterior will be dominated by 'merging' components (with $\mu_1 = \mu_2$ or segment b in Fig. 3A). The analysis highlights the influence of the prior on the mixing probabilities on the posterior. In our example $a = 1$ and there is one parameter in each component (e.g., μ_1 for the first component), so that $a > d/2$. The authors simulated small datasets only with $n \leq 10^3$. Our simulation of larger datasets ($n = 10^6$) shows a U-shaped posterior for α, and segment b does not dominate the posterior. We leave it to future research to investigate the impact of the data and prior on the posterior when data are generated under H_0 and analyzed under H_1.

Test of unidirectional introgression (UDI) between sister lineages

The test of introgression (under the UDI model) has been introduced in Introduction (Fig. 1; see also Fig. 4A&B). The null model H_0 assumes no gene flow, with the parameter vector $\boldsymbol{\theta}_0 = (\tau, \theta)$, while the alternative model H_1 assumes unidirectional introgression, with the parameter vector $\boldsymbol{\theta} = (\varphi, \tau_Y, \tau_R, \theta)$ (Fig. 1B). Here we assume the same population size θ for all species on the phylogeny, which has the same prior under H_0 and H_1. In models with different population sizes, the θ parameters have the same priors in the two models as well. There are three routes from H_1 to H_0, and the null point (one point in the parameter space of H_0) corresponds to three lines or line segments in the parameter space of H_1 given by the constraints

$$\Omega_0 = (\varphi = 0) \cup (\tau_Y = \tau_R) \cup (\varphi = 1) \quad (31)$$

(table 1(i), Fig. 1D). The null space (the parameter space for H_0) is represented by the three planes in Figure 1D.

In BPP, we assign a beta prior $\varphi \sim \text{Beta}(a, b)$, a gamma-Dirichlet prior on the split times (Yang and Rannala, 2010, eq. 2), which means that $\tau_R \sim G(\alpha_\tau, \beta_\tau)$, and given τ_R, $\beta = \tau_Y/\tau_R$ is $U(0, 1)$. We note that the condition on priors (eq. 5) holds for segments a and b, but not for segment c (table 1(i)). For segments a and b, the nuisance parameters are τ_R and θ), each with the same gamma prior in both H_0 and H_1. For segment c, the nuisance parameter τ_Y in H_1 is mapped onto τ in H_0, with different priors. In H_0 the prior is $\tau \sim G(\alpha, \beta)$, with the prior mean α/β. In H_1, we have $\tau_Y|\tau_R \sim U(0, 1)$ with $\tau_R \sim G(\alpha, \beta)$ so that

$$p(\tau_Y) = \int_0^\infty \frac{1}{\tau_R} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \tau_R^{\alpha-1} \, \mathrm{e}^{-\beta\tau_R} \, \mathrm{d}\tau_R = \frac{\beta}{\alpha-1}\left[1 - G(\tau_Y, \alpha - 1, \beta)\right], \quad (32)$$

where $G(y, \alpha, \beta)$ is the CDF for the gamma distribution $G(\alpha, \beta)$ (Flouri *et al.*, 2022). This has the prior mean $\alpha/(2\beta)$.

From eq. 31, we define the null region

$$\Omega_\epsilon = (\varphi < \epsilon_\varphi) \cup (1 - \beta < \epsilon_\beta) \cup (1 - \varphi < \epsilon_\varphi) \quad (33)$$

to calculate the Bayes factor, $B_{10} \approx \frac{\mathbb{P}(\Omega_\epsilon)}{\mathbb{P}(\Omega_\epsilon|\boldsymbol{x})}$, where $\mathbb{P}(\Omega_\epsilon) = \epsilon_\varphi + \epsilon_\beta + \epsilon_\varphi$ while $\mathbb{P}(\Omega_\epsilon|\boldsymbol{x})$ is the proportion of MCMC samples under the UDI model that fall in the null region (Ω_ϵ).

We use this approach in our simulation below to evaluate the performance of the Bayesian test, setting the differentials to be 0.01σ with σ to be the prior standard deviation; in other words we use $\epsilon_\varphi = \epsilon_\beta = 0.01 \cdot \frac{1}{\sqrt{12}} = 0.002886$.

The shape of the posterior surface and the impact of the prior. As mentioned in the Introduction, when large datasets generated under the null model of no gene flow (H_0, Fig. 1A) are analyzed under the UDI model (H_1, Fig. 1B), the posterior tends to be concentrated around the three segments (a, b, c in Fig. 1D&E) that collectively

represent the true model of no gene flow. Qualitatively, the asymptotics under the UDI model appears to be similar to that under the mixture models (GM1 and GM2), but it is unclear whether the asymptotic analysis of finite-mixture models by Rousseau and Mengersen (2011) applies to the introgression model. In the special case of sampling one sequence from the recipient species, the UDI model specifies a mixture of gene trees and the likelihood model is a mixture model with mixing probabilities φ and $1-\varphi$ for two components. When two or more sequences are sampled from the recipient species the data are not described by a mixture model anymore (see Meng and Kubatko, 2009 and correction in Jiao *et al.*, 2021).

Similarly to Rousseau and Mengersen (2011), we have observed high sensitivity of the posterior to the prior on the introgression probability (φ) under the UDI model when data are generated with no gene flow. We analyzed the data of Figure 1 under the UDI model using the prior $\varphi \sim \text{Beta}(a, a)$ with $a = \frac{1}{4}$, 1, and 10 (Fig. S3). The Beta(a, a) prior has a U-shape when $a < 1$, is flat when $a = 1$, and has a bell shape when $a > 1$. At $a = \frac{1}{4}$ or 1, the posterior has substantial mass close to segments a and c, with φ close to 0 or 1. At $a = 10$, the posterior shifts to segment b, away from 0 and 1. In Bayesian analyses under regular conditions, the prior typically becomes less and less important when the amount of data increases, and eventually the posterior is entirely dominated by the data when $n \to \infty$. Here the prior has an effect on the posterior even when $n \to \infty$ (Rousseau and Mengersen, 2011). Nevertheless, it should be noted that the impact of the prior is to shift the posterior mass among three segments of Figure 1D&E, all of which represent the same likelihood model and the same biological scenario of no gene flow.

Test of bidirectional introgression (BDI) between sister lineages

Under the BDI model there are two introgression probabilities, so that the parameter vector in H_1 is $\boldsymbol{\theta} = (\varphi_X, \varphi_Y, \tau_Y, \tau_R, \theta)$ (Fig. 4D). The null model H_0 is the MSC model with no gene flow as before, with $\boldsymbol{\theta}_0 = (\tau, \theta)$ for H_0. The common population size parameter θ is a nuisance parameter that exists in both models, with the same prior.

There are five routes by which H_1 can be reduced to H_0. In other words, one point in H_0 (the null point) corresponds to five lines (or plane) specified by equality constraints on four parameters in H_1: $\varphi_X, \varphi_Y, \tau_Y$, and τ_R (table 1(ii)),

$$\begin{aligned}\Omega_0 = &(\varphi_X = 0, \varphi_Y = 0) \cup (\varphi_X = 1, \varphi_Y = 1)\\ &\cup (\tau_Y = \tau_R)\\ &\cup (\varphi_X = 0, \varphi_Y = 1) \cup (\varphi_X = 1, \varphi_Y = 0).\end{aligned} \tag{34}$$

The condition on priors (eq. 5) holds for segments a–c but not for segments d and e (table 1(ii)). For the latter two, the nuisance parameter τ_Y in H_1, which has the prior of eq. 32, is mapped onto τ in H_0, with the prior $\tau \sim G(\alpha, \beta)$, so there is a mismatch.

From eq. 34, define the null region as

$$\begin{aligned}\Omega_\epsilon = &(\varphi_X < \epsilon_\varphi, \varphi_Y < \epsilon_\varphi)\\ &\cup (1-\varphi_X < \epsilon_\varphi, 1-\varphi_Y < \epsilon_\varphi)\\ &\cup (1-\beta < \epsilon_\beta)\\ &\cup (\varphi_X < \epsilon_\varphi, 1-\varphi_Y < \epsilon_\varphi)\\ &\cup (1-\varphi_X < \epsilon_\varphi, \varphi_Y < \epsilon_\varphi).\end{aligned} \tag{35}$$

As in the test of UDI, we set $\epsilon_\varphi = \epsilon_\beta = 0.002886$ to process the MCMC sample from the BDI model to calculate $\mathbb{P}(\Omega_\epsilon|\boldsymbol{x})$. For the prior probability, we have

$$\mathbb{P}(\Omega_\epsilon) = \epsilon_\varphi^2 + \epsilon_\varphi^2 + \epsilon_\beta + \epsilon_\varphi^2 + \epsilon_\varphi^2 \tag{36}$$

We use this approach in the computer simulation below.

Test of gene flow between sister lineages under continuous migration model (MSC-M)

We consider two variants of the continuous MSC-migration (MSC-M) model of gene flow between sisters: the isolation-with-migration (IM) model in which migration has been ongoing until the present time (Fig. 4C) and the isolation-with-initial-migration (IIM) model (Costa and Wilkinson-Herbots, 2017; see also Huang *et al.*, 2022, Fig. 1B), which assumes that gene flow occurs post speciation initially but ceased at a certain time in the past (τ_Y, Fig. 5B&C). The Bayesian test of gene flow under the IM model does not involve the irregularities discussed in this paper but the IIM model involves similar irregularities (table 2).

First we note that version 4.8 of BPP introduced a reparametrization so that the mutation-scaled migration rate $\varpi_{XY} = m_{XY}/\mu = 4M_{XY}/\theta_Y$ is used instead of the population migration rate $M_{XY} = m_{XY}N_Y$, where m_{XY} is the expected proportion of $X \to Y$ immigrants in the recipient population Y, so that M_{XY} and ϖ_{XY} are the expected number of $X \to Y$ immigrants per generation and per mutational time unit, respectively. Gamma priors are assigned to ϖ while θs are assigned independent gamma or inverse-gamma priors.

Under the IM model, H_0 is represented by $\varpi_{AB} = 0$ in H_1. This null value is at the boundary of the parameter space (which is not a problem for either the Bayes factor or the S-D approach for calculating it), but no other parameters become indeterminate. There do not exist other routes from H_1 to H_0. For example, letting $\varphi = \infty$ reduces H_1 to the model of one species (or two species with $\tau_r = 0$) but not H_0, which is the complete-isolation model for two species (Fig. 5A). The bidirectional migration model in the case of two species (BDM, with both ϖ_{AB} and ϖ_{BA}) is also identifiable with data of multiple samples per species per locus, unlike the discrete introgression model (BDI), which shows two unidentifiable modes (Yang and Flouri, 2022).

Under the IIM model, if gene flow is unidirectional (UDM), the situation for the test is very similar to that under the discrete UDI model, and there exist three routes by which H_1 is reduced to H_0: (a) $\varpi = 0$ (with τ_Y unidentifiable), (b) $\tau_Y = \tau_R$ (with ϖ unidentifiable), and (c) $\varpi = 1$ (with τ_R unidentifiable) (Fig. 5A&B, table

2(i)). Similarly to the test under UDI (Fig. 1), we can calculate the Bayes factor using the Savage-Dickey density ratio based on each of the three segments. For example, for segment a, $B_a \approx \frac{\mathbb{P}(\varpi<\epsilon_\varpi)}{\mathbb{P}(\varpi<\epsilon_\varpi|\boldsymbol{x})}$, with ϵ_ϖ to be 0.5% or 0.1% quantile for the gamma prior $\varpi \sim G(\alpha,\beta)$. Segment c may be used to calculate $B_c = \frac{\mathbb{P}(\varpi>E_\varpi)}{\mathbb{P}(\varpi>E_\varpi|\boldsymbol{x})}$, where E_ϖ may be set to the 99.5% or 99.9% quantile for the gamma prior.

Under the IIM model with bidirectional gene flow (BDM), there exist four routes by which H_1 is reduced to H_0 (Fig. 5A&C, table 2(ii)). Segments (a) and (b) are the same as in the case of UDM, but there are two other routes (segments c and d). If any of the two migration rates $(\varpi_{AB}, \varpi_{BA})$ is ∞, H_1 will become H_0, with $\tau_Y \to \tau$ and with the other migration rate becoming indeterminate (as is τ_R). We leave it to the future to explore the posterior surface when large datasets of no gene flow are analyzed under the IIM model.

Materials and Methods

Simulation under the null model of no gene flow

To verify our theory for applying the S-D approach to calculating the Bayes factor and to examine the false-positive rate of the Bayesian test of gene flow, we conducted a small simulation under the null model of MSC with no gene flow (Fig. 1A or Fig. 4A). We used the population size $\theta = 0.005$ and three values for the species split time: $\tau_R = \frac{1}{2}\theta$, θ, and 10θ, representing very shallow, shallow, and deep divergences, respectively. Note that $\tau = \frac{1}{2}\theta$ represents a very shallow phylogeny as $\frac{1}{2}\theta$ is the average coalescent time between two sequences from the same species. Each dataset consisted of $L = 500$ loci, with 10 haploid sequences per locus per species, and with the sequence length to be 500. The number of replicates was 100. The `simulate` option of BPP was used to generate gene trees for the loci and then to 'evolve' sequences along branches of the gene trees under the JC model (Jukes and Cantor, 1969) to generate sequence alignments.

Each dataset was analyzed using BPP under the MSC-I (UDI and DBI) and MSC-M models of Figure 4B–D to conduct the Bayesian test. In analyses under the MSC-I models, the age of the root τ_R was assigned a gamma prior $G(2, 400)$, with the prior mean 0.005, too large for data simulated using $\tau_R = \frac{1}{2}\theta$ and too small for $\tau_R = 10\theta$. Given τ_R, τ_X under UDI and BDI has the uniform distribution. The option `thetamodel = linked-all` in BPP was used so that the same population size (θ) was assumed for all populations on the species tree, and assigned the gamma prior $\theta \sim G(2, 400)$ with mean 0.005. Note that models with different θs for species on the phylogeny are available in BPP, but in this paper assumptions on θs are unimportant. The introgression probability φ (or both φ_Y and φ_X in the BDI model) was assigned the prior $U(0, 1)$.

We used a burn-in of 40,000 iterations, after which we took 10^5 samples, sampling every 2 iterations. Convergence is assessed by running the same analysis twice and confirming consistency between runs. Running time for analyzing one dataset was $\sim$ 11 hours using one thread.

The MCMC sample under H_1: MSC-I was processed to calculate the Bayes factor via the S-D density ratio. For the test of H_0: MSC against H_1: MSC-I (UDI), we used 1% of the prior standard deviations as the differentials, so that $\epsilon_\varphi = \epsilon_\beta = 0.01/\sqrt{12} = 0.002886$. From eq. 31, the null region is

$$\Omega_\epsilon = (\varphi < \epsilon_\varphi) \cup (1-\beta < \epsilon_\beta) \cup (1-\varphi < \epsilon_\varphi),$$

where $\beta = \tau_Y/\tau_R$ (eq. 31). We calculate the Bayes factor as $B_{10} \approx \frac{\mathbb{P}(\Omega_\epsilon)}{\mathbb{P}(\Omega_\epsilon|\boldsymbol{x})} = \frac{3\epsilon}{\mathbb{P}(\Omega_\epsilon|\boldsymbol{x})}$.

We also tested the idea of using segments a and b only, for which the condition on priors holds (eq. 5), with

$$\Omega_\epsilon = (\varphi < \epsilon_\varphi) \cup (1-\beta < \epsilon_\beta),$$

and $B_{10} \approx \frac{2\epsilon}{\mathbb{P}(\Omega_\epsilon|\boldsymbol{x})}$. This produced nearly identical results as the use of all three segments.

For testing BDI, we used the null region defined in eq. 34. As in the test of UDI, we set $\epsilon_\varphi = \epsilon_\beta = 0.002886$. We have (see eq. 36)

$$B_{10} \approx \frac{\mathbb{P}(\Omega_\epsilon)}{\mathbb{P}(\Omega_\epsilon|\boldsymbol{x})} = \frac{\epsilon_\varphi^2 + \epsilon_\varphi^2 + \epsilon_\beta + \epsilon_\varphi^2 + \epsilon_\varphi^2}{\mathbb{P}(\Omega_\epsilon|\boldsymbol{x})}.$$

If we use only segments a, b, c, for which the condition on priors hold (table 1(i)), we have

$$\begin{aligned}\Omega_\epsilon = &(\varphi_X < \epsilon_\varphi, \varphi_Y < \epsilon_\varphi)\\ &\cup (1-\varphi_X < \epsilon_\varphi, 1-\varphi_Y < \epsilon_\varphi)\\ &\cup (1-\beta < \epsilon_\beta),\end{aligned}$$

and $B_{10} \approx \frac{\epsilon_\varphi^2+\epsilon_\varphi^2+\epsilon_\beta}{\mathbb{P}(\Omega_\epsilon|\boldsymbol{x})}$. This produced nearly identical results as the use of all five segments.

The same data, simulated under H_0, were also analyzed under the migration model (H_1: MSC-M, Fig. 4C) to conduct the Bayesian test. The MSC-M model assumes that migration from $A \to B$ occurs continuously in every generation (Flouri *et al.*, 2023). We used two priors to assess the impact of the prior, $\varpi_{XY} \sim G(2, 0.01)$ with mean 200 and $G(2, 1)$ with mean 2. Like the analysis under the MSC-I models (UDI and BDI), we assumed the same population size for all species (`thetamodel = linked-all`). The S-D approach was used to calculate the Bayes factor B_{10}, using the null region Ω_ϵ: $\varpi < \epsilon_\varpi = 0.005\sigma, 0.01\sigma, 0.05\sigma$, where σ is the prior standard deviation. Then $B_{10} \approx \mathbb{P}(\Omega_\epsilon)/\mathbb{P}(\Omega_\epsilon|\boldsymbol{x})$, with $\mathbb{P}(\varpi < \epsilon_\varpi)$ given by the CDF for the gamma prior.

Simulation under models of gene flow

To assess the power of the Bayesian test of gene flow between sister lineages and the precision of Bayesian estimation of the rate of gene flow, we simulated data under the MSC-I and MSC-M models of Figure 4B&C, with $A \to B$ (or $X \to Y$) gene flow. Most settings were the same as for the simulation under the null model. As before we used three species split times: $\tau_R = \frac{1}{2}\theta$, θ,

and 10θ, with $\theta = 0.005$. Under the MSC-I model, the introgression time was set at $\tau_Y = \frac{1}{4}\theta$, while the introgression probability varied: $\varphi = 0.05, 0.1, 0.2, 0.3, 0.5, 0.7$. For the MSC-M model, the population migration rate (from $A \rightarrow B$) was $M = 0.05, 0.1, 0.2, 0.3, 0.5, 1$. Each dataset consists of $L = 500$ loci, with 10 + 10 sequences, with the sequence length 500. The number of replicates is 100.

Each replicate dataset was analyzed under the three models of gene flow of Figure 4B–D: MSC-I (UDI), MSC-I (BDI), and MSC-M. One population size (θ) was assumed for all species. The priors and MCMC settings were as above for analysis of data simulated under the null model. The Bayesian test of gene flow was conducted as before, using the S-D density ratio to calculate the Bayes factor.

Simulation under the Gaussian mixture model (GM2)

A C program is written to generate MCMC samples under the Gaussian-mixture model $H_1 : \alpha N(\mu_1, 1) + (1-\alpha)N(\mu_2, 1)$ to estimate parameters $\boldsymbol{\theta} = (\alpha, \mu_1, \mu_2)$ (Fig. 3**A**). A dataset of size $n = 10^5$ was simulated under $H_0 : N(\mu, 1)$ with $\mu = 0.5$. The MCMC algorithm includes sliding-window steps to update α, μ_1, and μ_2 separately, with another step updating μ_1 and μ_2 together. Window sizes are proportional to $n^{-\frac{1}{2}}$. Besides acceptance rates of the moves, one can also use features of the posterior to assess convergence, as the posterior means $\mathbb{E}(\alpha|\boldsymbol{x}) = \frac{1}{2}$, and $\mathbb{E}(\mu_1|\boldsymbol{x}) = \mathbb{E}(\mu_2|\boldsymbol{x})$. Running time for 10^6 MCMC iterations was 1h20m using one thread on a server.

Bayesian inference of gene flow between two lizard species

We analyzed a dataset of 1000 ddRAD loci from three Spiny lizard species from Baja California, *Sceloporus orcutti* (O), *S. hunsakeri* (H), and *S. licki* (L), published and analyzed by Gottscho *et al.* (2025). There are 16, 16, and 12 haploid sequences per locus for O, H, and L, respectively, with the sequence length to be ~90bps. The species phylogeny is shown in Figure 6A, with possible gene flow between *S. orcutti* and *S. hunsakeri*.

We assigned the gamma prior $\tau_r \sim G(2, 200)$ with mean 0.01 on the age of the root, $\theta \sim G(2, 400)$ with mean 0.005 for the population sizes, and the uniform prior $U(0, 1)$ on the introgression probabilities $(\varphi_{oh}, \varphi_{ho})$. For the MCMC, we used a burn-in of 40,000 iterations, after which we took 2×10^5 samples, sampling every 2 iterations. The run took 11hrs using 36 threads on a server. The MCMC samples were processed to calculate the Bayes factor for testing the hypothesis of no gene flow using the S-D density ratio, as described in the paper.

The same data were also analyzed under the MSC-M model (IM) with bidirectional migration. The mutation-scaled migration rates $(\varpi_{oh}, \varpi_{ho})$ were assigned the gamma prior $G(2, 0.1)$ with mean 20. Other settings were the same as for the BDI model. Running time was 15hrs using 36 threads.

We also fitted the IIM model, specified as an instance of the MSC-M model by using a ghost lineage in the species tree for which no data are available,

```
((O, (H, ghost) h) s, L) r;
```

and by specifying migration between lineage O and h (the parent node of H and ghost) (Fig. 6**A**). The option `thetamodel = linked-mscm` was specified to force branches Hh and hs to share the same θ_H. Other settings are the same as for the IM model.

Results

Bayesian test of gene flow between sister species has low false positives

Results for the Bayesian test of gene flow using data simulated under the null model (MSC, Fig. 4A) are summarized in table 3. We discuss the results with $\epsilon_\varphi, \epsilon_\beta$ in MSC-I and ϵ_ϖ in MSC-M set to 0.01σ, where σ is the prior standard deviation. Use of 0.005σ and 0.05σ produced very similar results.

The Bayes factor B_{10} may be calibrated with reference to posterior model probabilities $\mathbb{P}(H_1|\boldsymbol{x})$ (table 4). At the 5% level (or $B_{10} > 19$), the false positive rate of the test is 0 for all but two settings, where the rate is 1% and 2%, lower than the nominal 5% (table 3). Here we are examining a frequentist property (the type-I error rate) of a Bayesian test, so one does not expect a match between the error rate and the significance level. Nevertheless, the false-positive rate of the Bayesian test is noted to be well below the nominal significance level.

Note that the Bayesian test makes a 'stronger' inference than the LRT in that it can strongly reject H_1 and support H_0 whereas the LRT may fail to reject H_0 but will never strongly support it. At the 5% level ($B_{10} < 0.0526$), H_1 is rejected in 72% and 100% of datasets simulated at deep divergence ($\tau = 10\theta$) when the alternative hypothesis is MSC-I UDI and MSC-M, respectively (table 3). Strong support for H_0 rarely occurs at shallow divergence ($\tau = \frac{1}{2}\theta$ or θ), suggesting that the data are far more informative at deep divergence. These results are consistent with the simulation of Thawornwattana *et al.* (2026).

Posterior means and the 95% highest posterior density (HPD) credible intervals (CIs) for parameters in the three alternative models of gene flow (UDI, BDI, and MSC-M; Fig. 4B–D) are summarized in Figure S4. We first consider the UDI model. The population size parameter θ is well estimated, but the other parameters $(\tau_R, \tau_Y, \varphi_Y)$ are not. In particular the introgression probability φ_Y has CIs that cover nearly the whole range $(0, 1)$. While wide CIs might indicate that the data are uninformative about gene flow, here a stronger conclusion may be more appropriate. Note that the summaries in Figure S4 do not accommodate the fact that all of the three segments a, b, and c of Figure 1D represent the scenario of no gene flow. For example, both $\varphi = 0$ and $\varphi = 1$, as well as any value for φ (when $\tau_Y = \tau_R$), represent the true model of no gene flow and are thus plausible, but the posterior mean and the HPD CI of φ generated from the MCMC sample are implausible. Similarly under the BDI model the neighborhoods of the

five segments/regions that represent the null model of no gene flow (table 1(ii)) represent the same biological scenario of no gene flow, and this fact is not accommodated in the summaries for φ_X and φ_Y in Figure S4.

As discussed above, the model of gene flow is rejected in many of these datasets, especially at deep divergence (table 3). In such cases, the correct conclusion is strong evidence against gene flow (if $B_{10} < 0.01$ or 0.05), rather than a lack of information in the data. We suggest that if the estimated introgression probability between sister lineages has large CIs, one should rely on the Bayesian test to assess the evidence for gene flow, as discussed in the paper.

The MSC-M model (the IM model) does not involve the irregularities of the MSC-I models discussed in this paper. Posterior summaries of parameters under the model (θ, τ_R, M) are all well-behaved and easy to interpret (Fig. S4). In particular, estimates of M are close to 0 (the true value), supporting the null model of no gene flow. The data are far more informative (with better estimates of M) on the deep tree than on the shallow trees.

Bayesian test of gene flow between sister species has high power

Results of the Bayesian test when data were simulated under the MSC-I and MSC-M models of Figure 4B&C, with $A \to B$ (or $X \to Y$) gene flow are summarized in Figure 7. First we consider data simulated under UDI (Fig. 7A–C). When the data are analyzed under the correct UDI model (I-UDI), the test has ~100% power except for the very shallow tree ($\tau_R = \frac{1}{2}\theta$) and low rates of gene flow ($\varphi \leq 0.1$). For data simulated under UDI, the BDI model is over-parametrized but not misspecified. The test under the BDI showed similar performance to the use of the UDI model. We implemented the MSC-M model under two priors, $\varpi \sim G(2, 0.01)$ with mean 200 and $\varpi \sim G(2, 1)$ with mean 2. The test using the MSC-M model, in particular, under the $\varpi \sim G(2, 1)$ prior (Fig. 7C′, I-M) had lower power than under UDI or BDI. In particular, power dropped when the introgression probability is very high. We note that when $\varphi > \frac{1}{2}$, most sequences from B are traced to the introgression parent A. It should be noted that in all datasets and all models, $B_{10} > 1$ (Fig. 7A–C), suggesting that the Bayesian test supports gene flow, even if the support may not reach the cutoff of $B_{10} > 19$ or 99.

Next we consider data simulated under the MSC-M model (Fig. 7D–F). The test has full power at the 1% level (with $B_{10} > 99$) under the two MSC-I models (M-UDI, M-BDI) for all three species split times and for all levels of migration rates. Misspecification of the mode of gene flow (i.e., the true model assumes continuous gene flow while the analysis model is discrete) has not had any major impact. The test under MSC-M had full power except for the combination of $M = 1$ and the $\varpi \sim G(2, 1)$ prior. We note that just as $\varphi \approx 1$ in MSC-I represents the scenario of isolation with no gene flow, $M = \infty$ represents the scenario of panmixia, with no gene flow.

Overall there is a greater amount of gene flow in data simulated under the MSC-M model. The expected amount of introgression under the MSC-M model, measured by the probability that any sequence from species B is traced to population A (irrespective of the migration time), when one traces the genealogical history of the sampled sequences backward in time, is

$$\varphi_0 = 1 - \mathrm{e}^{-\frac{4M_{AB}}{\theta_B}\tau_R} \tag{37}$$

(Huang *et al.*, 2022, eq. 10). At the low rate of $M = 0.05$, $\varphi_0 = 0.095, 0.181, 0.865$ for the three divergence levels (Fig. S5), higher than the low rate $\varphi = 0.05$ in the MSC-I (UDI) model. This should explain why all tests have full power at the low rate in data simulated under MSC-M. At the high rate of $M = 1$, we have $\varphi_0 = 0.865, 0.982, 1.000$, higher than $\varphi = 0.7$ in the UDI model. Data simulated under the MSC-M model with $\varphi_0 \approx 1$ may be expected to resemble data simulated under UDI with $\varphi = 1$ (which is equivalent to the MSC model of no gene flow; see Fig. 1C). This may explain the loss of power of the Bayesian test to detect gene flow at very high migration rates (M).

Compared with the true values used in the simulation, $\varpi = 40, 80, 160, 240, 400$ and 800 at the six values of M, the prior $\varpi \sim G(2, 1)$ has a mean that is far too small, creating conflicts between the prior and the data and leading to reduced power (Fig. 7D–F′). In practical data analysis, $G(2, 0.1)$ with the prior mean 20 may be a good choice.

Bayesian estimation of the rate of introgression between sister species is reliable only at deep species divergences

We examine the precision and accuracy of parameter estimation in data analyzed in Figure 7, simulated under models with gene flow. In particular, we are interested in the rate of gene flow (φ under MSC-I and M under MSC-M; Fig. 4B–D). The results are summarized in Figure 8.

First we consider data simulated under the UDI model (Fig. 8A–C). Under the correct UDI model, species split time τ_R and the population size (θ) are well estimated. Note that we assumed the same population size for all species on the tree, so there is a huge amount of information in the data about θ. The introgression time τ_Y and the introgression probability φ involved large uncertainties when the true φ is small (< 0.2, say). When φ is very small, the null model of no gene flow is nearly correct, and the three segments of Figure 1C (where $\varphi = 0$, 1, or any value between 0–1) are expected to fit the data nearly equally well; in other words, there exist ridges in the posterior surface close to the three segments of Figure 1C, and the data are not informative about φ. The poor estimation of τ_Y and φ form a sharp contrast with the precise and accurate estimation of τ_R and θ. Furthermore, the species split time has a large impact, with the CIs for φ to be much narrower for the deep tree (with $\tau_R = 10\theta$) than for the very shallow tree (with $\tau_R = \frac{1}{2}\theta$) (Fig. 8).

The analysis under BDI (Fig. 8B, I-BDI) produces similar estimates of the species split time and introgression time (τ_R, τ_Y) to those under UDI, but the estimates of introgression probabilities (φ_X, φ_Y) show interesting patterns, different from the analysis under UDI (the I-UDI setting). For data generated at low φ in the true UDI model, both φ_X and φ_Y are poorly estimated. This is similar to the poor estimation of φ_Y under UDI, due to the issue with the summary: at the low rate, the null model of no gene flow is nearly correct, and the four lines and a plane under the BDI model (eq. 34, table 1(ii)), where φ_X and φ_Y are either 0 or 1 or anywhere over (0, 1), are expected to fit the data nearly equally well. At high φ in the true UDI model, the BDI model underestimates φ_Y (which has the true value φ) and overestimates φ_X (which has the true value 0), having difficulty in determining the direction of introgression. This may reflect the label-switching unidentifiability problem for the BDI model, as discussed by Yang and Flouri (2022). Under the BDI model, without constraints on θ, two sets of parameter values, $\boldsymbol{\theta} = (\varphi_X, \varphi_Y, \theta_X, \theta_Y)$ and $\boldsymbol{\theta}' = (1-\varphi_X, 1-\varphi_Y, \theta_Y, \theta_X)$, are unidentifiable (Yang and Flouri, 2022, Fig. 1). Similar unidentifiability involving φ_X and φ_Y occurs when all populations are assumed to have the same size (θ).

Next we consider data simulated under the MSC-M model (Fig. 8D–F). Estimates under UDI and BDI at shallow divergence ($\tau_R = \frac{1}{2}$) and low migration rate (M) are very poor. As discussed above, when the null model of no gene flow is nearly correct, our summaries in Figure 8 do not account for the fact that different parts of the parameter space represent the same biological scenario of no gene flow. When the migration rate is higher, the estimates improve, in particular at high divergence. However, when M is very high, the model will again be close to the null model of no gene flow. In general the MSC-I model does not recover all the gene flow that has occurred. The results are similar to previous simulations in which data were generated under the MSC-M model and analyzed under the MSC-I model (Huang *et al.*, 2022, φ against M under the IM model).

In the M-M setting, the MSC-M model with $A \to B$ migration is used to both simulate and analyze the data (Fig. 8D–F). This case shows a complex pattern with respect to the species split time (τ_R) and the migration rate M. When M is low (0.05 or 0.1), both θ and τ are well estimated, as is the migration rate M. When $0.2 \le M \le 0.5$, the estimates are good for the shallow trees only ($\tau_R = \frac{1}{2}\times\theta$ or θ), but poor for the deep tree. The prior on ϖ affected the posterior estimate (Fig. 8F&F′). Note that at $M \ge 0.2$, $\varphi_0 = 1.000$ on the deep tree (Fig. S5), and at such extreme amount of gene flow, the estimate of τ_R may collapse.

Asymptotic analysis to assess the impact of species split time

To understand the effect of species split time on the information content in the data concerning the introgression probability φ under the UDI model (Fig. 4B), we consider the simple case of only two sequences per locus, with one sequence per species. The large-sample behavior of the estimation method (when the number of loci $L \to \infty$) is analytically tractable (Huang *et al.*, 2022). We assume one θ for all populations, so there are 4 parameters in the model, $\boldsymbol{\theta} = (\varphi, \tau_Y, \tau_R, \theta)$. Let $\boldsymbol{\theta}^* = (\varphi^*, \tau_Y^*, \tau_R^*, \theta^*)$ be the true parameter values. The alignment at each locus is summarized as the observed number (x) of differences at n sites. Let the probability of observing x differences at n sites be $p(x; \boldsymbol{\theta})$, given by Huang *et al.* (2022, eq. S9). When the number of loci $L \to \infty$, the per-locus log likelihood becomes

$$\ell(\boldsymbol{\theta}) = \sum_{x=0}^{n} p(x; \boldsymbol{\theta}_*) \log p(x; \boldsymbol{\theta}). \tag{38}$$

Note that here n is the sequence length so that the number of variable sites (x) has the range $(0, n)$.

We fix the population size parameter θ at the true value, as it is typically well estimated. For given τ_Y and φ, we estimate τ_R by maximizing ℓ, and calculate the optimized ℓ as a function of the other two parameters (Fig. 9). We used three values for τ_R: $\frac{1}{2}\theta$, θ, and 10θ, and three values for φ: 0.01, 0.1, and 0.6. The log-likelihood contour for τ_Y and φ is shown in Figure 9.

When $\varphi = 0.01$ and the true $\tau_R = \frac{1}{2}\theta$, the log-likelihood surface has an elongated foot shape, similar to the Bayesian posterior surface for the two parameters of Figure 1F. Information content on φ improved when the species split time is larger ($\tau_R = 10\theta$) or when the rate of gene flow is higher ($\varphi = 0.1$ or 0.6). In particular, as φ increased, the contour lines became more circular, resulting in narrower CIs for both parameters (Fig. 9).

In the case of two sequences per locus and treating the coalescent time as data, Thawornwattana *et al.* (2023, eq. 2) give the Fisher information for estimating φ_Y as

$$I \approx \frac{P_X}{\varphi_Y(1-\varphi_Y P_X)}, \tag{39}$$

where $P_X = 1 - \mathrm{e}^{-\frac{2}{\theta_X}(\tau_R - \tau_Y)}$ is the probability that two sequences (a, b) entering population X coalesce in population X rather than in population R (Fig. 4B). The data are more informative about φ_Y if P_X is larger (i.e, if species divergence is deeper with $\tau_R \gg \tau_Y$ and/or if population size θ_X is smaller).

Inference of gene flow between Sceloporus *lizard species*

We analyzed a ddRAD dataset from three *Sceloporus* lizard species, *Sceloporus orcutti* (O), *S. hunsakeri* (H), and *S. licki* (L) to test for gene flow between the sister species *S. orcutti* and *S. hunsakeri* (Fig. 6A). We ran BPP under the bidirectional introgression (BDI) and migration (BDM) models. First we note that estimates of parameters shared between the two models were very similar (table 5). The introgression time under BDI model was nearly zero, indicating that gene flow between the two species may be ongoing or stopped very recently. The rate of gene flow was small, with $\varphi_{oh} = 0.0044$ (with the 95% HPD CI 0.0013–0.0079) and $\varphi_{ho} = 0.0022$ (0.0008–0.0038) (Fig. 6B, table

5). The BDM model assumes ongoing gene flow, with $M_{oh} = M_{ho} = 0.0033$ migrant individuals in each direction per generation (table 5). Those rates translate to $\varphi_{0,o\to h} = 1 - e^{-4\times 0.003433/0.003388\times 0.007606} = 0.0304$, and $\varphi_{0,h\to o} = 0.0125$ according to eq. 37, suggesting that the migration model recovered more gene flow than the MSC-I model. We also fitted the IIM model, in which gene flow occurs post species split but stops at a later time. This produced nearly identical estimates to the IM model assuming ongoing gene flow (table 5). Overall the posterior estimates suggest a small but significant amount of gene flow in each direction, and gene flow is either ongoing or stopped a short time ago.

We then applied the Bayesian test. Under the BDI model, one can test $H_0 : \varphi_{oh} = 0, \varphi_{ho} = 0$ against $H_1 : \varphi_{oh} \neq 0, \varphi_{ho} \neq 0$ (table 1(ii)). Rejection of H_0 means introgression in at least one direction. We used each of the five segments for which the condition on priors holds to calculate the Bayes factor, and all calculations led to strong support for gene flow (with $B \approx \infty$).

Figure 6d illustrates the test of $H_0 : \varphi_{oh} = 0$ against $H_1 : \varphi_{oh} \neq 0$, with φ_{ho} treated a nuisance parameter (present in both H_0 and H_1). There exist three segments, each of which can be used to calculate B_{10}, as described in table 1(i). Here we used the first segment, with $B_{10} \approx \frac{\mathbb{P}(\varphi_{ho}<\epsilon)}{\mathbb{P}(\varphi_{ho}<\epsilon|\boldsymbol{x})}$. The data supported the $h \to o$ introgression, but the introgression probability was very low. However, the low introgression probabilities caused challenges to the S-D approach as B was sensitive to the choice of ϵ.

The test under the BDM model (the IM model) does not involve irregularities. In Figure 6E, we used the same MCMC sample under BDM for Figure 6C to calculate the Bayes factor for testing three null hypotheses. The first is $H_0 : \varpi_{oh} = 0$, against $H_1 : \varpi_{oh} > 0$, with ϖ_{ho} treated as a nuisance parameter. The Bayes factor is given as $B_{oh} \approx \frac{\mathbb{P}(\varpi_{oh}<\epsilon)}{\mathbb{P}(\varpi_{oh}<\epsilon|\boldsymbol{x})}$, where $\mathbb{P}(\varpi_{oh} < \epsilon|\boldsymbol{x})$ was estimated by the proportion of MCMC samples in which $\varpi_{oh} < \epsilon$ while the prior probability $\alpha = \mathbb{P}(\varpi_{oh} < \epsilon) = G(\epsilon; 2, 0.1)$, the CDF of the gamma distribution. In fact we fix $\alpha = 0.001, 0.005$ and obtain $\epsilon = 0.454, 1.035$ via the quantile function. In the second test we have $H_0 : \varpi_{ho} = 0$ against $H_1 : \varpi_{ho} > 0$. The third test is for $H_0 : \varpi_{oh} = 0, \varpi_{ho} = 0$ against $H_1 : \varpi_{oh} > 0, \varpi_{ho} > 0$. We have $B_2 \approx \frac{\mathbb{P}(\varnothing)}{\mathbb{P}(\varnothing|\boldsymbol{x})}$, calculated by defining $\varnothing = (\varpi_{oh} < \epsilon, \varpi_{ho} < \epsilon)$. The three Bayes factors (B_{oh}, B_{ho}, B_2) are plotted against the tail probability α in Figure 6E. All of them are $\gg 1$, strongly supporting gene flow although there is a lack of precision when B is very large.

Discussion

Bayesian testing of gene flow between sister lineages and calculation of the Bayes factor

Considered a problem of statistical hypothesis testing (Jeffreys, 1935), the test of introgression between sister lineages involves a whole suite of challenging nonstandard features, including boundary problems (the null value $\varphi = 0$ is at the boundary of the parameter space for H_1), indeterminate parameters (e.g., when the introgression probability $\varphi = 0$, the time of introgression becomes unidentifiable), and in particular multiple routes from the alternative hypothesis to the null hypothesis (e.g., Fig. 1C). While these do not invalidate the definition and use of the Bayes factor for hypothesis testing, they pose challenges to the use of the S-D density ratio to calculate the Bayes factor. In this study, we show that each route from H_1 to H_0 can be used to calculate the Bayes factor via the S-D density ratio. When the condition on priors is satisfied over all routes, a combined estimate using all routes is also possible. Note that one draws the correct conclusion if the calculated Bayes factor falls into the right interval, and high precision may not be very important (table 4).

Our simulation demonstrates that our theories can be implemented in practice, and the resulting Bayesian test has low false positives and high power for testing gene flow between sister lineages.

For parameter estimation, we see that the posterior surface under introgression (MSC-I) models may involve nearly flat ridges when there is no gene flow or the rate of gene flow is very low. There are also label-switching unidentifiability issues under the bidirectional model (BDI) (Yang and Flouri, 2022). When the rate of introgression is low, one should exercise care when interpreting simple summaries of the MCMC samples (such as posterior means and CIs), and rely instead on the Bayes factor to assess the evidence for gene flow (Fig. S4).

Here we note a few advantages of the Bayes factor, relative to frequentist hypothesis testing such as the LRT (Yang, 2026). The Bayes factor penalizes parameter-rich models automatically. It works whether the compared models are nested or nonnested, and when the compared models have an irregular parameter space. The Bayes factor is also able to reject the alternative hypothesis with great force.

We note that many methods exist for calculating the marginal likelihood (see, for review, Fourment *et al.*, 2020). Numerically stable algorithms, such as path-sampling or thermodynamic integration (Ogata, 1989; Gelman and Meng, 1998; Lartillot and Philippe, 2006), stepping-stones (Xie *et al.*, 2011), and nested sampling (Skilling, 2006), require at least 50 runs of the MCMC algorithm for each of the two compared models (Fourment *et al.*, 2020, table 1). The S-D approach to computing the Bayes factor requires only one run of the MCMC algorithm under the alternative model of gene flow (H_1) (Dickey, 1971; Ji *et al.*, 2023), and involves $\sim$ 100 times less computation than those sample-based methods, offering a major computational advantage. It is thus useful to extend the S-D approach to nonstandard conditions.

The irregularities discussed in this paper concern the test of gene flow between sister lineages under the discrete introgression models (UDI and BDI) and under the continuous IIM model (tables 1 & 2). Bayesian test of gene flow between non-sister lineages is not affected by those issues. Note also that the irregularities affect estimation of parameters for gene flow between sister species only when gene flow is absent or its rate is very

low. At high rates or when there is clear evidence of gene flow and our objective is to estimate parameters such as the introgression probability, the irregularities are not relevant. When the MCMC samples the rest of the parameter space in H_1, away from the three lines in Figure 1D, different points in the space represent distinct biological scenarios and are identifiable.

The theory developed in this paper applies as well if introgression is between sister lineages that are ancestral branches on the species phylogeny (Fig. 10). Whether one or both of the sister lineages involved are ancestral, there are three routes for reducing the UDI introgression model (H_1) to the null model of no gene flow (H_0). For example, in comparison of the two models of Figure 10A&B, the null point also consists of three lines or line segments, given by eq. 31 (Fig. 1C). The same holds true for the test comparing the two models of Figure 10C&D, in which case both sister lineages are ancestral. Similarly in the test under the BDI introgression model, there are five routes for reducing H_1 to H_0, as in the case where both involved species are extant (eq. 34).

Alternative tests of gene flow between sister lineages

The irregularities of the testing problem discussed in this paper also impact frequentist hypothesis testing such as the LRT, and indeed often cause insurmountable difficulties for which no known solutions exist (Chernoff, 1954; Self and Liang, 1987; Mitchell *et al.*, 2019; Brazzale and Mameli, 2024). In the case where H_0 corresponds to H_1 with one parameter of interest at the boundary of the space, the null distribution for the LRT statistic is a 50:50 mixture of 0 and χ^2_1 (Chernoff, 1954). The problem of indeterminate parameters is harder, and in general the null distribution is unknown. The problem of multiple routes for reducing H_1 to H_0 does not appear to have been studied in the statistics literature.

In population genetics, the programs ∂a∂i (Gutenkunst *et al.*, 2009) and FASTSIMCOAL2 (Excoffier *et al.*, 2013) calculate a composite likelihood for data of joint site frequency spectrum (SFS). The likelihood is composite as the method ignores the lack of independence among linked SNP sites, or the lack of independence among population pairs when multiple populations are analyzed. The use of composite likelihood ignoring correlation and the use of data summaries that are not sufficient statistics make it hard to define the null distribution of the test statistic; even without the irregularities discussed here, the usual χ^2 distribution may not apply. In practice, nonparametric resampling methods such as the bootstrap are used instead (Excoffier *et al.*, 2013).

Supplementary Material and Data Availability

Supplemental information is available at https://doi.org/10.5281/zenodo.xxxxxxxx. Supplementary data (sequence alignments and control files for running BPP are available through the Dryad Digital Repository at 10.5061/dryad.zkh1893pp.

Acknowledgments

We thank Dr Adam Leaché for providing the spiny lizard data, and Drs Mengyu Li and Jun Liu for bringing to our attention Rousseau and Mengersen (2011). We are indebted to three anonymous reviewers for many constructive comments and suggestions.

Funding

This study has been supported by Biotechnology and Biological Sciences Research Council (BBSRC) grants (BB/T003502/1, BB/X007553/1) and Natural Environment Research Council (NERC) grant (NE/X002071/1) to Z.Y., by Natural Science Foundation of China (NSFC) grant (12101295), Guangdong Natural Science Foundation grant (2022A1515011767), and Shenzhen Training Project of Excellent Scientific & Technological Talents grant (RCYX20221008093033012) to X.J., and by China Natural Science Foundation grants (T2122017 and 32070685) and China National Key R&D Program (2020YFA0712700) to T.Z.

Conflict of Interest

The authors declare no conflict of interests.

References

Blischak, P. D., Chifman, J., Wolfe, A. D., and Kubatko, L. S. 2018. HyDe: a Python package for genome-scale hybridization detection. *Syst. Biol.*, 67(5): 821–829.

Brazzale, A. R. and Mameli, V. 2024. Likelihood asymptotics in nonregular settings: A review with emphasis on the likelihood ratio. *Statist. Sci.*, 39: 322–345.

Chernoff, H. 1954. On the distribution of the likelihood ratio. *Ann. Math. Stat.*, 25: 573–578.

Costa, R. J. and Wilkinson-Herbots, H. 2017. Inference of gene flow in the process of speciation: An efficient maximum-likelihood method for the isolation-with-initial-migration model. *Genetics*, 205(4): 1597–1618.

Dalquen, D., Zhu, T., and Yang, Z. 2017. Maximum likelihood implementation of an isolation-with-migration model for three species. *Syst. Biol.*, 66: 379–398.

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

Excoffier, L., Dupanloup, I., Huerta-Sanchez, E., Sousa, V. C., and Foll, M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genet.*, 9(10): e1003905.

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.

Flouri, T., Huang, J., Jiao, X., Kapli, P., Rannala, B., and Yang, Z. 2022. Bayesian phylogenetic inference using relaxed-clocks and the multispecies coalescent. *Mol. Biol. Evol.*, 39(8): msac161.

Flouri, T., Jiao, X., Huang, J., Rannala, B., and Yang, Z. 2023. Efficient Bayesian inference under the multispecies coalescent with migration. *Proc. Nat. Acad. Sci. U.S.A.*, 120(44): e2310708120.

Fourment, M., Magee, A. F., Whidden, C., Bilge, A., Matsen, F. A., and Minin, V. N. 2020. 19 dubious ways to compute the marginal likelihood of a phylogenetic tree topology. *Syst. Biol.*, 69(2): 209–220.

Gelman, A. and Meng, X. 1998. Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Stat. Sci.*, 13: 163–185.

Gottscho, A. D., Hollingsworth, B. D., Espinal, J. L., Leaché, A. D., Reeder, T. W., and de Queiroz, K. 2025. Comparative phylogeography of lizards (squamata: Phrynosomatidae) in Baja California and expansion of *Callisaurus draconoides* within the North American deserts. *J. Biogeography*, page 10.1101/2025.04.29.650089.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H., Hansen, N. F.and Durand, E. Y., Malaspinas, A. S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prufer, K., Meyer, M., Burbano, H. A., Good, J. M., Schultz, R., Aximu-Petri, A., Butthof, A., Hober, B., Hoffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E. S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Doronichev, V. B., Golovanova, L. V., Lalueza-Fox, C., de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R. W., Johnson, P. L., Eichler, E. E., Falush, D., Birney, E., Mullikin, J. C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D., and Paabo, S. 2010. A draft sequence of the Neandertal genome. *Science*, 328: 710–722.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., and Bustamante, C. D. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet*, 5(10): e1000695.

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. 2022. Phylogenomic approaches to detecting and characterizing introgression. *Genetics*, 220(2): 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 gene flow between species under misspecified models. *Mol. Biol. Evol.*, 39(12).

Jeffreys, H. 1935. Some tests of significance, treated by the theory of probability. *Proc. Cam. Phil. Soc.*, 31: 203–222.

Ji, J., Jackson, D. J., Leaché, A. D., and Yang, Z. 2023. 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. *Syst. Biol.*, 72(2): 446–465.

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.*

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

Kong, S., Swofford, D. L., and Kubatko, L. S. 2025. Inference of phylogenetic networks from sequence data using composite likelihood. *Syst. Biol.*, 74(1): 53–69.

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

Lewis, P., Holder, M., and Holsinger, K. 2005. Polytomies and Bayesian phylogenetic inference. *Syst. Biol.*, 54: 241–253.

Meng, C. and Kubatko, L. 2009. Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: a model. *Theo. Popul. Biol.*, 75(1): 35–45.

Mitchell, J. D., Allman, E. S., and Rhodes, J. A. 2019. Hypothesis testing near singularities and boundaries. *Electron. J. Stat.*, 13(1): 2150–2193.

Ogata, Y. 1989. A Monte Carlo method for high dimensional integration. *Numer. Math.*, 55: 137–157.

O'Hagan, A. and Forster, J. 2004. *Kendall's Advanced Theory of Statistics: Bayesian Inference*. Arnold, London.

Payseur, B. A. and Rieseberg, L. H. 2016. A genomic perspective on hybridization and speciation. *Mol. Ecol.*, 25(11): 2337–2360.

Rousseau, J. and Mengersen, K. 2011. Asymptotic behaviour of the posterior distribution in overfitted mixture models. *J. R. Stat. Soc. B*, 73: 689–710.

Roux, C., Fraisse, C., Romiguier, J., Anciaux, Y., Galtier, N., and Bierne, N. 2016. Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biol.*, 14(12): e2000234.

Self, S. and Liang, K.-Y. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Stat. Assoc.*, 82: 605–610.

Skilling, J. 2006. Nested sampling for general Bayesian computation. *Bayesian Anal*, 1: 833–859.

Solis-Lemus, C. and Ane, C. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genet.*, 12(3): e1005896.

Solis-Lemus, C., Bastide, P., and Ane, C. 2017. PhyloNetworks: a package for phylogenetic networks. *Mol. Biol. Evol.*, 34(12): 3292–3298.

Sousa, V. and Hey, J. 2013. Understanding the origin of species with genome-scale data: modelling gene flow. *Nat. Reviews Genet.*, 14(6): 404–414.

Suchard, M. A., Weiss, R. E., and Sinsheimer, J. S. 2003. Testing a molecular clock without an outgroup: derivations of induced priors on branch-length restrictions in a bayesian framework. *Syst. Biol.*, 52: 48–54.

Thawornwattana, Y., Huang, J., Flouris, T., Mallet, J., and Yang, Z. 2023. Inferring the direction of introgression using genomic sequence data. *Mol. Biol. Evol.*, 40(8): msad178.

Thawornwattana, Y., Flouris, T., Mallet, J., and Yang, Z. 2025. Inference of gene flow between species from genomic data when the mode, direction and lineages are misspecified. *Mol. Biol. Evol.*, 42: 1–18.

Thawornwattana, Y., Rannala, B., and Yang, Z. 2026. On the robustness of Bayesian inference of gene flow to intragenic recombination and natural selection. *Mol. Biol. Evol.*, page https://doi.org/10.1093/molbev/msaf327.

Verdinelli, I. and Wasserman, L. 1995. Computing bayes factors using a generalization of the savage-dickey density ratio. *J. Am. Stat. Assoc.*, 90(430): 614–618.

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

Xie, W., Lewis, P. O., Fan, Y., Kuo, L., and Chen, M.-H. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst. Biol.*, 60: 150–160.

Yang, Z. 2007. Fair-balance paradox, star-tree paradox and Bayesian phylogenetics. *Mol. Biol. Evol.*, 24: 1639–1655.

Yang, Z. 2026. Bayesian phylogenetic methods. In L. Knowels, editor, *Encyclopaedia of Evolutionary Biology*, pages 62–76. Elsevier, New York, 2nd edition.

Yang, Z. and Flouri, T. 2022. Estimation of cross-species introgression rates using genomic data despite model unidentifiability. *Mol. Biol. Evol.*, 39(5): 10.1093/molbev/msac083.

Yang, Z. and Rannala, B. 2005. Branch-length prior influences Bayesian posterior probability of phylogeny. *Syst. Biol.*, 54: 455–470.

Yang, Z. and Rannala, B. 2010. Bayesian species delimitation using multilocus sequence data. *Proc. Natl. Acad. Sci. U.S.A.*, 107: 9264–9269.

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.

Zhu, T. and Yang, Z. 2012. Maximum likelihood implementation of an isolation-with-migration model with three species for testing speciation with gene flow. *Mol. Biol. Evol.*, 29: 3131–3142.

Table 1: Parameters of interest (ω), nuisance parameters (λ), and indeterminate parameters (ξ) and condition on priors in Bayesian test of introgression

Segment	Parameters of interest ($\omega = \omega_0$)	Indeterminate (ξ)	Nuisance (λ)	Condition on λ (eq. 5)
(**i**) Unidirectional introgression (UDI, Fig. 1C)				
(a)	$\varphi = 0$	(τ_Y)	(τ_R, θ)	yes
(b)	$\beta = \frac{\tau_Y}{\tau_R} = 1$	(φ)	(τ_R, θ)	yes
(c)	$\varphi = 1$	(τ_R)	(τ_Y, θ)	no
(**ii**) Bidirectional introgression (BDI, Fig. 1D)				
(a)	$\varphi_X = 0, \varphi_Y = 0$	(τ_Y)	(τ_R, θ)	yes
(b)	$\varphi_X = 1, \varphi_Y = 1$	(τ_Y)	(τ_R, θ)	yes
(c)	$\beta = \frac{\tau_Y}{\tau_R} = 1$	(φ_X, φ_Y)	(τ_R, θ)	yes
(d)	$\varphi_X = 0, \varphi_Y = 1$	(τ_R)	(τ_Y, θ)	no
(e)	$\varphi_X = 1, \varphi_Y = 0$	(τ_R)	(τ_Y, θ)	no

Note.— For UDI, in segment c nuisance parameter τ_Y in H_1 is mapped onto τ in H_0 (Fig. 1A&B). For BDI, segments a, b, d, and e are lines, while segment c is a plane, and in segments d and e, nuisance parameter τ_Y in H_1 is mapped to τ_R in H_0. In last column, 'yes' means that the condition on priors (eq. 5) holds for the segment and 'no' means it does not.

Table 2: Parameters of interest (ω), nuisance parameters (λ), and indeterminate parameters (ξ) and condition on priors in the Bayesian test of gene flow under the isolation-with-initial-migration (IIM) models

Segment	Parameters of interest ($\omega = \omega_0$)	Indeterminate (ξ)	Nuisance (λ)	Condition on λ
(**i**) Unidirectional migration (UDM, Fig. 5B)				
(a)	$\varpi_{AB} = 0$	(τ_Y)	(τ_R, θ)	yes
(b)	$\beta = \frac{\tau_Y}{\tau_R} = 1$	(ϖ_{AB})	(τ_R, θ)	yes
(c)	$\varpi_{AB} = \infty$	(τ_R)	(τ_Y, θ)	no
(**ii**) Bidirectional migration (BDM, Fig. 5C)				
(a)	$\varpi_{XY} = 0, \varpi_{YX} = 0$	(τ_Y)	(τ_R, θ)	yes
(b)	$\beta = \frac{\tau_Y}{\tau_R} = 1$	$(\varpi_{AB}, \varpi_{BA})$	(τ_R, θ)	yes
(c)	$\varpi_{AB} = \infty$	(τ_R, ϖ_{BA})	(τ_Y, θ)	no
(d)	$\varpi_{BA} = \infty$	(τ_R, ϖ_{AB})	(τ_Y, θ)	no

Note.— See legend to table 1.

Table 3: Bayes factor (B_{10}) for testing introgression in 100 replicate datasets simulated under the null model of no gene flow (H_0, MSC, Fig. 4A), and analyzed under models of gene flow (H_1, MSC-I and MSC-M, Fig. 4B–D)

	(A) MSC-I, UDI			(B) MSC-I, BDI			(C) MSC-M		
B_{10}	$\frac{1}{2}\theta$	θ	10θ	$\frac{1}{2}\theta$	θ	10θ	$\frac{1}{2}\theta$	θ	10θ
(i) With $\epsilon_\varphi, \epsilon_\varpi$ set at 0.005σ									
(0, 1/99)	0	0	49	0	0	2	0	0	96
(1/99, 1/19)	1	4	31	3	0	34	0	4	4
(1/19, 1)	89	89	20	90	94	62	69	92	0
(1, 19)	10	7	0	7	5	2	20	4	0
(19, 99)	0	0	0	0	1	0	0	0	0
(99, ∞)	0	0	0	0	2	0	7	0	0
(ii) With $\epsilon_\varphi, \epsilon_\varpi$ set at 0.01σ									
(0, 1/99)	0	0	45	0	0	0	0	0	95
(1/99, 1/19)	0	3	27	0	0	34	0	2	5
(1/19, 1)	90	92	28	93	95	66	80	95	0
(1, 19)	10	5	0	7	4	0	18	3	0
(19, 99)	0	0	0	0	1	0	0	0	0
(99, ∞)	0	0	0	0	0	0	2	0	0
(iii) With $\epsilon_\varphi, \epsilon_\varpi$ set at 0.05σ									
(0, 1/99)	0	0	31	0	0	1	0	0	94
(1/99, 1/19)	1	0	34	3	1	45	0	0	6
(1/19, 1)	92	95	35	90	95	53	79	97	0
(1, 19)	7	5	0	7	4	1	20	3	0
(19, 99)	0	0	0	0	0	0	0	0	0
(99, ∞)	0	0	0	0	0	0	1	0	0

Note.– B_{10} is calculated using the S-D density ratio with $\epsilon_\varphi = \epsilon_\beta$ (in MSC-I) and ϵ_ϖ (in MSC-M) set at $0.005\times, 0.01\times, 0.05\times\sigma$, where σ is the prior standard deviation; i.e., $\sigma = \frac{1}{\sqrt{12}} = 0.28867$ for $\varphi, \beta \sim U(0,1)$ and $\sigma = 1.414$ for $\varpi \sim G(2,1)$. The Bayes factor is grouped into six bins according to the cutoffs of table 4.

Table 4: Calibration of the Bayes factor via posterior model probabilities

$\log B_{10}$	B_{10}	$\mathbb{P}(H_1\|\boldsymbol{x})$	Interpretation
$(-\infty, -4.60)$	$(0, 1/99)$	$(0, 1\%)$	very strong support for H_0
$(-4.60, -2.94)$	$(1/99, 1/19)$	$(1\%, 5\%)$	strong support for H_0
$(-2.94, 0)$	$(1/19, 1)$	$(5\%, 50\%)$	support for H_0
$(0, 2.94)$	$(1, 19)$	$(50\%, 95\%)$	support for H_1
$(2.94, 4.60)$	$(19, 99)$	$(95\%, 99\%)$	strong support for H_1
$(4.60, \infty)$	$(99, \infty)$	$(99\%, 100\%)$	very strong support for H_1

Note.— With equal prior probabilities for models ($\pi_0 = \pi_1 = \frac{1}{2}$), the posterior probabilities for models are $\mathbb{P}(H_1|\boldsymbol{x}) = \frac{1}{1+1/B_{10}}$ and $\mathbb{P}(H_0|\boldsymbol{x}) = 1 - \mathbb{P}(H_1|\boldsymbol{x})$. Note that $\log 19 = 2.94$ and $\log 99 = 4.60$ (cf: the log likelihood difference of $\frac{1}{2}\chi^2_{1,\alpha} = 1.92$ and 3.32 at the $\alpha = 1\%$ and 5% levels).

Table 5: Posterior means and 95% HPD CIs for parameters under the BDI and BDM models obtained from BPP analyses of the lizard data

Parameters	MSC-I (BDI)	MSC-M (BDM, IM)	MSC-M (BDM, IIM)
τ_r	0.0079 (0.0075, 0.0084)	0.0080 (0.0076, 0.0084)	0.0080 (0.0076, 0.0084)
τ_s	0.0074 (0.0070, 0.0079)	0.0076 (0.0071, 0.0081)	0.0077 (0.0072, 0.0081)
$\tau_o = \tau_h$	0.0003 (0.0001, 0.0005)		0.0001 (0.0000, 0.0004)
θ_O	0.0079 (0.0075, 0.0083)	0.0079 (0.0075, 0.0083)	0.0079 (0.0075, 0.0083)
θ_H	0.0034 (0.0032, 0.0037)	0.0034 (0.0031, 0.0036)	0.0034 (0.0031, 0.0036)
θ_L	0.0055 (0.0052, 0.0059)	0.0056 (0.0052, 0.0059)	0.0055 (0.0052, 0.0059)
θ_r	0.0073 (0.0064, 0.0082)	0.0071 (0.0062, 0.0080)	0.0071 (0.0062, 0.0080)
θ_s	0.0046 (0.0007, 0.0088)	0.0046 (0.0004, 0.0097)	0.0032 (0.0004, 0.0065)
φ_{oh}	0.0044 (0.0013, 0.0079)		
φ_{ho}	0.0022 (0.0008, 0.0038)		
ϖ_{oh}		4.06 (1.91, 6.23)	4.75 (2.06, 7.60)
ϖ_{ho}		1.66 (0.63, 2.72)	1.80 (0.57, 3.09)
M_{oh}		0.0034 (0.0017, 0.0053)	0.0040 (0.0018, 0.0064)
M_{oh}		0.0033 (0.0013, 0.0054)	0.0035 (0.0011, 0.0061)

Note.— In the isolation-with-migration (IM) model, gene flow has been ongoing post speciation until the present time, while in the isolation-with-initial-migration (IIM) model, gene flow occurs during the time period from τ_s to τ_h and stops at time τ_h (Fig. 6A).

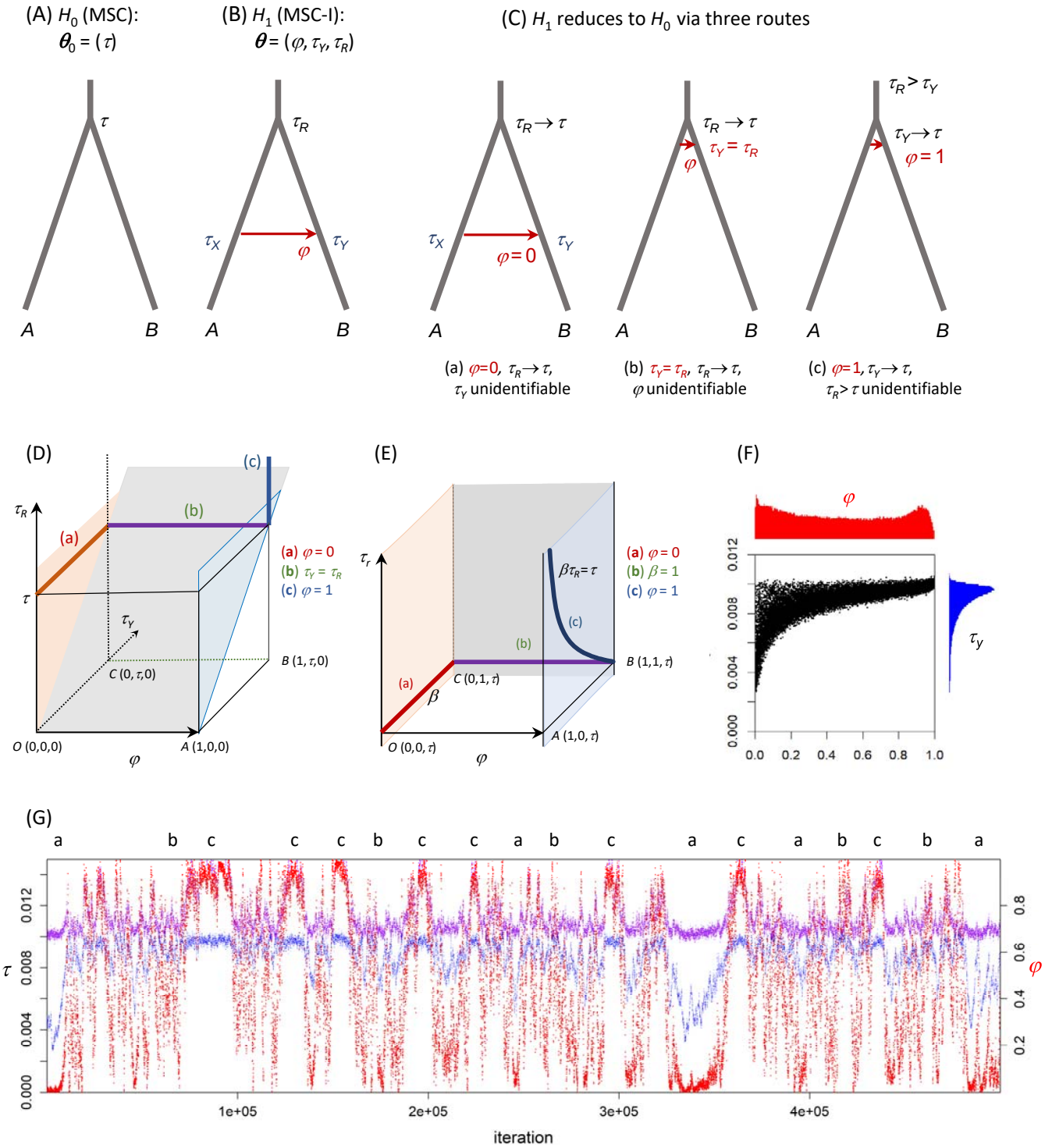

Fig. 1: (**A**, **B**) The null (H_0) and alternative (H_1) hypotheses in the test of (unidirectional) introgression from A to B. The parameter vector is $\boldsymbol{\theta}_0 = (\tau)$ under H_0 and $\boldsymbol{\theta} = (\varphi, \tau_Y, \tau_R)$ under H_1, with the population size parameter (θ) suppressed. (**C**) H_1 reduces to H_0 via three routes (constraints): a, b, and c (table 1(i)). (**D**) Parameter space of H_1, showing the null point as three lines a, b, c, which correspond to the three routes of panel **C**, and the null space (the parameter space for H_0) as three planes. (**E**) Reparametrization of H_1, with $\beta = \tau_Y/\tau_R$ used instead of τ_Y, so that $\boldsymbol{\theta}' = (\varphi, \beta, \tau_R)$. (**F**) The 2-D scatter plot (for τ_Y and φ) and (**G**) trace plots for the three parameters in H_1 from an MCMC sample generated in a BPP analysis of a dataset simulated under H_0: purple for τ_R, blue for τ_Y and red for φ (the 2nd y-axis). Labels on top of the trace plot (**G**) indicates that the chain is in the neighborhood of segments a, b, c of panel **D**. For example, close to segment a, $\varphi \approx 0, \tau_R \approx \tau$, while τ_Y is indeterminate. The dataset consists of 500 loci, with 10 sequences per species per locus, and with 500 sites in each sequence, simulated under H_0 with $\tau = 0.01$ and $\theta = 0.005$. The data are analyzed under H_1, using the priors $\theta \sim G(2, 200)$, $\tau_R \sim G(2, 200)$, and $\varphi \sim U(0, 1)$. The same data are analyzed using different priors on φ in Figure S3.

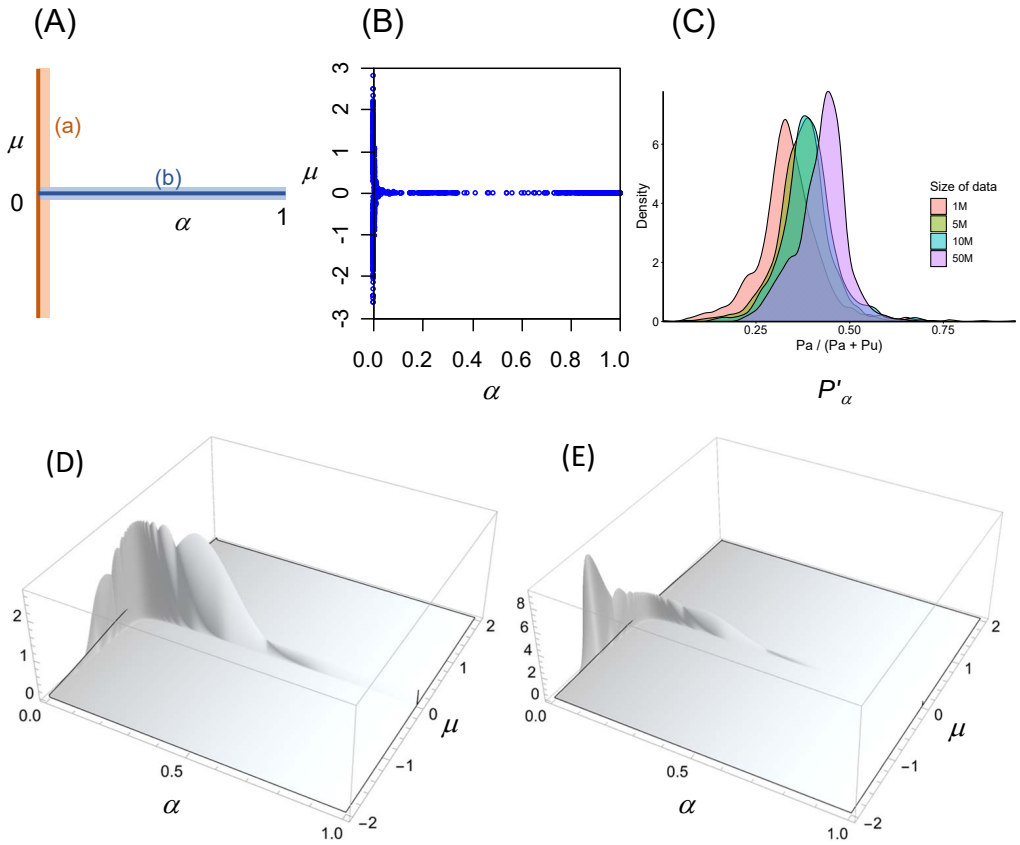

Fig. 2: (**A**) Parameter space for the Gaussian mixture model $H_1 : \alpha N(\mu, 1) + (1 - \alpha)N(0, 1)$ is the α–μ plane, with $0 \le \alpha \le 1$ and $-\infty < \mu < \infty$. There are no unknown parameters in $H_0 : N(0, 1)$, so that the null space is the null point and is represented by the two axes: (a) $\alpha = 0$ (with μ unidentifiable) and (b) $\mu = 0$ (with α unidentifiable). (**B**) MLEs $(\hat{\alpha}, \hat{\mu})$ under H_1 from 10^3 datasets (of size $n = 10^6$) simulated under H_0. The MLE corresponds to the highest ℓ in ten runs of the BFGS algorithm. Note that all points on the two axes have identical likelihood so that the MLE is one unique point only if it is not on the axes. (**C**) Density for posterior probability for segment a (the μ-axis) at $n = 10^6, 5 \times 10^6, 10^7$, and 5×10^7, estimated by simulating 1000 datasets under H_0. Gaussian quadrature is used to calculate P_α and P_μ with $\epsilon_\alpha = 0.01/\sqrt{12}, \epsilon_\mu = 0.01$, which are rescaled to give $P'_\alpha = P_\alpha/(P_\alpha + P_\mu)$. The average of $P_\alpha + P_\mu$ is 0.5491, 0.6849, 0.7348, 0.8340, at the four values of n. Mathematica gave identical results to Gaussian quadrature for $n \le 2000$ but encounters numerical problems for larger n. (**D**, **E**) Unnormalized posterior surface (eq. 17) for two datasets of size $n = 10^6$ simulated under H_0. In the first (**D**), the sample mean is $\bar{x} = -0.001931$, and the MLE under H_1 is $(\hat{\alpha}, \hat{\mu}) = (1, -0.001931)$ with the log likelihood (relative to that for H_0) $\Delta\ell = 1.864$. Multiple local peaks exist close to the α-axis, including (0, any) with $\Delta\ell = 0$, and (0.328984, −0.005867) and (0.791298, −0.002440) both with $\Delta\ell = 1.864$. In the second (**E**), the sample mean is $\bar{x} = -0.001830$. The MLE is (0.000927, −1.571305) close to the μ-axis, with $\Delta\ell = 4.381$, but local peaks exist close to the α-axis, such as (0.717088, −0.002553), (0.802798, −0.002280), and (0.909107, −0.002013) all with $\Delta\ell = 1.675$.

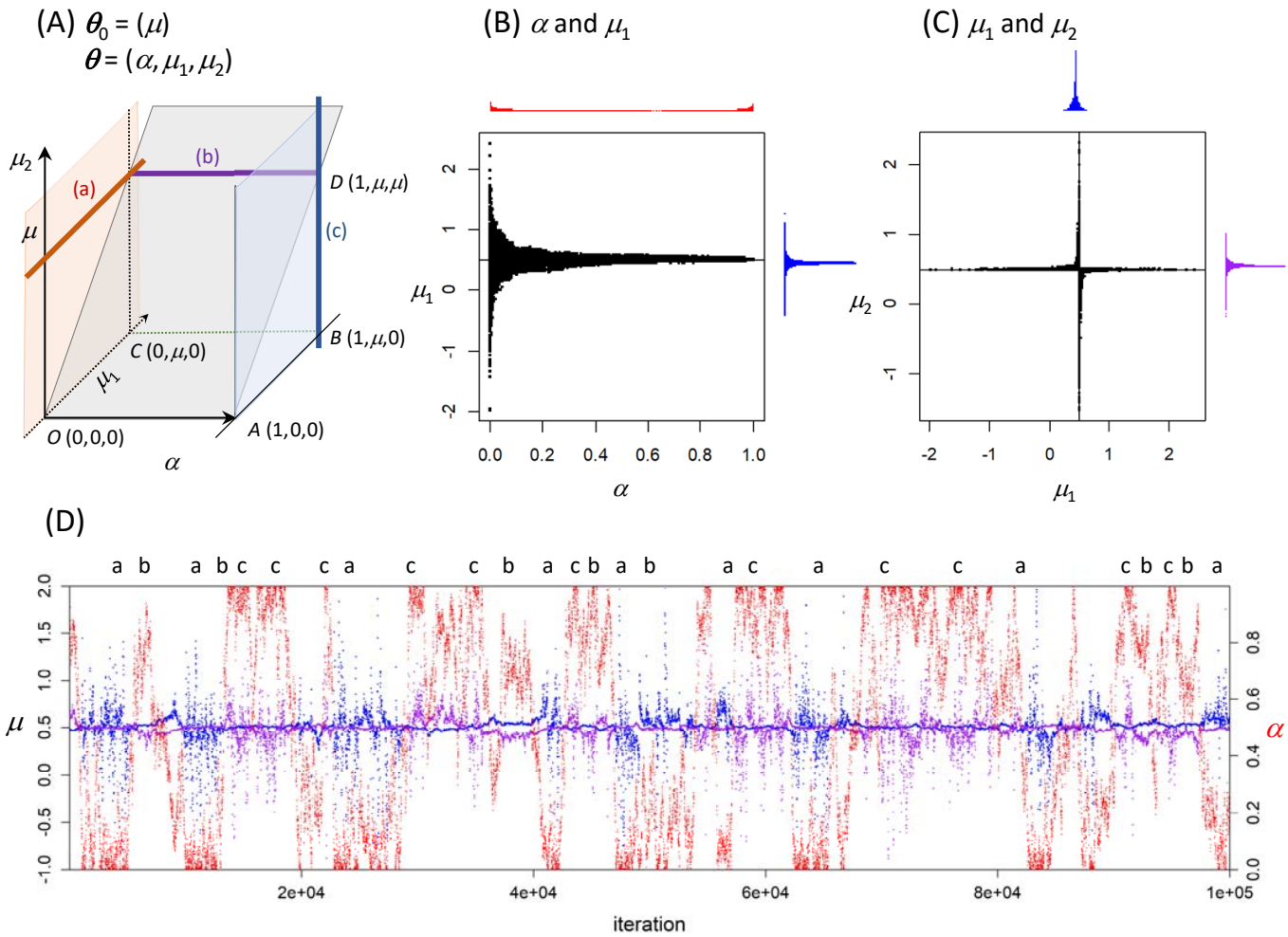

Fig. 3: (**A**) The 3-D parameter space for the Gaussian mixture model $H_1 : \alpha N(\mu_1, 1) + (1-\alpha) N(\mu_2, 1)$, with parameter vector $\boldsymbol{\theta} = (\alpha, \mu_1, \mu_2)$. The null model is $H_0 : N(\mu, 1)$, with $\boldsymbol{\theta}_0 = (\mu)$. The parameter space of H_0 (the null space) corresponds to the three planes while one point in the parameter space of H_0 (the null point or a particular μ value such as $\mu = 0.5$) corresponds to three lines or line segments: (a) $\alpha = 0$ (with μ_1 unidentifiable and with $\mu_2 \to \mu$), (b) $\mu_1 = \mu_2$ (with $0 \le \alpha \le 1$ unidentifiable and with $\mu_1 = \mu_2 \to \mu$), and (c) $\alpha = 1$ (with μ_2 unidentifiable and with $\mu_1 \to \mu$). (**B**, **C**) 2-D scatter plots and (**D**) trace plots of α (red), μ_1 (blue) and μ_2 (purple) from an MCMC analysis under H_1 of a dataset of size $n = 10^5$, simulated under H_0 with $\mu = 0.5$. Labels a, b, c on top of the trace plot (**D**) indicates that the chain is around the three segments of panel **A**.

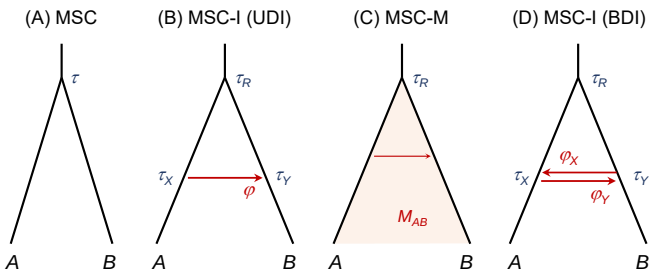

Fig. 4: (**A**) Unidirectional introgression model (MSC-I, UDI) with $A \to B$ (or $X \to Y$) introgression occurring at time $\tau_X \equiv \tau_Y$ and introgression probability φ_Y. (**B**) Continuous migration model (MSC-M) with $A \to B$ migration at the population migration rate M_{AB} per generation. (**C**) Bidirectional introgression model (MSC-I, BDI) with introgression probabilities φ_X and φ_Y and introgression time $\tau_X \equiv \tau_Y$. The MSC-I UDI and MSC-M models (**A**, **B**) are used to simulate data while all three models of gene flow (**B**-**D**) are used to analyze data.

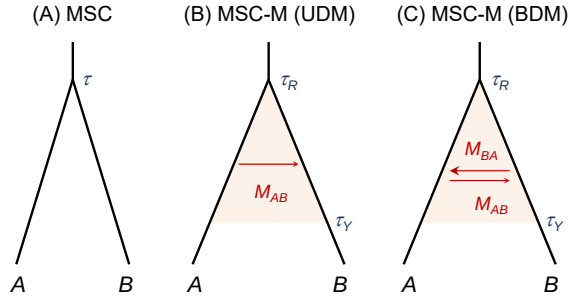

Fig. 5: (**A**) The null model of no gene flow (H_0: MSC). (**B**, **C**) Unidirectional and bidirectional migration (UDM, BDM) models with gene flow stopping at time τ_Y. Both instances of the MSC-M model assumes isolation with initial migration (IIM) (Costa and Wilkinson-Herbots, 2017; see also Huang *et al.*, 2022, Fig. 1B).

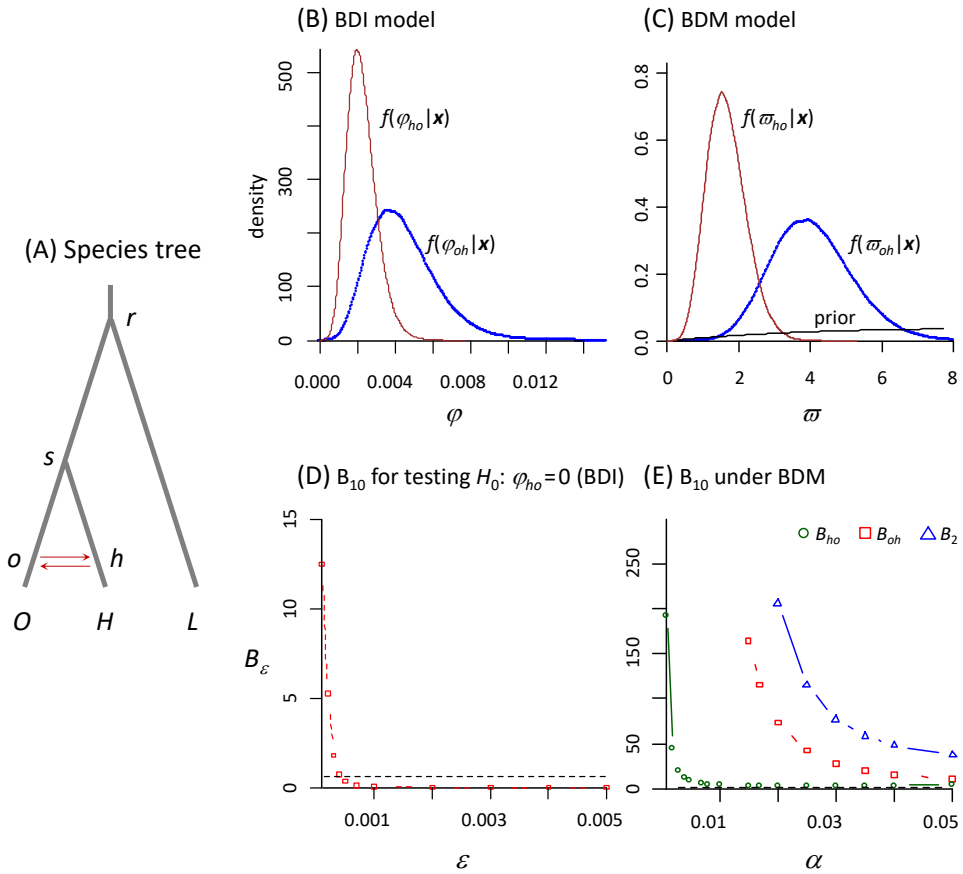

Fig. 6: (**A**) Phylogeny for three Spiny lizard species, *Sceloporus orcutti* (O), *S. hunsakeri* (H), and *S. licki* (L), with gene flow between O and H. (**B**) Posterior densities for introgression probabilities φ_{oh} and φ_{ho} obtained in BPP analysis of the ddRAD data under the BDI model. The prior $\varphi \sim U(0,1)$ is invisible. (**C**) Posterior densities for migration rates (ϖ_{oh}, ϖ_{ho}) under the BDM model. The prior $\varpi \sim G(2, 0.1)$ is shown as well. See table 5 for estimates of other parameters under the two models. (**D**) Approximate Bayes factor ($B_{10,\epsilon}$) for testing $H_0 : \varphi_{ho} = 0$ against $H_1 : \varphi_{ho} \neq 0$, with φ_{oh} treated as a nuisance parameter, calculated using an MCMC sample under BDI of panel B. (**E**) $B_{10,\epsilon}$ for testing $H_0 : \varpi_{ho} = 0$; $H_0 : \varpi_{oh} = 0$; or $H_0 : \varpi_{ho} = 0, \varpi_{oh} = 0$ using the MCMC sample under BDM of panel **D**, plotted against the tail probability of the prior, $\alpha = \mathbb{P}(\varpi < \epsilon)$.

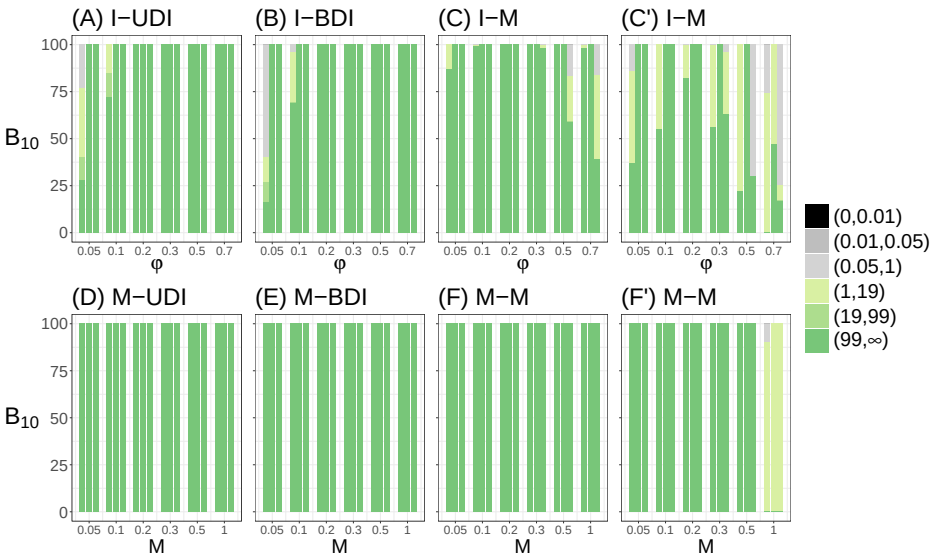

Fig. 7: Power of the Bayesian test of gene flow between sister lineages. Bayes factor (B_{10}) calculated using data simulated under the MSC-I or MSC-M models of figure 4**B**&**C**, and analyzed under the three models of figure 4**B**–**D**. The setting is in the format of simulation model-analysis model: for example, in M-UDI, data are simulated under MSC-M and analyzed under MSC-I (UDI) (Fig. 4**B**&**C**). The MSC-M model was implemented under two priors, $\varpi \sim G(2, 0.01)$ with mean 200 in panels **C** & **F**, and $\varpi \sim G(2, 1)$ with mean 2 in panels **C**′ & **F**′. The stacked bar plots show the proportions of datasets (out of 100 replicates) in which B_{10} falls in different intervals: 0–0.01 (strong rejection of gene flow), 0.01–0.05 (rejection of gene flow), 0.05–1 (preference for no gene flow), 1–19 (preference for gene flow), 19–99 (support for gene flow), and 99–∞ (strong support for gene flow). The three bars in each group correspond to three divergence levels, with $\tau_R = \frac{1}{2}\theta$ (very shallow), θ (shallow), and 10θ (deep). B_{10} is calculated via the S-D density ratio with $\epsilon_\varphi, \epsilon_\beta$ in MSC-I and ϵ_ϖ in MSC-M set to 0.01σ, where σ is the prior standard deviation; use of 0.005σ and 0.05σ produces very similar results.

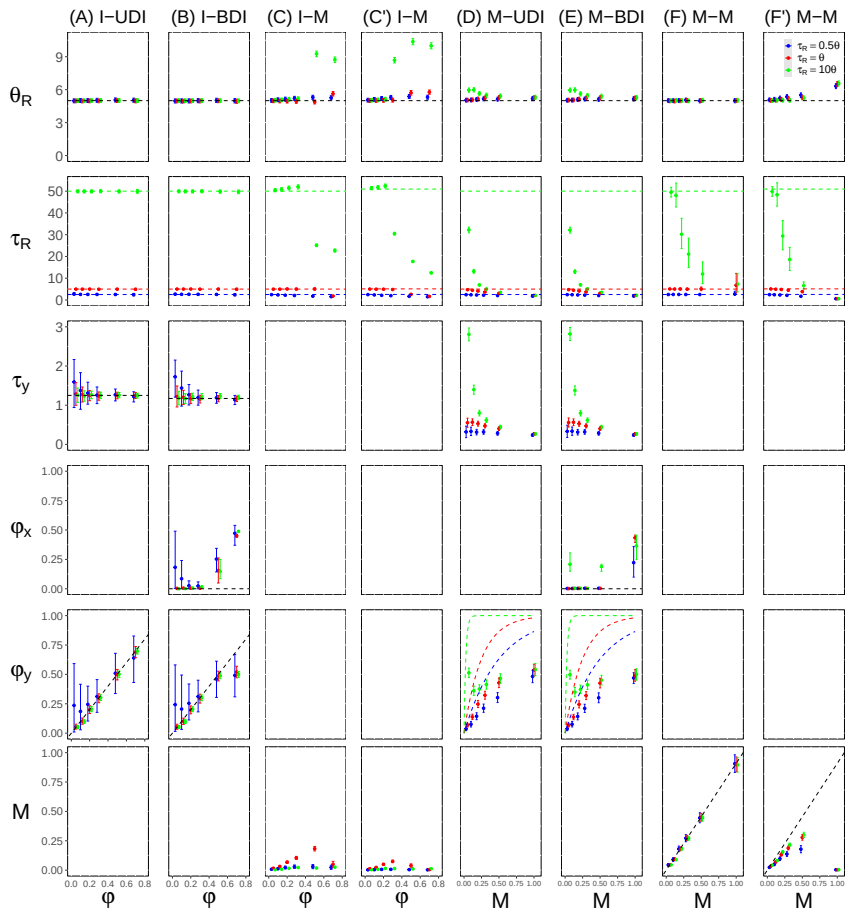

Fig. 8: Average posterior means and 95% HPD CIs of parameters in BPP analysis of data of figure 7. See legend to figure 7. Dotted lines in the plots for φ_y in the M-UDI and M-BDI settings represent the expected amount of gene flow in the true MSC-M model (φ_0 in eq. 37).

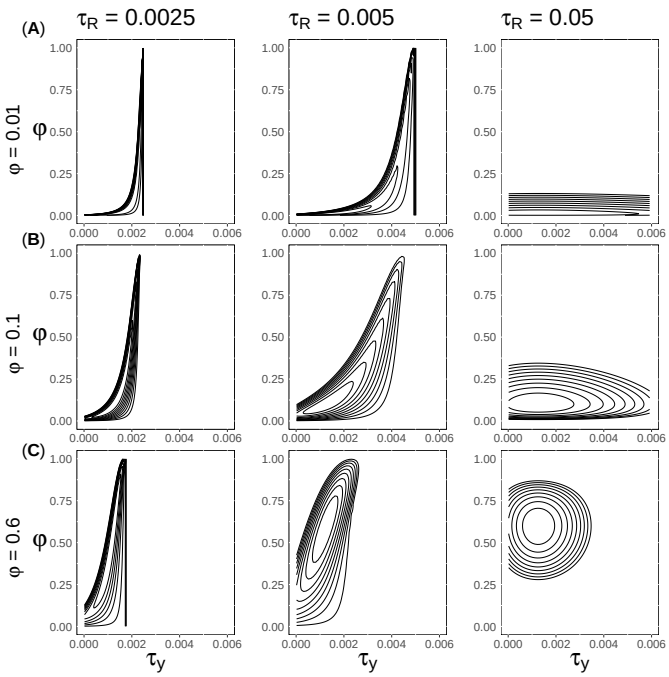

Fig. 9: Contour plot of the log likelihood (eq. 38) as a function of τ_Y and φ under the MSC-I UDI model (Fig. 4**B**) for the limiting case of $L = \infty$ loci, generated using the true parameter values $\varphi = 0.01, 0.1$ or 0.6, and $\tau_R = \frac{1}{2}\theta$, θ, or 10θ, with $\theta = 0.005$ and $\tau_Y = \frac{1}{4}\theta$. Each locus consists of two sequences (one from each species), summarized as the number of differences at n sites. Here $\theta = 0.005$ is fixed at the true value, while τ_R is optimized given φ and τ_Y; fixing τ_R at the true value produces nearly identical results.

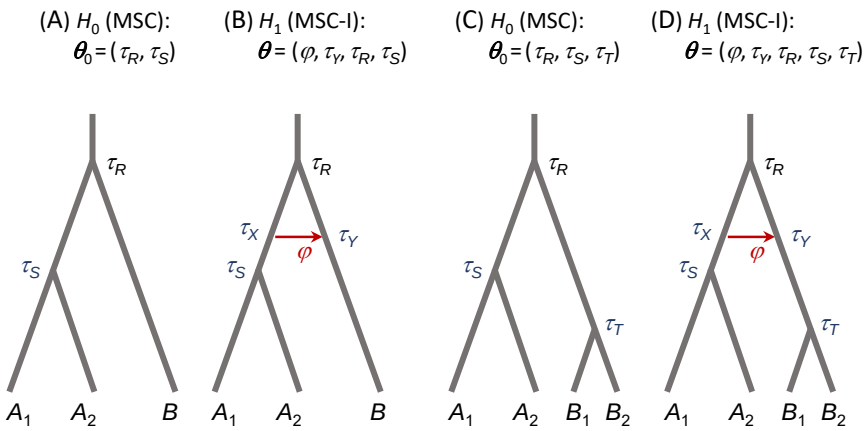

Fig. 10: (**A**, **C**) Null (H_0) and (**B**, **D**) alternative (H_1) hypotheses for testing (unidirectional) introgression that involves one or two ancestral (sister) species. Population-size parameters (θ) are not shown.