Journal (2050), Vol. 00, No. 00, pp. 1–21
DOI:00.0000/xxxxxx/xxxxxxx

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

Jiayi Ji,[1] Donavan J. Jackson,[2] Adam D. Leaché (orcid: 0000-0001-8929-6300),[2] and Ziheng Yang (orcid: 0000-0003-3351-7981)[1,*]
[1]Department of Genetics, Evolution and Environment, University College London, London, WC1E 6BT, UK
[2]Department of Biology and Burke Museum of Natural History and Culture, University of Washington, Box 351800, Seattle, WA 98195-1800, USA

Received on xxxx, revised on xxxx, accepted on xxxx

In the past two decades genomic data have been widely used to detect historical gene flow between species in a variety of plants and animals. The *Tamias quadrivittatus* group of North America chipmunks, which originated through a series of rapid speciation events, are known to undergo massive amounts of mitochondrial introgression. Yet in a recent analysis of targeted nuclear loci from the group, no evidence for cross-species introgression was detected, indicating widespread cytonuclear discordance. The study used the heuristic method HyDe to detect gene flow, which may suffer from low power. Here we use the Bayesian method implemented in the program BPP to reanalyze these data. We develop a Bayesian test of introgression, calculating the Bayes factor via the Savage-Dickey density ratio using the Markov chain Monte Carlo (MCMC) sample under the model of introgression. We take a stepwise approach to constructing an introgression model by adding introgression events onto a well-supported binary species tree. The analysis detected robust evidence for multiple ancient introgression events affecting the nuclear genome, with introgression probabilities reaching 63%. We estimate population parameters and highlight the fact that species divergence times may be seriously underestimated if ancient cross-species gene flow is ignored in the analysis. We examine the assumptions and performance of HyDe, and demonstrate that it lacks power if gene flow occurs between sister lineages or if the mode of gene flow does not match the assumed hybrid speciation model with symmetrical population sizes. Our analyses highlight the importance of using adequate statistical methods to reach reliable biological conclusions concerning cross-species gene flow.

Bayesian test | BPP | chipmunks | introgression | MSci | multispecies coalescent | Savage-Dickey density ratio

Introduction

Genomic sequence data are a rich source of information concerning the history of species divergences and cross-species gene flow. The past two decades have seen widespread use of genomic data to infer hybridization or introgression (Mallet *et al.*, 2016). Gene flow has been detected in a variety of species including Arabidopsis (Arnold *et al.*, 2016), butterflies (Martin *et al.*, 2013), corals (Mao *et al.*, 2018), lizards (Finger *et al.*, 2022), birds (Ellegren *et al.*, 2012), and mammals (Kumar *et al.*, 2017; Chan *et al.*, 2013; Shi and Yang, 2018). The studies have considerably enriched our understanding of the evolutionary dynamics of introgressed genes, and the role of introgression in speciation and ecological adaptation (Payseur and Rieseberg, 2016; Martin and Jiggins, 2017).

A number of statistical methods have been developed to analyze genomic sequence data to detect gene flow between species and to estimate its strength (as measured by the introgression probability or migration rate). Heuristic or summary methods are based on summaries of the multilocus sequence data and include the popular *D*-statistic or ABBA-BABA test (Patterson *et al.*, 2012), HyDe (Blischak *et al.*, 2018), and SNaQ (Solis-Lemus and Ane, 2016). The *D*-statistic and HyDe use the site-pattern counts for a species quartet to test for the presence of gene flow between non-sister species, while SNaQ uses the frequencies of estimated gene tree topologies. Likelihood methods use the multilocus sequence alignments directly and include the Bayesian implementations of the introgression model in PhyloNet/MCMC-seq (Wen and Nakhleh, 2018), *BEAST (Zhang *et al.*, 2018), and BPP (Flouri *et al.*, 2020), as well as the maximum-likelihood and Bayesian implementations of the continuous-migration model (also known as the isolation-with-migration or IM model) (Nielsen and Wakeley, 2001; Zhu and Yang, 2012; Dalquen *et al.*, 2017; Hey *et al.*, 2018). See Jiao *et al.* (2021) for a recent review. In theory, likelihood methods are expected to be more powerful because they use all information in the data about the model and

[*] Correspondence: z.yang@ucl.ac.uk

parameters. However, summary and likelihood methods for inferring cross-species gene flow are seldom applied to the same real datasets with their utilities evaluated, partly because likelihood methods typically involve intensive computation and may not be computationally feasible for genome-scale datasets. In this regard, it is noteworthy that the BPP implementation of the multispecies-coalescent-with-introgression (MSci) model has been successfully applied to genomic datasets of more than 10,000 loci (Flouri *et al.*, 2020, table 1; Thawornwattana *et al.*, 2022, table S4).

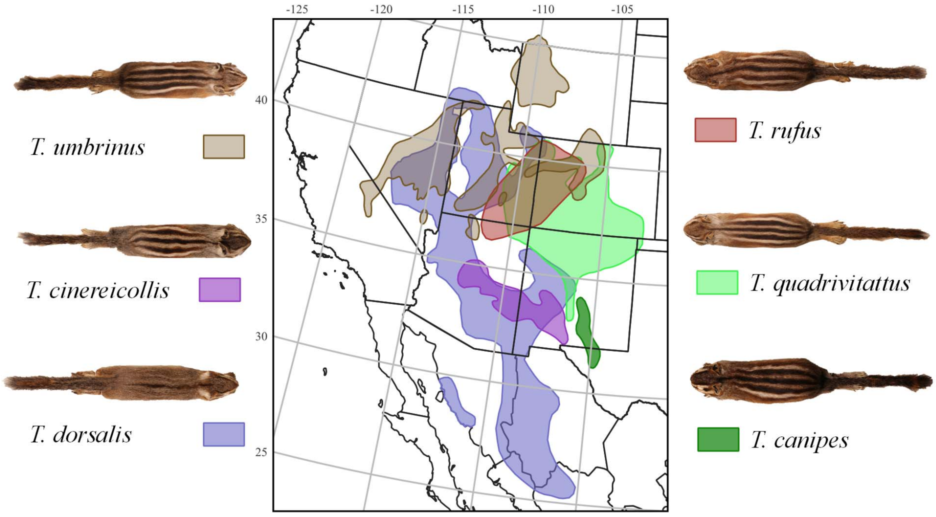


Figure 1: Geographic distributions of the six chipmunk species in the *Tamias quadrivittatus* group, based on data downloaded from the IUCN (`https://www.iucnredlist.org/`).

Table 1. Summary of evidence for mitochondrial introgression in the *T. quadrivittatus* group (Sullivan *et al.*, 2014)

Species	Region	Distribution	Introgression	Source
T. bulleri	M	Allopatric	No	
T. canipes (C)	GB/RM	Allopatric	No	
T. cinereicollis (I)	GB/RM	Parapatric	Yes	Not assignable
T. dorsalis (D)	GB/RM	Parapatric	Yes	C/U/Q/Not assignable
T. durangae	M	Allopatric	No	
T. palmeri	GB/RM	Allopatric	Untested	
T. quadrivittatus (Q)	GB/RM	Parapatric	Yes	Not assignable
T. rufus (R)	GB/RM	Allopatric	No	
T. umbrinus (U)	GB/RM	Parapatric	Yes	Not assignable

Note.— Geographic regions include Great Basin (GB), Rocky Mountains (RM), and Mexico (M). Single letter codes are for the six species included in the nuclear data analysis.

The *Tamias* chipmunks (*sensu lato*, but see Patterson and Norris, 2016) are a diverse group of at least 23 distinct species, occupying a variety of habitats in the western United States. Molecular phylogenetic studies have revealed a complex history of radiative speciations and cross-species gene flow involving morphologically and ecologically diverse lineages (Good and Sullivan, 2001; Good *et al.*, 2003).

The *Tamias quadrivittatus* group of chipmunks currently consists of nine species that are distributed across the Great Basin along with the central and southern Rocky Mountains in North America (fig. 1). Previous work on *Tamias* has highlighted the importance of genital morphology, specifically the baculum (a bone found in the penis) in male chipmunks, as a reliable indicator of species limits (Patterson and Thaeler Jr, 1982; White, 2010). The biogeographic history of the group likely included large range fluctuations that have periodically resulted in isolation and secondary contact among species, which would have affected opportunities for hybridization and/or introgression (Good *et al.*, 2003). The current distributions of species in the group has extensive regions of overlap and broad parapatry in ecological transition zones (fig. 1), with instances of both allopatry and parapatry, and the determinants of current distributions are thought to be related primarily to competitive exclusion and ecological preference (Brown, 1971; Heller, 1971; Root *et al.*, 2001). The system provides an exciting opportunity to investigate the effects of introgression on genetic variation within and between species.

Hybridization between chipmunk species has been widely reported based on discrepancies between mtDNA, nuclear DNA, and morphology (Good and Sullivan, 2001; Good *et al.*, 2003, 2008; Hird *et al.*, 2010). Work in the past decade has documented widespread mitochondrial introgression among species of the group (Reid *et al.*, 2012; Sullivan *et al.*, 2014; Sarver *et al.*, 2017, 2021), which is often asymmetrical, possibly due to bacular morphology, which has been identified in at least six species (Good *et al.*, 2003, 2008; Reid *et al.*, 2012; Sullivan *et al.*, 2014). Recent work on six species in the *T. quadrivittatus* group found that four of them exhibited clear evidence of introgressed mitochondrial DNA: *T. cinereicollis*, *T. dorsalis*, *T. quadrivittatus*, and *T. umbrinus* (table 1). The cliff chipmunk (*T. dorsalis*) was involved in local introgression with multiple other species, receiving mtDNA from whichever congeneric chipmunk it came into contact with. However, populations of *T. dorsalis* that are geographically isolated carry mtDNA haplotypes that are unique to the species (Sullivan *et al.*, 2014; Sarver *et al.*, 2017). Range overlap in transition zones plays an important role in mitochondrial introgression in *Tamias* (Brown, 1971; Bi *et al.*, 2019).

Sarver *et al.* (2021) used a targeted sequence-capture approach to sequence thousands of nuclear loci (mostly genes or exons) to estimate the species phylogeny of the *T. quadrivittatus* group and to infer possible nuclear introgression. The program HYDE (Blischak *et al.*, 2018) was used to infer gene flow. Surprisingly, no significant evidence for gene flow involving the nuclear genome was detected between any species in the group, despite the evidence for widespread mitochondrial introgression. We note that HYDE, like the *D*-statistic, uses the four-taxon site-pattern counts pooled across the genome as data, and does not use information in the variation in genealogical history across the genome caused by the stochastic fluctuation of coalescent and introgression (Lohse and Frantz, 2014; Jiao *et al.*, 2021; Zhu and Yang, 2021). As a result, neither the *D*-statistic nor HYDE can detect gene flow between sister species or populations. Importantly, HYDE is designed to estimate the relative genetic contributions of the two parental species which hybridized to form a third species. When applied to detect other modes of gene flow, it makes restrictive assumptions about the direction of gene flow, and about species divergence times and population sizes that may be unrealistic (see fig. 7 below). The performance of HYDE when its model assumptions are violated is unexplored.

To examine whether the lack of evidence for nuclear introgression in the analysis of Sarver *et al.* (2021) may be due to the lack of power of HYDE, here we re-analyze the data of Sarver *et al.* (2021) using the BPP program (Flouri *et al.*, 2018, 2020), which includes a Bayesian implementation of the MSci model. Borrowing ideas from stepwise regression or Bayesian variable selection, we add introgression events sequentially onto the binary species tree to construct a joint MSci model with multiple introgression events. We develop a Bayesian test of introgression, calculating the Bayes factor for comparing the null model of no introgression against the alternative model of introgression via the Savage-Dickey density ratio (Dickey, 1971), using a Markov chain Monte Carlo (MCMC) sample under the MSci model. This may have a computational advantage over cross-model MCMC algorithms such as reversible jump MCMC (Green, 1995) or calculation of Bayes factors using thermodynamic integration (Gelman and Meng, 1998; Lartillot and Philippe, 2006). Our re-analysis revealed robust evidence for several ancient introgression events affecting the nuclear genome in the *Tamias* group, involving both sister species and nonsister species. We examine the model assumptions underlying HYDE and use computer simulation to demonstrate that the opposite conclusions reached in the two analyses may be explained by the lack of power of HYDE to detect gene flow. We then assess the impact of ignoring introgression on estimation of population parameters, highlighting serious biases in species divergence time estimation when introgression exists and is ignored. Our results highlight the power of coalescent-based likelihood methods in the analysis of genomic datasets to infer the history of species divergence and gene flow.

Theory: Bayesian test of introgression

Bayes factor is given by the Savage-Dickey density ratio in comparisons of nested hypotheses

One can test for the presence of cross-species gene flow by comparing the introgression (MSci) model with the corresponding multispecies coalescent (MSC) model with no gene flow. The model of no gene flow (H_0) is a special case of the introgression model (H_1), with H_1 reducing to H_0 when the introgression probability is 0.

The commonly used device for Bayesian model comparison is the Bayes factor, which is the ratio of the marginal likelihood values under the two compared models. When the two models are nested, the Bayes factor is given by the Savage-Dickey density ratio (Dickey, 1971). In general, suppose we wish to compare the null model $H_0 : \phi = \phi_0$ against the alternative model $H_1 : \phi \neq \phi_0$, and suppose that both models have common (nuisance) parameters λ, while parameters ξ in H_1 become unidentifiable when $\phi = \phi_0$. The parameter vector is λ for H_0 and (ϕ, λ, ξ) for H_1. Given data x, let the likelihood be $L_0(\lambda)$ under H_0 and $L(\phi, \lambda, \xi) = p(x|\phi, \lambda, \xi)$ under H_1, with $L(\phi_0, \lambda, \xi) = L_0(\lambda)$ as the two models are nested. Let the prior be $\pi_0(\lambda)$ under H_0 and $\pi(\phi, \lambda, \xi) = \pi(\phi)\pi(\lambda|\phi)\pi(\xi|\phi, \lambda)$ under H_1. The Bayes factor in support of H_1 over H_0 is defined as

$$B_{10} = \frac{m}{m_0} = \frac{\iiint \pi(\phi,\lambda,\xi)L(\phi,\lambda,\xi)\,\mathrm{d}\phi\mathrm{d}\lambda\mathrm{d}\xi}{\int \pi_0(\lambda)L_0(\lambda)\,\mathrm{d}\lambda}, \tag{1}$$

where m_0 and m are the marginal likelihoods for the two models respectively.

Under the assumption that the priors on the common parameters (λ) agree between the two models, with

$$\pi(\lambda|\phi_0) = \pi_0(\lambda), \tag{2}$$

B_{10} can be expressed as the ratio of the prior and posterior densities for ϕ in H_1, both evaluated at the null value ϕ_0:

$$B_{10} = \frac{m}{m_0} = \frac{\pi(\phi_0)}{\pi(\phi_0|x)}, \tag{3}$$

where $\pi(\phi|x) = \iint \pi(\phi,\lambda,\xi|x)\,\mathrm{d}\xi\mathrm{d}\lambda$ is the marginal posterior density of ϕ.

Proof. Rewrite the prior $\pi_0(\lambda)$ and likelihood $L_0(\lambda)$ under H_0 as densities under H_1.

$$\begin{aligned} B_{10} &= \frac{m}{\int \pi_0(\lambda)L_0(\lambda)\,\mathrm{d}\lambda} \\ &= \frac{m}{\int \pi(\lambda|\phi_0)L_0(\lambda)\,\mathrm{d}\lambda} \\ &= \frac{m}{\int\int \frac{\pi(\phi_0,\lambda,\xi)}{\pi(\phi_0)}L(\phi_0,\lambda,\xi)\,\mathrm{d}\xi\mathrm{d}\lambda} \\ &= \frac{\pi(\phi_0)}{\int\int \frac{1}{m}\pi(\phi_0,\lambda,\xi)L(\phi_0,\lambda,\xi)\,\mathrm{d}\xi\mathrm{d}\lambda} \\ &= \frac{\pi(\phi_0)}{\iint \pi(\phi_0,\lambda,\xi|x)\,\mathrm{d}\xi\mathrm{d}\lambda} \\ &= \frac{\pi(\phi_0)}{\pi(\phi_0|x)}. \end{aligned} \tag{4}$$

Thus eq. 3 holds even if there exist nuisance parameters (λ) in both models, if the null values (ϕ_0) are at the boundary of the parameter space in H_1, and if some parameters in H_1 (ξ) become unidentifiable when the parameters of interest take the null values (when $\phi = \phi_0$). The proof above is more general than that given by Dickey (1971), which does not deal with the unidentifiability of ξ. Note that such irregular conditions cause considerable difficulties for likelihood ratio test (LRT), leading to unknown null distributions for the test statistic (e.g., Self and Liang, 1987). It is interesting that they do not cause any difficulty for the Bayesian test.

If the condition on the priors (eq. 2) does not hold, a correction factor may be applied (Verdinelli and Wasserman, 1995). This is not needed in our application.

Calculation of the Savage-Dickey density ratio

The prior density $\pi(\phi_0)$ of eq. 3 is typically available analytically. The posterior density $\pi(\phi_0|x)$ can be estimated using a kernel density smoothing procedure using the MCMC sample under H_1 (Silverman, 1986). This means that calculation of B_{10} using eq. 3 requires running the MCMC under H_1 only and no cross-model algorithms such as reverse-jump MCMC (Green, 1995) are needed. Note that within-model MCMC typically has better mixing properties than cross-model algorithms (Yang, 2014, pp. 247-260).

Suppose $(\phi^{(1)}, \phi^{(2)}, \cdots, \phi^{(N)})$ are an MCMC sample from the posterior $\pi(\phi|x)$. These are the ϕ values sampled during the MCMC, with the values for other parameters (λ and ξ) simply ignored. The kernel density estimator at the point ϕ_0 is

$$\hat{\pi}(\phi_0|x) = \frac{1}{Nh}\sum_{i=1}^{N} K\Big(\frac{\phi_0 - \phi^{(i)}}{h}\Big), \tag{5}$$

where $K(\cdot)$ is the kernel smoothing function and h is the smoothing parameter or window width. A good choice of h is

$$h = 0.9 \cdot \min\Big(\mathrm{SD}, \frac{\text{inter-quartile range}}{1.34}\Big) \times N^{-\frac{1}{5}} \tag{6}$$

(Silverman, 1986, eq. 3.30-3.31, p.47). The kernel function K is typically symmetrical around 0, with points further away from ϕ_0 make less contribution to the density at ϕ_0. For example, the Gaussian kernel is given as

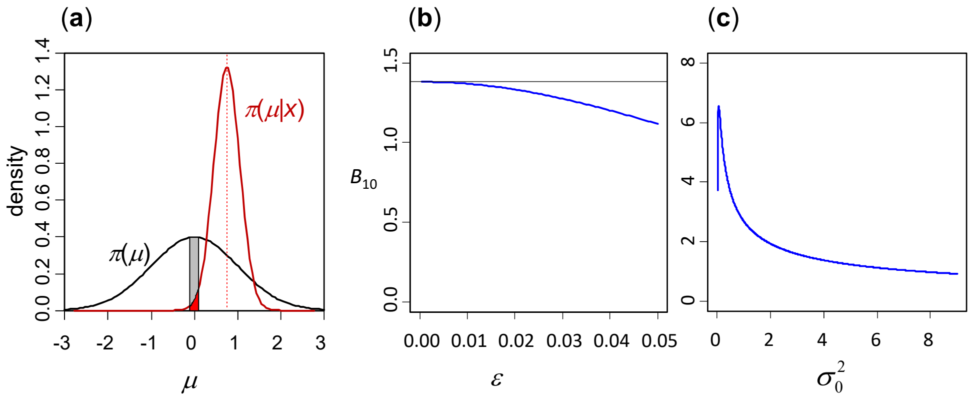


Figure 2: (**a**) Bayes factor expressed as the Savage-Dickey density ratio in the test of the null hypothesis $H_0 : \mu = 0$ against the alternative hypothesis $H_1 : \mu \neq 0$, using a data sample from $\mathbb{N}(\mu, 1)$. The black and red curves represent the prior and posterior densities for μ in H_1, and the small interval (of width ε) in the parameter space for H_1 is the null interval $\varnothing$ (or interval of null effects), representing H_0. The prior and posterior probabilities over the null interval (the gray and red areas) depend on the interval width (ε), but when $\varepsilon \to 0$, their ratio converges to the Bayes factor $B_{10} = \frac{\pi(\mu_0)}{\pi(\mu_0|x)}$. If the area of null effects shrinks greatly when we move from the prior to the posterior, the data contain strong evidence against H_0. (**b**) Approximate Bayes factor $B_{10,\varepsilon} = \frac{\mathbb{P}(\varnothing)}{\mathbb{P}(\varnothing|x)}$ (eq. 8) plotted against ε for a dataset of size $n = 100$ with the sample mean $\bar{x} = 0.258$. The prior is $\mu \sim \mathbb{N}(0, \sigma_0^2)$ with $\sigma_0 = 2$ (twice the sampling standard deviation). When $\varepsilon \to 0$, $B_{10} = 1.381$. (**c**) Bayes factor (eqs. 1 or 13) plotted against the prior variance σ_0^2 for the same dataset showing the sensitivity of B_{10} to the prior on the parameter of interest (μ). Note that in this dataset (with $\sqrt{n}|\bar{x}| = 2.58$) H_0 is rejected by the LRT with p-value 1%.

$$K(t) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-t^2/2}. \tag{7}$$

However, this approach may be awkward to apply if the prior or posterior density at the null value, $\pi(\phi_0)$ or $\pi(\phi_0|x)$, is 0 or ∞. In this paper, we use a more intuitive way of deriving the Savage-Dickey density ratio of eq. 3, which also provides an approach to its calculation. This treats the problem of testing as a problem of estimation, and assesses how likely the parameter of interest (ϕ) differs from the null value (ϕ_0). Define a null region or region of null effects, $\varnothing : |\phi - \phi_0| < \varepsilon$, inside which ϕ is very close to ϕ_0. The null region is a small part of the parameter space for H_1 that represents H_0 (fig. 2). We then define a Bayes factor to represent the evidence for H_1

$$B_{10,\varepsilon} = \frac{1 - \mathbb{P}(\varnothing|x)}{\mathbb{P}(\varnothing|x)} \Bigg/ \frac{1 - \mathbb{P}(\varnothing)}{\mathbb{P}(\varnothing)} \approx \frac{\mathbb{P}(\varnothing)}{\mathbb{P}(\varnothing|x)}, \tag{8}$$

as $1 - \mathbb{P}(\varnothing) \approx 1$ and $1 - \mathbb{P}(\varnothing|x) \approx 1$ for small ε. When $\varepsilon \to 0$, $\mathbb{P}(\varnothing) \to \pi(\phi_0)\Delta$ and $\mathbb{P}(\varnothing|x) \to \pi(\phi_0|x)\Delta$, where the differential Δ is the size of the null region, so that $B_{10,\varepsilon} \to \frac{\pi(\phi_0)}{\pi(\phi_0|x)}$, as in eq. 3. Thus the same conclusion is reached whether the problem is considered a testing problem (eqs. 1 or 3) or an estimation problem (eq. 8).

The approach is illustrated in figure 2 using the simple problem of testing $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ using a sample of size n from $\mathbb{N}(\mu, 1)$. The data are summarized as the sample mean $|\bar{x}|$. We assign the prior $\mu \sim \mathbb{N}(0, \sigma_0^2)$ under H_1. The posterior is then $\mu|x \sim \mathbb{N}(\mu_1, \sigma_1^2)$, with $\mu_1 = \frac{n\bar{x}}{n+1/\sigma_0^2}$ and $\frac{1}{\sigma_1^2} = n + \frac{1}{\sigma_0^2}$. The prior and posterior probabilities of the null interval are $\mathbb{P}(\varnothing) = \mathbb{P}\{|\mu| < \varepsilon\} = 1 - 2\phi\left(-\frac{\varepsilon}{\sigma_0}\right) \approx \pi(\mu_0)\Delta$ and $\mathbb{P}(\varnothing|x) = \phi\left(\frac{\varepsilon - \mu_1}{\sigma_1}\right) - \phi\left(\frac{-\varepsilon - \mu_1}{\sigma_1}\right) \approx \pi(\mu_0|x)\Delta$, with the differential to be the width of the null interval, $\Delta = 2\varepsilon$.

The above theory applies generally to Bayesian testing of nested hypotheses. Examples include comparison of different species delimitation models (e.g., one-species versus two-species models) (Yang and Rannala, 2010) and test of migration between species (e.g., two species with and without migration) (Nielsen and Wakeley, 2001).

Test of introgression

When we use the Savage-Dickey density ratio (eq. 3) to test introgression, the nuisance parameters include species divergence times (τ) and population sizes (θ) on the species tree. Since we use the same priors on τ and θ in models with and without introgresion, independent of the introgression probabilities (φ), the assumption of eq. 2 holds. We consider two tests with different assumptions about the population size parameters (fig. 3). In test 1, the

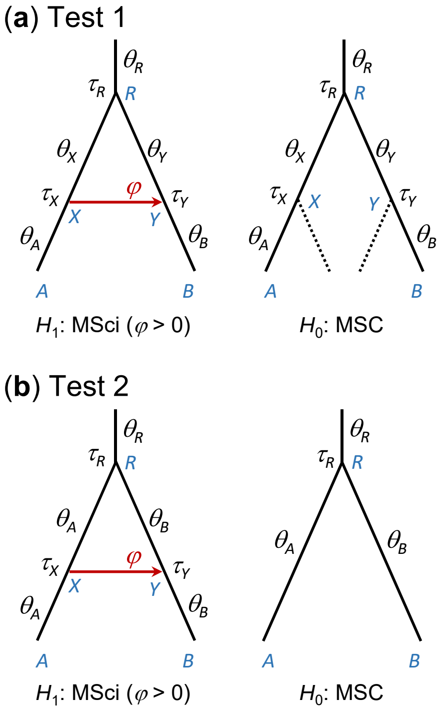


Figure 3: Parameters in the alternative and null hypotheses in two Bayesian tests of introgression (i.e., test of $H_0 : \varphi = 0$ against $H_1 : \varphi > 0$). The parameter of interest is the introgression probability φ. In test 1 (**a**), the shared parameters are $\lambda = (\tau_R, \tau_X = \tau_Y, \theta_A, \theta_B, \theta_R, \theta_X, \theta_Y)$. In test 2 (**b**), the shared parameters are $\lambda = (\tau_R, \theta_A, \theta_B, \theta_R)$ while $\xi = (\tau_X = \tau_Y)$ in H_1 becomes unidentifiable at the null value $\varphi_0 = 0$. Here only the two species involved in introgression are shown. Including other species on the species tree adds the same set of parameters to the null and alternative hypotheses.

MSci model assigns different θ parameters on the two segments of a branch broken by an introgression event; for example, in figure 3**a** branch RA is broken into two branches RX and XA and assigned θ_X and θ_A, respectively. The null model of no gene flow will have two θ parameters for the branch as well. Such a model can be implemented in BPP by including ghost species in the MSC model from which no sequences are sampled (fig. 3**a**). In the second test, the MSci model assigns the same θ parameter for a branch on the species tree before and after an introgression event (which can be specified using the control variable `thetamodel = linked-msci` in BPP) (fig. 3**b**). When the introgression probability takes the null value (0) in H_1, the introgression time τ_X becomes unidentifiable. The proof of eq. 4 applies to both scenarios. In this study, we used test 1. Note that calculating the Bayes factor using the Savage-Dickey density ratio (eqs. 3 or 8) requires an MCMC sample from H_1 and does not require any analysis or MCMC run under H_0.

In our BPP analysis, the introgression probability φ is assigned a beta prior beta(a,b), and the null hypothesis corresponds to $\varphi_0 = 0$ in H_1. Let the null region be $\emptyset : \varphi < \varepsilon$. Then $\mathbb{P}(\emptyset) = \mathbb{P}(\varphi < \varepsilon)$ in eq. 8 is given by the cumulative distribution function (CDF) for beta(a,b), while $\mathbb{P}(\emptyset|x)$ is simply the proportion of the sampled φ values that are $< \varepsilon$. Intuitively, the null region $\emptyset : \varphi < \varepsilon$ in H_1 represents absence of introgression (as the introgression probability φ is negligibly small), $\frac{1-\mathbb{P}(\emptyset)}{\mathbb{P}(\emptyset)}$ is the prior odds in favor of gene flow, while $\frac{1-\mathbb{P}(\emptyset|x)}{\mathbb{P}(\emptyset|x)}$ is the posterior odds, and B_{10} measures the change in the odds in favour of gene flow when we move from the prior to the posterior. We used $\varepsilon = 0.01$ and confirm that use of $\varepsilon = 0.001$ gave very similar results. A cut-off of 20 for B_{10} may be considered strong evidence in support of H_1 (corresponding to 95% posterior for H_1 if the prior model probabilities for H_0 and H_1 are $\frac{1}{2}$ each), while 100 means extremely strong evidence (corresponding to 99% posterior for H_1).

Materials and Methods

Chipmunk genomic data

The dataset, generated and analyzed by Sarver *et al.* (2021), includes 1060 nuclear loci from six chipmunk species: *T. rufus* (R), *T. canipes* (C), *T. cinereicollis* (I), *T. umbrinus* (U), *T. quadrivittatus* (Q) and *T. dorsalis* (D) (with 5, 5, 9, 10, 11, 11 individuals, respectively), as well as the outgroup *T. striatus* (3 individuals). We included all individuals whether or not their mtDNA was likely to be introgressed. Due to lack of a reference genome, Sarver *et al.* (2021) assembled genomic loci (targeted genes or exons) into contigs using an approach called Assembly by Reduced Complexity (ARC). Filters were then applied to remove missing data (contigs not present across all individuals) and sequences with likely assembly errors. The procedure generated a dataset of 1060 loci (1060 ARC contigs, Sarver *et al.*, 2021), with sequence length ranging from 14 to 1026 bp among loci and the number of variable sites from 0.33% to 15.2%.

High-quality heterozygous sites in the data, as identified by high mapping quality and depth of coverage, are represented using IUPAC ambiguity codes. They are accommodated using the analytical integration algorithm implemented in BPP (Flouri *et al.*, 2018; Gronau *et al.*, 2011). This takes the unphased genotype sequences as data and averages over all possible heterozygote phase resolutions, using their relative likelihoods based on the sequence alignment at the locus as weights (Huang *et al.*, 2021).

Species tree estimation for the T. quadrivittatus group

We used BPP version 4 (Flouri *et al.*, 2018; Rannala and Yang, 2017) to estimate the species tree under the MSC model without gene flow. This is the A01 analysis (`speciesdelimitation=0`, `speciestree=1`) (Yang, 2015).

We assigned inverse-gamma (IG) priors to parameters in the MSC model: $\theta \sim$ IG(3, 0.002) with mean 0.001 for population size parameters and $\tau_0 \sim$ IG(3, 0.01) with mean 0.005 for the age of the root. The shape parameter $\alpha = 3$ means that those priors are diffuse, while the prior means are based on estimates from preliminary runs. Note that both θ and τ are measured in the expected number of mutations per site. The inverse gamma is a conjugate prior for θ and allows the θ parameters to be integrated out analytically, leading to a reduction of parameter space and improved mixing of the MCMC algorithm. We conducted 10 replicate MCMC runs, using different starting species trees. Each run generated 2×10^5 samples, with a sampling frequency of 2 iterations, after a burn-in of 16,000 iterations. Each run took about 70 hours using one thread on a server with Intel Xeon Gold 6154 3.0GHz processors. Convergence was confirmed by consistency between runs. All runs converged to the same species tree (fig. 4**a**), with $\sim$ 100% posterior probability, which had the same topology as the tree inferred by Sarver *et al.* (2021).

Stepwise construction of the introgression model

As the species tree is well supported, apparently unaffected by cross-species introgression, we used the species tree to build an introgression model with multiple introgression events. Our procedure is similar to stepwise regression, the step-by-step method for constructing a regression model that involves adding or removing explanatory variables based on a criterion such as an F-test or t-test.

Our procedure has two stages. In the first stage, we used BPP to fit a number of introgression models, each with only one introgression event, and rank candidate introgression events by their strength (indicated by the introgression probability φ). The analyses of Sarver *et al.* (2021) suggest that mitochondrial introgression affected mostly four species: *T. umbrinus* (U), *T. dorsalis* (D), *T. quadrivittatus* (Q) and *T. cinereicollis* (I). We considered introgression events involving all possible pairs among those four species, as well as another species, QI, the common ancestor of *T. cinereicollis* and *T. quadrivittatus* (fig. 4**a**). The dataset of 1060 loci was analyzed under an MSci model with only one introgression event, estimating the introgression probability (φ) and introgression time (τ). We assign the same inverse-gamma priors on θ and τ as above, and beta(1,1) or $\mathbb{U}(0,1)$ for the introgression probability φ. Two replicate runs were conducted for each analysis to confirm consistency between runs, and MCMC samples from the two runs were then combined to produce posterior estimates of parameters. This analysis provides a ranking of the introgression events by the introgression probability. We calculated the Bayes factor for testing $H_0 : \varphi_0 = 0$ given by the Savage-Dickey density ratio (eq. 3), using the null interval $\emptyset = (0, 0.01)$ (eq. 8); use of $(0, 0.001)$ produced virtually identical results. Only introgresssion events with $B_{10} \geq 20$ were considered further.

In the second stage, we added introgression events onto the binary species tree (fig. 4**a**) sequentially in the order of decreasing strength (introgression probability). To reduce the computational cost and to examine the robustness of the analysis, this step was applied to two subsets of the 1060 loci: the first half and the second half, each of 530 loci. The priors used for population sizes and root age were as above. With multiple introgression events in the model, we extended the MCMC runs to be k-times as long if the model involved k introgression events. Three replicate runs were performed to check consistency between runs. Samples from the replicate runs were then combined to produce posterior summaries. At each step, the added introgression event was retained if it met the same cutoff as above in either of the two data subsets.

Our procedure produced a joint introgression model with three unidirectional introgression events. The joint model was then applied to the full dataset of 1060 loci to estimate the population parameters including introgression probabilities, introgression times, species divergence times, and population sizes (fig. 4**b**), using the same prior

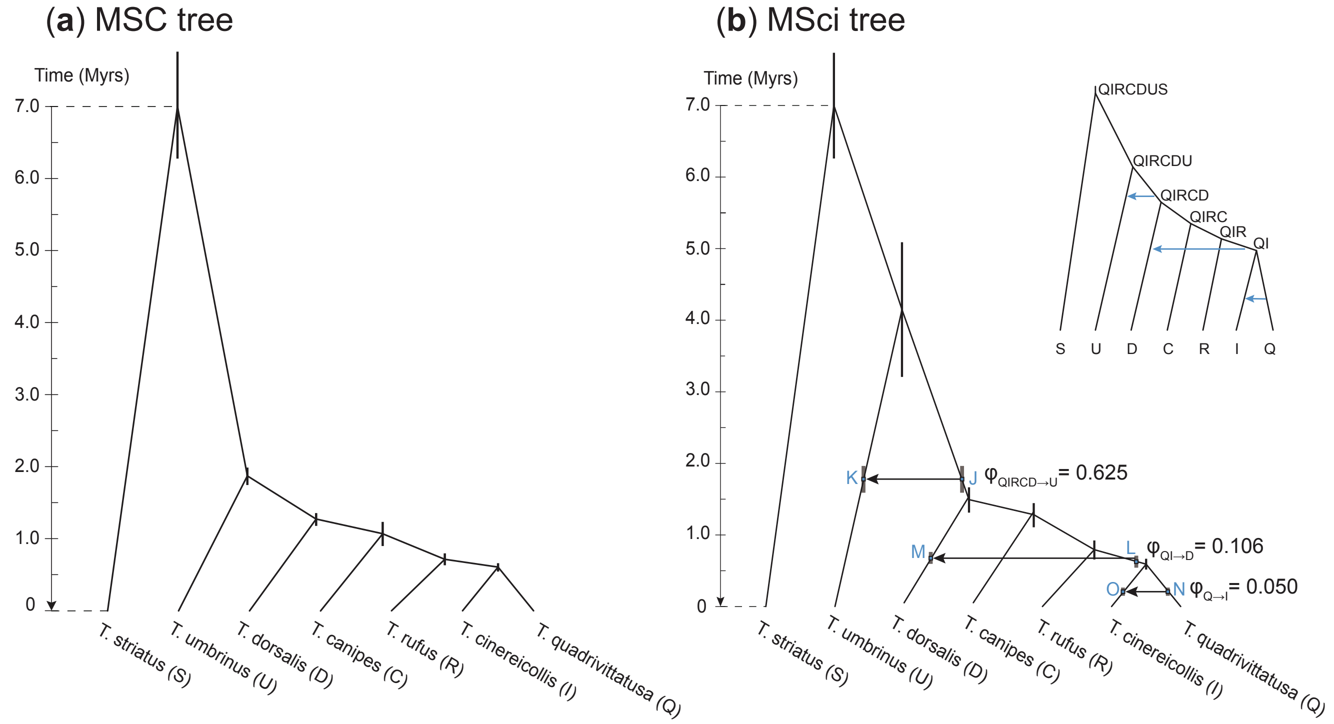


Figure 4: (**a**) Species tree for the *T. quadrivittatus* group with *T. striatus* used as the outgroup. Branch lengths represent the posterior means of divergence times (τ) estimated from BPP analysis of the full data of 1060 loci under the MSC model with no gene flow, with node bars indicating the 95% HPD intervals. A minimum divergence time of 7 Myrs for the outgroup *T. striatus* is used to convert the τ estimates into absolute times. **b**) The joint introgression model constructed in this study with three unidirectional introgression events, showing parameter estimates from BPP analysis of the full data of 1060 loci. Nodes created by introgression events are labeled, with the labels used to identify parameters in table S3. The MSci model includes 6 species divergence times and 3 introgression times (τ), 19 population size parameters (θ), and 3 introgression probabilities (φ).

settings. We conducted 3 replicate runs, using a burn-in of 50,000 iterations and then taking 10^6 samples, sampling every 2 iterations. Each run took 200 hrs.

Results

Species tree estimation for the T. quadrivittatus group

We analyzed the full data of 1060 loci under the MSC model without gene flow to estimate the species tree. The ten replicate runs using different starting species trees converged to the same maximum *a posteriori* probability (MAP) tree, with posterior probability $\sim$ 100% (fig. 4**a**). Sarver *et al.* (2021) recovered the same species tree topology in their analysis of the same data using ASTRAL (Mirarab and Warnow, 2015) and SVDQUARTETS (Chifman and Kubatko, 2014), although with weaker support for some nodes, e.g., concerning the placement of *T. rufus*. The differences in support may be due to the fact that ASTRAL and SVDQUARTETS use summaries of the multilocus sequence data that are not sufficient statistics, and are thus less efficient than the full likelihood method implemented in BPP (Xu and Yang, 2016; Zhu and Yang, 2021).

Stepwise construction of the introgression model

In the first stage of our procedure, we fitted introgression models, each involving one introgression event, using the full dataset of 1060 loci. We considered introgression events between every contemporary pair of the five species: *T. cinereicollis* (I), *T. dorsalis* (D), *T. quadrivittatus* (Q), and *T. umbrinus* (U), and the ancestral species QI (fig. 4**a**). Introgression events that passed our cutoff ($B_{10} \geq 20$) are listed in table 2. Introgression from QI into D had the highest probability, $>$ 10%, while six more events had $\varphi > 5\%$: Q→D, D→QI, QI→U, I→D, Q→I, and I→Q. We note that introgressions between Q and I, and between QI and D, was significant in both directions and the estimated introgressions times were close (table 2). We thus replaced the two unidirectional introgression events by one bidirectional introgression in further analyses (model D in Flouri *et al.*, 2020).

Table 2. Posterior means and 95% HPD CIs (in parentheses) for introgression probability (φ) and introgression time (τ) in the separate introgression analysis

	Introgression	φ	τ ($\times 10^{-3}$)	B_{10}
*	QIRCD → U	0.6215 (0.3907, 0.8243)	0.896 (0.784, 1.004)	∞
*	QI → D	0.1187 (0.0866, 0.1499)	0.337 (0.311, 0.367)	∞
	Q → D	0.0779 (0.0509, 0.1026)	0.297 (0.253, 0.328)	∞
	D → QI	0.0707 (0.0384, 0.1058)	0.337 (0.302, 0.366)	∞
	QI → U	0.0624 (0.0269, 0.1020)	0.408 (0.353, 0.457)	21.27
	I → D	0.0579 (0.0332, 0.0862)	0.265 (0.217, 0.318)	∞
*	Q → I	0.0568 (0.0315, 0.0750)	0.098 (0.073, 0.121)	∞
	I → Q	0.0533 (0.0153, 0.0969)	0.111 (0.077, 0.156)	∞
	D → U	0.0214 (0.0022, 0.0483)	0.276 (0.178, 0.474)	0.04
	Q → U	0.0198 (0.0037, 0.0389)	0.296 (0.209, 0.367)	0.05
	D → I	0.0180 (0.0092, 0.0275)	0.155 (0.123, 0.192)	0.39
	D → Q	0.0177 (0.0058, 0.0315)	0.184 (0.117, 0.347)	0.10
	U → QI	0.0097 (0.0022, 0.0181)	0.371 (0.322, 0.410)	0.01
	I → U	0.0069 (0.0015, 0.0136)	0.158 (0.098, 0.223)	0.00
	U → D	0.0066 (0.0024, 0.0112)	0.235 (0.176, 0.300)	0.00
	U → Q	0.0061 (0.0008, 0.0127)	0.200 (0.119, 0.294)	0.00
	U → I	0.0037 (0.0009, 0.0071)	0.147 (0.090, 0.207)	0.00

Note.— The species tree of figure 4**a** is used, with a single introgression event assumed in each analysis. The full dataset of 1060 loci is analyzed using BPP to estimate the introgression probability (φ) and the introgression time (τ), together with the species divergence times (τ) and population sizes (θ) on the species tree. Introgression events with $B_{10} < 20$ (D → U and below) are not considered further in the stepwise approach of constructing the joint introgression model. The three introgression events that are selected in the joint introgression model are marked with asterisks. Bayes factor $B_{10} = \infty$ occurs if all φ values in the MCMC sample are $> \varepsilon = 1\%$.

The time of QI→U introgression was estimated to be 0.000408, very close to the species divergence time at node QIR (0.000417) (fig. 4**a**), suggesting that the introgression was probably a more ancient event. Note that if an introgression event is assigned incorrectly to a daughter branch to the lineage truly involved in introgression, one would expect the estimated introgression time to collapse onto the species divergence time. We thus attempted to place the introgression onto more ancient ancestral branches on the species tree (fig. 4**a**) and finally identified the lineage involved in introgression to be the ancestral species QIRCD. The QIRCD→U introgression had an estimated time that was away from the species divergence times, and the estimated introgression probability (62%) was the highest (table 2).

In the second stage, we added introgression events identified in table 2 onto the binary species tree of figure 4**a**, in the order of their introgression probabilities (table S1). This was applied to two data subsets (the full data split into two halves). While our procedure allows introgression events already in the model to drop out when new introgressions are added to the model, this did not happen in the analysis of the *Tamias* dataset. Instead the most important introgression events identified in stage 1 remained to be most important in the joint introgression models constructed in stage 2. Note that multiple introgression events may not be independent. An introgression event significant in stage 1 may not be significant anymore when other introgression events are already included in the model. For example, when the QI→D introgression was already included in the model, none of the introgressions Q→D, D→QI, I→D and I→Q was significant. Those introgressions may be expected to lead to similar features in the sequence data, such as reduced sequence divergences between Q or I and D. Similarly, introgression probability for an introgression event often became smaller when other introgressions were added in the model. However, the opposite may occur as well. For example, $\varphi_{QIRCD \to U}$ was estimated to be 54-63% when this was the only introgression assumed in the model, but increased to 59-69% when other introgression events were added in the model (table S1).

Results for the two data subsets were largely consistent, especially concerning introgression events with high introgression probabilities. We thus arrived at a joint introgression model with three unidirectional introgression events (fig. 4**b**, table S1).

We examined the impact of the prior for φ on the Bayesian test of introgression. We calculated the Bayes factor B_{10} using the full dataset of 1060 loci under the prior $\varphi \sim \text{beta}(\alpha, \beta)$, with $\alpha = 0.2, 1, 5$ and $\beta = 0.2, 1, 5$, generating nine prior settings (table S2). Note that beta(α, β) has the mean $\mathbb{E}(\varphi) = \frac{\alpha}{\alpha+\beta}$ and variance $\mathbb{V}(\varphi) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$. In particular, the prior mean varied from 0.0385 for beta(0.2, 5) to 0.961 for beta(5, 0.2). The Bayes factor B_{10} was ∞ for all three introgression probabilities in the joint model, insensitive to the prior on φ (table S2).

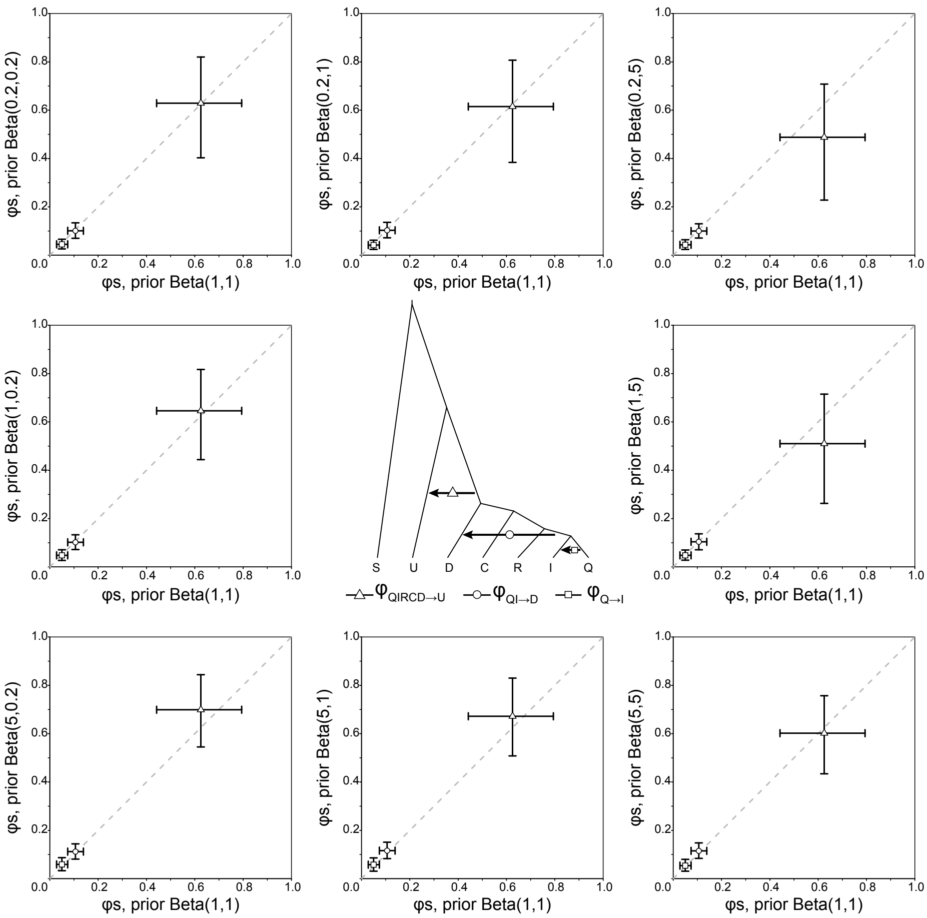


Figure 5: Posterior means and 95% HPD CIs for the three introgression probabilities (φ) obtained from BPP analyses of the full data of 1060 loci using different beta priors, φ ~beta(α, β).

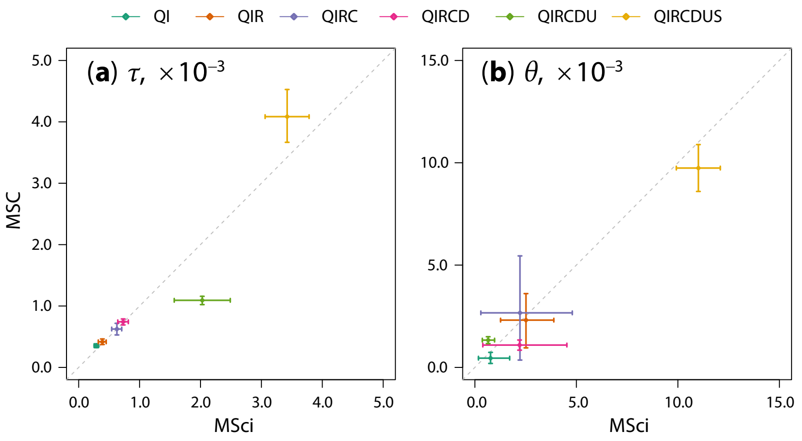
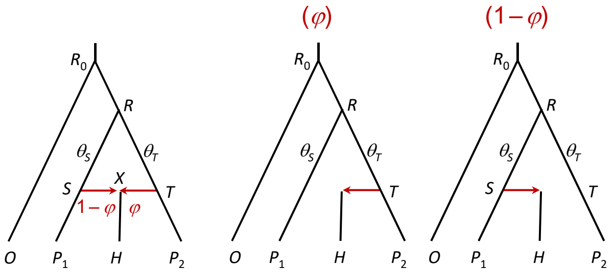


Figure 6: Scatterplot of posterior means and 95% HPD CIs (**a**) for the six species divergence times (τ) and (**b**) for the six ancestral population sizes (θ) in the MSC and MSci models of figure 4 obtained from BPP analyses of the full data of 1060 loci. Note that both τ and θ are measured in the expected number of mutations per site.

Estimation of introgression probabilities and species divergence/introgression times

Finally, we fitted the joint introgression model of figure 4**b** to the full data of 1060 loci, as well as the two halves, with parameter estimates shown in table S3. The fitted model is very parameter-rich, partly as we assign different θ parameters for different branches on the species tree: for example, branch Q in figure 4**b** is broken into two segments by the introgression event, Q→I, which are assigned two independent θ parameters. As a result, population sizes for ancestral species tend to be poorly estimated, especially for those populations with a very short time duration.



(**a**) Hybrid speciation model (**b**) Species tree S_1 (**c**) Species tree S_2

Figure 7: (**a**) HYDE assumes a hybrid-speciation model with the additional assumption of equal population sizes, or a symmetrical inflow model, with $\tau_S = \tau_T$ and $\theta_S = \theta_T$ (Blischak *et al.*, 2018). (**b, c**) Two parental species trees S_1 and S_2 induced by the hybridization model of (**a**). Site patterns are a mixture over the two species trees.

These patterns are consistent with simulation studies that examine the information content in multi-locus datasets (Huang *et al.*, 2020).

The estimated introgression probabilities from the full data are 0.625 with the 95% highest probability density (HPD) credibility interval (CI) to be (0.442, 0.794) for $\varphi_{QIRCD \to U}$, 0.106 (0.074, 0.139) for $\varphi_{QI \to D}$, and 0.050 (0.028, 0.074) for $\varphi_{Q \to I}$. The introgression probability $\varphi_{QIRCD \to U}$ involved considerable uncertainty, with a large CI, possibly because the introgression is ancient and is between sister species, making it hard to estimate its strength, so that the dataset of 1060 loci may be too small.

We evaluated the impact of the prior for φ on parameter estimation in the analysis of the full dataset, using $\alpha = 0.2, 1, 5$ and $\beta = 0.2, 1, 5$ in the prior $\varphi \sim \text{beta}(\alpha, \beta)$ (fig. 5). The prior had some effects on $\varphi_{QIRCD \to U}$, with the prior mean being more important than the prior variance. Under beta(0.2, 5) with the prior mean 0.0385, the posterior mean was lower, and the CI wider. Under beta(5, 0.2) with the prior mean 0.961, the posterior mean was higher, and the CI narrower. However, the posterior CIs overlapped considerably among the different priors, and overall the impact of the prior for φ on the estimate of $\varphi_{QIRCD \to U}$ was minor. Estimates of $\varphi_{QI \to D}$ and $\varphi_{Q \to I}$ were insensitive to the prior used (fig. 5).

Accommodating gene flow in the model had significant impacts on estimation of the time of divergence between species involved in gene flow (figs. 4 & 6). While estimates of times for the recent divergences ($\tau_{QI}, \tau_{QIR}, \tau_{QIRC}$, and τ_{QIRCD}) were nearly identical between the MSC model ignoring gene flow and the MSci model incorporating gene flow, the estimated age of the *T. quadrivittatus* clade (τ_{QIRCDU}) was much greater under MSci than under MSC (fig. 6). This can be explained by the fact that the MSC model ignored the QIRCD→U introgression, which had introgression probability 62.5%. Note that sequence divergence between any pair of species X and Y has to be older than species divergence ($t_{XY} > \tau_{XY}$), and as a result, the minimum (rather than average) sequence divergence dominates the estimate of species divergence time. If gene flow is present between species and is ignored in the model, the reduced sequence divergence due to gene flow will be misinterpreted as recent species divergence, leading to underestimation of species divergence time. This effect has been noted in previous simulations (Leaché *et al.*, 2014).

The estimated age of the root of the species tree ($\tau_{QIRCDUS}$) was slightly smaller under MSci than under MSC. However, $\tau_{QIRCDUS}$ is negatively correlated with the population size ($\theta_{QIRCDUS}$) so that both parameters have large uncertainties (Burgess and Yang, 2008).

Sullivan *et al.* (2014, fig. 1) used the minimum divergence time of 7 Ma for the outgroup species *T. striatus*, based on fossil teeth thought to belong to *Tamias* found in the late Miocene, reported in Dalquest *et al.* (1996), to date the *T. quadrivittatus* clade to 1.8 Ma in a maximum-likelihood concatenation analysis of four nuclear genes, and to 1.2 Ma (with 95% CI 0.6–2.2) in a *BEAST (Heled and Drummond, 2010) analysis of the same data. Concatenation analysis is known to be biased as it does not accommodate the stochastic variation of gene tree topologies and divergence times among loci due to the coalescent process (Ogilvie *et al.*, 2017). We used the same calibration to rescale the estimates of τ under the MSC and MSci models (fig. 4). The minimum age for the *T. quadrivittatus* clade was 1.9 Ma (with 95% HPD CI to be 1.8–2.0) under the MSC model, comparable to the *BEAST estimate under the same model (fig. 4**a**). Under the MSci model, the estimated minimum age was 4.1 Ma (with CI be 3.2–5.1) (fig. 4**b**), much older than the estimates under the MSC model without gene flow. Note that here the CIs accommodate the uncertainty due to finite amounts of sequence data but not uncertainties in the fossil calibration.

Model assumptions underlying HYDE

Whereas the analyses of nuclear data by Sarver *et al.* (2021) using HYDE detected no significant signal of introgression at all, our BPP analyses of the same data revealed strong evidence of multiple introgression events, involving both sister and non-sister species (fig. 4**b**). To understand the opposing conclusions reached in the two analyses, here we examine the model assumptions underlying HYDE. We then use simulation to compare the performance of HYDE and BPP under conditions that are representative of the *Tamias* data but may violate the assumptions of HYDE.

HYDE was developed under the hybrid-speciation model of figure 7**a**, with $\tau_S = \tau_X = \tau_T$, and $\theta_S = \theta_T$ (Blischak *et al.*, 2018). Formulated for quartet data, with one sequence from each of the four species, it uses the counts or frequencies of three parsimony-informative site patterns: $iijj$, $ijji$, $ijij$, to estimate the genetic contributions of the two parental species to the hybrid species: φ and $1-\varphi$. Here pattern $ijkl$ means a site with nucleotides i, j, k, l in O, P_1, H, P_2, respectively (fig. 7**a**). Under this model, the probabilities of gene trees and site patterns are both given by a mixture over the two binary species trees S_1 and S_2 (called *parental species trees*), with mixing probabilities φ and $1-\varphi$ (fig. 7**b**&**c**). Given species tree S_1, the matching pattern $iijj$ has a larger probability (say, a) than the other two mismatching patterns (each with probability b, say, with $b < a$). Given species tree S_2, the matching pattern $ijji$ has probability a while the two mismatching patterns have b each. The symmetry assumptions ($\tau_S = \tau_T$ and $\theta_S = \theta_T$) ensure that a, b for tree S_1 are equal to a, b for S_2. By averaging over the two species trees, the site pattern probabilities under the hybridization model are given as

$$\begin{aligned} p_{iijj} &= \varphi a + (1-\varphi) b \\ p_{ijij} &= \varphi b + (1-\varphi) b = b \\ p_{ijji} &= \varphi b + (1-\varphi) a. \end{aligned} \tag{9}$$

Setting those probabilities to the observed frequencies ($\hat{p}$) and eliminating a and b from the system of equations gives the estimate

$$\hat{\varphi} = \frac{\hat{p}_{iijj} - \hat{p}_{ijij}}{\hat{p}_{iijj} - 2\hat{p}_{ijij} + \hat{p}_{ijji}}, \tag{10}$$

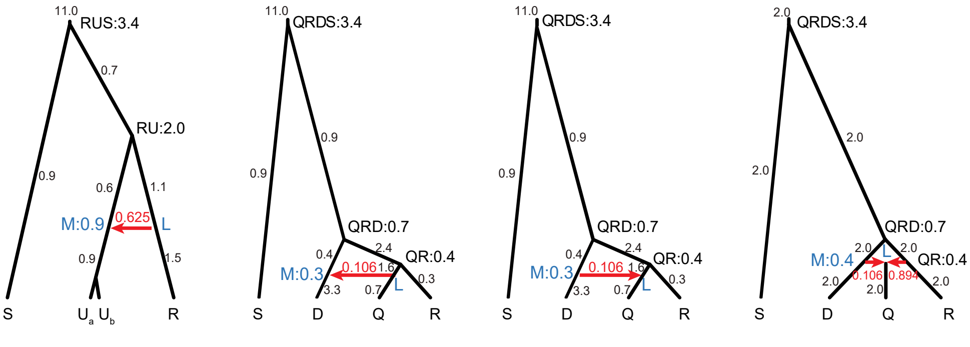
This is eq. 3 in Blischak *et al.* (2018), although the derivation here is simpler than that of Kubatko and Chifman (2019). Note that the theory works if $\tau_S = \tau_T > \tau_X$ and $\theta_S = \theta_T$, so that the method may be used under model A of Flouri *et al.* (2020, fig. 1) with the symmetry assumption. The null hypothesis of no hybridization/introgression ($H_0 : \varphi = 0$) can be tested by applying a normal approximation to the site-pattern counts (Kubatko and Chifman, 2019).

To see which of the two assumptions ($\tau_S = \tau_T$ and $\theta_S = \theta_T$) has more impact, note that a change in τ is comparable with the same amount of change in $\frac{2}{\theta}$. Coalescent may occur in population *RS* (if the H sequence takes the left parental path in the model of fig. 7a), at the rate $\frac{2}{\theta_S}$ over time period $\tau_R - \tau_S$, and it may occur in population *RT* (if the H sequence takes the right parental path), at the rate $\frac{2}{\theta_T}$ over time period $\tau_R - \tau_T$. If $\frac{2(\tau_R - \tau_S)}{\theta_S} = \frac{2(\tau_R - \tau_T)}{\theta_T}$, the probability of coalescent (given that two sequences enter populations S or T) will be the same in the two populations. However, the probabilities of the site patterns depend on the time of coalescent as well as its occurrence. Thus for eq. 10 to be valid, both the rates and the times have to be identical: $\tau_S = \tau_T$ and $\theta_S = \theta_T$.

Note that HYDE or the D-statistic cannot be used to infer gene flow between sister lineages. One might think that HYDE or D could be applicable if two sequences were sampled from the recipient lineage to form a quartet. However this is not the case. With ancient introgression, the two sequences from the same lineage are interchangeable and have the same average genomic distance to the outgroup sequence. Suppose P_1 and H in figure 7**a** are two sequences from the same lineage. Then site patterns $iijj$ and $ijij$ will have the same probability even if $\varphi > 0$.

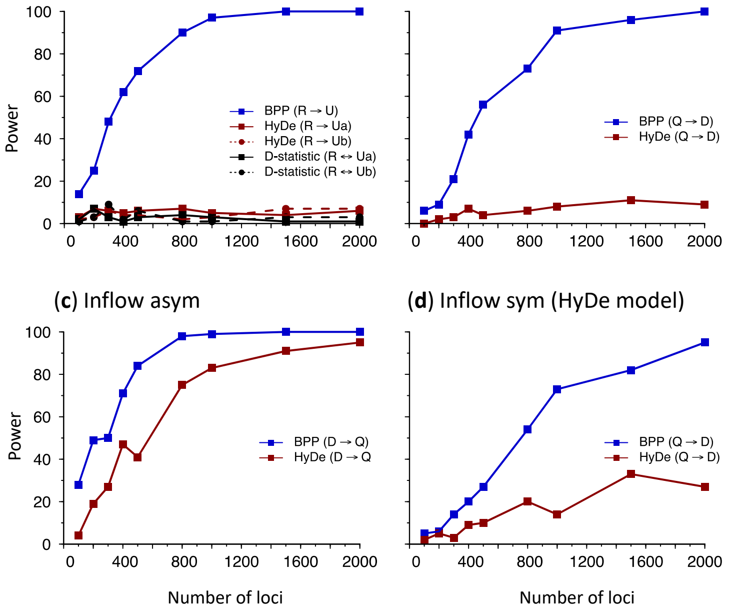
Simulations to examine the performance of HYDE

Our examination of assumptions underlying HYDE suggests that HYDE may not be suitable for testing gene flow in the *Tamias* data. The strongest introgression in the *Tamias* data detected using BPP was between sister species, with $\varphi_{QIRCD\to U} = 0.625$ (fig. 4**b**). This is unidentifiable by HYDE. The next introgression involved outflow with $\varphi_{QI\to D} = 0.106$, whereas HYDE assumes inflow. The third introgression was again between sister species, with $\varphi_{Q\to I} = 0.050$. To verify those expectations and to explore the performance of HYDE and BPP under different scenarios of gene flow, we conducted simulations using four different model settings (fig. 8**a**-**d**), based on parameter estimates obtained from the *Tamias* data (fig. 4**b**, table S3). Gene trees and sequence alignments at multiple loci were generated using the `simulate` option of BPP. HYDE analysis was conducted using PAUP (Swofford, 2003).



(**a**) sister species (**b**) outflow asym (**c**) inflow asym (**d**) inflow sym (HyDe)

Figure 8: Introgression models (species trees with introgression) used for simulating data to evaluate the performance of HyDe and BPP. **(a)** Species tree for three species (R, U and S) with $R \to U$ introgression at the rate of $\varphi = 0.625$, and with S to be the outgroup, based on BPP estimates from the *Tamias* data (fig. 4**b**, table S3). Population sizes (θ) are next to the branches and species divergence times (τ) are next to the nodes. Two sequences are sampled from species U. When the data are analyzed using HyDe, either Ua or Ub is specified as the hybrid lineage. (**b**) Outflow model for three species (D, Q, R), with S to be the outgroup, with introgression from Q to D at the rate $\varphi = 0.106$ (table S3). (**c**) Inflow asymmetrical model for three species, with asymmetrical divergence times and population sizes. (**d**) Inflow symmetrical model for three species, with $\tau_M = \tau_{QR}$ and $\theta_M = \theta_{QR}$ (see fig. 7**a**). Note that only model (**d**) matches the assumption of HyDe.



(**a**) Sister species (**b**) Outflow asym (**c**) Inflow asym (**d**) Inflow sym (HyDe model)

Figure 9: Power of detecting gene flow by HyDe and BPP in 100 replicate datasets simulated under the models of figure 8.

The data were also analyzed using BPP. The results are summarized in figure 9.

Model a (fig. 8**a**) assumes gene flow between sister lineages, based on the introgression event from QIRCD→U in the *Tamias* data (fig. 4**b**). It was suggested that by including multiple sequences from the recipient lineage, HyDe or the *D*-statistic might be used to detect gene flow between sister lineages. We used species R and U, with introgression rate $\varphi_{R\to U} = 0.625$, including two sequences (Ua and Ub) from the recipient species U, while S was used as the outgroup. The divergence times (τ) and population sizes (θ) were based on the real data (table S3). When multiple branches in the full tree (fig. 4b) were merged into one branch in the tree of figure 8**a**, θ for the merged branch was calculated as a weighted average, with the branch lengths as weights. As our objective in this case was to confirm the lack of power of HyDe (and the *D*-statistic), we simulated large datasets, each with $L = 8000$ loci. The sequence length was 500 sites, and the number of replicates was 100. When the data were analyzed using HyDe and the D-statistic, the quartet tree (((Ua, Ub), R), S) was used, with Ua or Ub labelled the 'hybrid' lineage. The same data were analyzed using BPP under the MSci model with three species (fig. 8**a**).

As expected, HyDe and the D-statistic had no power to detect gene flow between sister lineages: indeed, the power of HyDe and D was not higher than the significant level (fig. 9, table S4). Note that a test that ignores data and produces 5% positives at random will have 5% of power. Also HyDe did not produce reliable estimates of φ; in about half of the datasets, the estimate was outside the range $(0,1)$.

Model b (fig. 8**b**) was based on the next strongest introgression in the *Tamias* data, with $\varphi_{QI\to D} = 0.106$ (fig. 4**b**). We used species D, Q, R, with S as the outgroup. This is a case of outflow, when gene flow from an ingroup species Q to a more distant species D. Our examination of the assumptions made by HyDe suggests that HyDe can be used to detect inflow, but not outflow. We generated datasets of various sizes with $L = 500, 2000$ or 8000 loci. The other settings were the same as for model a. When the data were analyzed using HyDe, Q was designated the 'hybrid' lineage while R and D were the two parents. HyDe performed poorly (fig. 9**b**), with very low power and frequent invalid estimates of φ (table S5).

Model c (fig. 8**c**) was the same as model b but the direction of gene flow was reversed. The model was then a case of inflow, as assumed by HyDe. However, species divergence times and population sizes did not satisfy the symmetry requirements of HyDe (in other words, $\tau_M \neq \tau_{QR}$ and $\theta_M \neq \theta_{QR}$). In this case, HyDe had considerable power in detecting gene flow (fig. 9**c**). However, the estimates of φ by HyDe involved large biases, apparently converging to ≈ 0.32 when the true value was 0.106 (table S5). This positive bias is apparently because coalescent occurs at a higher rate or over longer time period on the M branch than on the QR branch in figure 8**c**, with $\frac{\tau_{QRD}-\tau_M}{\theta_M} > \frac{\tau_{QRD}-\tau_{QR}}{\theta_{QR}}$. In the opposite case, the bias should be negative.

Model d (fig. 8**d**) was the same as model c with inflow but in addition we enforced the symmetry assumptions, so that species Q was a hybrid species formed by hybridization between D and R. This is the hybrid speciation model assumed by HyDe, and the method performed well (fig. 9**d**). Its power was lower than that for BPP, as expected from statistical theory, but improved with the increase of data, rising from 10% at $L = 500$ loci to 90% at 8000 loci. The parameter estimate appeared to be consistent, converging to the correct value (0.106) when the number of loci increased, and there were not many invalid estimates (table S5). Those results are consistent with previous simulations, which evaluated the performance of HyDe when all its assumptions were met and found the method to perform well (Blischak *et al.*, 2018; Flouri *et al.*, 2020).

In summary, our simulations suggest that it is important to apply HyDe to detect the correct mode of gene flow (that is, gene flow between non-sister lineages, and inflow instead of outflow) (fig. 8d). Furthermore, the symmetry assumptions are important for HyDe to produce reliable estimates of introgression probability. When all model assumptions are met, HyDe performed well. However, HyDe had no power to detect gene flow between sister lineages, and very low power to detect outflow.

In all four models (fig. 8**a-d**), the Bayesian test using BPP had good power (fig. 9, tables S4&S5). Furthermore, the posterior means and 95% HPD CIs for parameters in the introgression models b-d were well-behaved (fig. 10). While HyDe can estimate only two parameters from the site-pattern counts (the internal branch length in coalescent units on the species tree and the introgression probability), the BPP analysis of the same data estimates all parameters in the model. The species divergence/introgression times were all well estimated with small CIs (fig. 10). The introgression probability was accurately estimated with narrow CIs when ≥ 500 loci were used. Population size parameters for short branches were poorly estimated due to lack of coalescent events in those populations.

Table 3. False positive rate of BPP and HyDe tests and average estimates of introgression probability in 100 simulated replicates

	BPP			HyDe			
# loci	Error Rate ($\alpha = 1\%$)	Error Rate ($\alpha = 5\%$)	$\hat{\varphi}\pm$SD	Error Rate ($\alpha = 1\%$)	Error Rate ($\alpha = 5\%$)	$\hat{\varphi}\pm$SD	Proportion of invalid estimates
Inflow asym (fig. 8**c**)							
500	0%	0%	0.019 ± 0.011	1%	7%	0.140 ± 0.108	52%
2000	0%	0%	0.009 ± 0.004	5%	13%	0.094 ± 0.061	52%
8000	0%	0%	0.004 ± 0.002	2%	7%	0.038 ± 0.032	51%
Inflow sym (fig. 8**d**, HyDe model)							
500	0%	0%	0.032 ± 0.016	0%	3%	0.064 ± 0.048	49%
2000	0%	0%	0.014 ± 0.006	1%	2%	0.039 ± 0.029	55%
8000	0%	0%	0.006 ± 0.003	0%	3%	0.022 ± 0.016	49%

Note.— Data were simulated using the species trees of figure 8**c-d** but with $\varphi = 0$.

Table 4. LRT and Bayesian tests in the normal example in two datasets

Data	LRT	Bayesian test		
$\sqrt{n}\lvert\bar{x}\rvert$	p-value	Prior	B_{10}	$\mathbb{P}(H_1\lvert x)$
1.96	$p=0.05$	$\sigma_0=1$	0.359	0.264
1.96	$p=0.05$	$\sigma_0=2$	0.262	0.208
1.96	$p=0.05$	$\sigma_0=10$	0.120	0.107
2.58	$p=0.01$	$\sigma_0=1$	0.408	0.290
2.58	$p=0.01$	$\sigma_0=2$	0.300	0.230
2.58	$p=0.01$	$\sigma_0=10$	0.138	0.122

Note.— The Bayes factor B_{10} is calculated assuming data size $n=100$ in eq. 13, while the posterior model probability is given by eq. 14. Note that the p-value for the LRT is 5% (or 1%) in the dataset with $\sqrt{n}|\bar{x}|=1.96$ (or 2.58).

We also examined the false positive rate (type-I error rate) of the HYDE and Bayesian tests, by simulating data using the inflow-asym (fig. 8**c**) and inflow-sym (fig. 8**d**) models but with $\varphi=0$ fixed so that there was no introgression in the true model. The results are summarized in table 3. Under the inflow-asym model, HYDE had higher false positive rate than the nominal significant level. For example, at the 5% significance level, the false positive rate was 7%, 13%, and 7% in datasets of 500, 2000, and 8000 loci, respectively. The high rate may be explained by the violation of the symmetry assumptions for HYDE. Under the inflow-sym model (or the HYDE model), the rate was 3%, 2%, and 3%, all within the allowed 5% (table 3). Thus HYDE performed well when its assumptions were met and had elevated false positives when the assumptions were violated. In all settings, the false positive rate of the Bayesian test was estimated to be ~ 0%. This is consistent with the expectation that the Bayesian test may be more conservative (with lower false positive rate and lower power) than the LRT (see discussions later).

Finally, to assess the information content in datasets of the size of the *Tamias* data, we used parameter estimates from the full dataset (fig. 4**b**, table S3) to simulate two datasets of the same size as the original, with 5, 5, 9, 10, 11, 11, 3 unphased sequences per locus for species R, C, I, U, Q, D and S, respectively. The sequence length was 200 sites. We analyzed the datasets under the same MSci model of figure 4**b** using BPP to estimate all parameters. The estimates from the two datasets were similar, so we present those from one of them in table S3. At this data size, BPP achieved relatively good precision and accuracy. The posterior means were close to the true values, and the CIs were also similar to those calculated from the real data. Similarly to analyses of the real data, divergence times and population sizes for modern species were well estimated, but ancestral population sizes, in particular those for populations of short time duration, were more poorly estimated.

Discussion

Criteria for testing gene flow

Hypothesis testing or model selection involves arbitrariness, and classical hypothesis testing and Bayesian model selection applied to the same data may produce strongly opposed conclusions, a situation known as Jeffreys's paradox (Jeffreys, 1939; Lindley, 1957). Furthermore, Bayesian model selection is known to be sensitive to priors on model parameters, especially on parameters that are not shared between the models under comparison. See Yang (2014, pp.194-7) for a discussion of those issues. Here we review different strategies for testing, using as example a simple problem of testing the null hypothesis $H_0:\mu=0$ against the alternative $H_1:\mu\neq 0$, using a data sample, $x=\{x_1,x_2,\cdots,x_n\}$, from the normal distribution $\mathbb{N}(\mu,1)$. We assume that a false positive error (of falsely rejecting H_0 when it is true) is more serious than a false negative error (of failing to reject H_0 when it is false). The data can be summarized as the sample mean $\bar{x}$, with the likelihood given by $\bar{x}\sim\mathbb{N}(0,1/n)$ under H_0 and $\bar{x}\sim\mathbb{N}(\mu,1/n)$ under H_1. Let $\phi(x;\mu,\sigma^2)$ be the probability density function (PDF) for $\mathbb{N}(\mu,\sigma^2)$ and $\Phi(\cdot)$ be the CDF for $\mathbb{N}(0,1)$.

In hypothesis testing, the p-value can be calculated from the fact that under H_0, $\sqrt{n}|\bar{x}|\sim\mathbb{N}(0,1)$ or $n|\bar{x}|^2\sim\chi^2_1$. At the $\alpha=5\%$ significance level, we reject H_0 if

$$2\Delta\ell = 2\log\frac{\phi(\bar{x};\bar{x},\frac{1}{n})}{\phi(\bar{x};0,\frac{1}{n})} = n|\bar{x}|^2 > \chi^2_{1,5\%} = 3.84. \tag{11}$$

Alternatively one may consider this as an estimation problem and construct a confidence interval (CI) for μ and

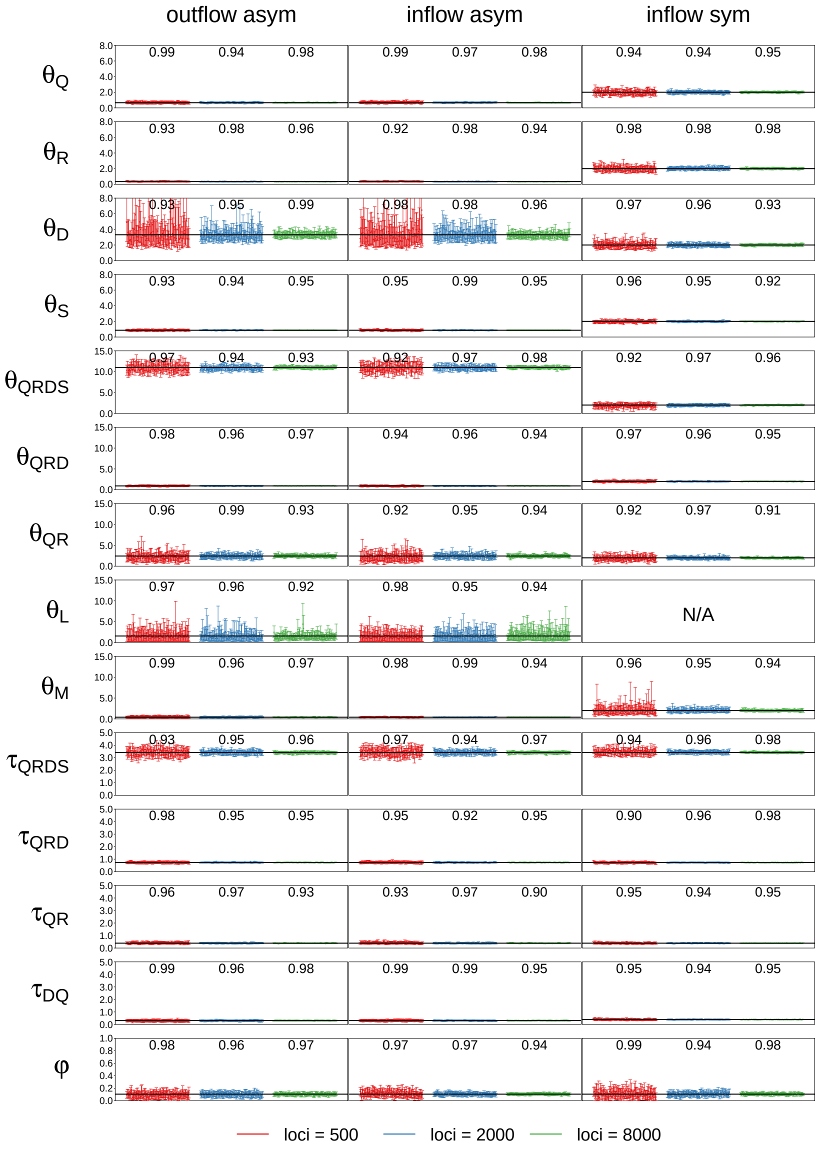


Figure 10: Posterior means and 95% HPD CIs for parameters in the three introgression models of figure 8: (**b**) outflow asym, (**c**) inflow asym and (**d**) inflow sym (HyDe model), in BPP analyses of 100 replicate datasets, each with 500, 2000, or 8000 loci. Note that in model (d) inflow sym, all populations had the same size (θ) although separate θ parameters were estimated for different populations when the data were analyzed using BPP. Parameters τ and θ are multiplied by 10^3. The number above the CI bars is the coverage or the probability that the CI includes the true value.

reject H_0 if the CI excludes the null value 0. This is equivalent to the LRT.

In a Bayesian analysis, we consider two approaches. The first is to examine whether the posterior 95% credibility interval (CI) for μ under H_1 excludes the null value 0. We assign the prior $\mu \sim \mathbb{N}(0, \sigma_0^2)$ under H_1. The posterior is then $\mu|x \sim \mathbb{N}(\mu_1, \sigma_1^2)$, with mean $\mu_1 = \frac{n\bar{x}}{n+1/\sigma_0^2}$ and precision $\frac{1}{\sigma_1^2} = n + \frac{1}{\sigma_0^2}$. Here the reciprocal of variance is known as precision. The sample precision is n and the prior precision is $1/\sigma_0^2$, while the posterior precision is the sum of the two. The 95% CI for μ is given as $\mu_1 \pm 1.96\sigma_1$ so that the CI excludes 0 (in which case we reject H_0) if $|\mu_1| > 1.96\sigma_1$, or if

$$n|\bar{x}|^2 > 3.84\left[1 + 1/(n\sigma_0^2)\right]. \tag{12}$$

The second approach is to use the Bayes factor to compare the null and alternative hypotheses.

$$
\begin{aligned}
B_{10} = \frac{\mathbb{P}(\bar{x}|H_1)}{\mathbb{P}(\bar{x}|H_0|} &= \frac{\phi(\bar{x}; 0, \frac{1}{n} + \sigma_0^2)}{\phi(\bar{x}; 0, \frac{1}{n})} \\
&= \frac{1}{\sqrt{1 + n\sigma_0^2}} \cdot \exp\left\{ \frac{n\bar{x}^2}{2\left[1 + 1/(n\sigma_0^2)\right]} \right\},
\end{aligned} \tag{13}
$$

(e.g., Yang, 2006, eq. 5.21).

The Bayes factor is closely related to (and 'calibrated' using) the posterior model probability. If the two models are assigned equal prior probabilities ($\pi_0 = \pi_1 = \frac{1}{2}$), the posterior model probability is

$$
\mathbb{P}(H_1|x) = \frac{B_{10}}{1 + B_{10}}, \tag{14}
$$

so that a 95% cut-off on $\mathbb{P}(H_1|x)$ corresponds to $B_{10} = 19$, and H_0 is rejected based on the Bayes factor if and only if

$$
n|\bar{x}|^2 > \log\left\{19\sqrt{1 + n\sigma_0^2}\right\} \times 2\left[1 + 1/(n\sigma_0^2)\right]. \tag{15}
$$

While the LRT (eq. 11) depends on $\sqrt{n}|\bar{x}|$ only, both the posterior CI (eq. 12) and the Bayes factor (eq. 15) depend in addition on $n\sigma_0^2$. Note that the three criteria (eqs. 11, 12, & 15) have the ordering

$$
\begin{aligned}
3.84 &< 3.84\left[1 + 1/(n\sigma_0^2)\right] \\
&< \log\left\{19\sqrt{1 + n\sigma_0^2}\right\} \times 2\left[1 + 1/(n\sigma_0^2)\right].
\end{aligned} \tag{16}
$$

Thus the LRT has more power and higher false positive rate than the posterior CI while the Bayesian test based on the Bayes factor is the most conservative. The result reflects the general perception that the LRT tends to reject the null hypothesis and favour parameter-rich models too often, especially in large datasets. Note that if H_0 is true, the false positive rate of the LRT stays at 5% when the sample size $n \to \infty$, whereas in the Bayesian analysis, the true model H_0 will dominate, with $\mathbb{P}(H_0|x) \to 1$ and $B_{10} \to 0$ when $n \to \infty$.

Example calculations are given in table 4 for two datasets with $\sqrt{n}|\bar{x}| = 1.96$ or 2.58 and $n = 100$. In both datasets, H_0 is rejected by the LRT (at the 5% and 1% levels, respectively), but the Bayes factor and the posterior model probabilities favour H_0 over H_1, with $B_{10} < 1$ and $\mathbb{P}(H_1|x) < \frac{1}{2}$.

This analysis suggests that the difference in power between HYDE and BPP are due to the inefficient use of information in the data by HYDE, not to the different statistical philosophies. An LRT for testing introgression applied to the multilocus sequence alignments may be expected to have more power (and higher false positive rate) than the Bayesian test based on the Bayes factor.

The power of heuristic and likelihood methods to detect introgression

When applied to the *Tamias* dataset, HYDE and BPP produced opposite conclusions concerning gene flow. Our examination of the model assumptions for HYDE and our simulations suggest that this is because gene flow with the strongest signal in the *Tamias* group, either between sister species or involving outflow, may be of the wrong type or in the wrong direction for HYDE. Here we review and summarize the major issues with HYDE.

First, both HYDE and the *D*-statistic pool sites across loci when counting site patterns, so that the site-pattern counts are genome-wide averages. Cross-species gene flow creates genealogical variation across the genome, with the probabilistic distribution of the gene trees and coalescent times specified by parameters in the MSC model with gene flow, such as species divergence times, population sizes, and rates of gene flow (Barton, 2006; Lohse and Frantz, 2014). As a result, there is important information concerning gene flow in the variance of site-pattern counts among loci, but this information is ignored by those methods. In other words, sites at the same locus share the genealogical history under the assumption of no within-locus recombination (see Zhu *et al.*, 2022 for an evaluation of the impact of this assumption on MSC-based analyses), and their differences reflect the stochastic fluctuation of the mutation process. Sites at different loci in addition may have different genealogical histories, reflecting the stochastic nature of the process of coalescent and introgression. When sites are pooled across loci, those two sources of variation are confounded, leading to loss of information (Shi and Yang, 2018; Zhu and Yang, 2021). As a consequence, certain forms of introgression, such as introgression between sister lineages, are unidentifiable by *D* or HYDE, while estimation of introgression rates between non-sister species suffers from larger variances (Jiao *et al.*, 2021).

Second, HYDE makes restrictive assumptions about gene flow. The underlying model is one of hybrid speciation with identical population sizes or equivalently the inflow model with symmetrical species divergence times and population sizes (fig. 7**a**, with $\tau_S = \tau_T$ and $\theta_S = \theta_T$) (Blischak *et al.*, 2018; Kubatko and Chifman, 2019). Our simulation suggests that HYDE can indeed infer gene flow/hybridization and produce reliable estimates of introgression probability under this model (fig. 9**d** & table S5; see also Blischak *et al.*, 2018; Flouri *et al.*, 2020). However, introgression in the wrong direction or violation of the symmetry assumptions may lead to loss of power and biased or invalid estimates by HYDE (fig. 9**b&c**, table S5).

Third, the approaches taken by HYDE to accommodate multiple samples per species and heterozygote sites in diploid genomes may be problematic. When multiple samples are available in the species quartet, HYDE counts site patterns in all combinations of the quartet. Let the numbers of sequences for species O, P_1, H, P_2 be n_O, n_1, n_H, n_2. There are then $n_O \times n_1 \times n_H \times n_2$ combinations in which one sequence is sampled per species, and HYDE counts site patterns in all of them (Blischak *et al.*, 2018). This ignores the lack of independence among the quartets and exaggerates the sample size. At the same time, multiple samples from the same species are never compared with each other, which should provide important information about the population size for that species. In a likelihood method such as BPP, all sequences at the same locus, both from the same species and from different species, are related through a gene tree, and genealogical information at the locus is used.

Similarly heterozygote sites are not treated properly in HYDE. If the site pattern is AGRG, with R representing an A/G heterozygote, HYDE adds 0.5 each to the site patterns $ijjj$ (for AGGG) and $ijij$ (for AGAG) (Blischak *et al.*, 2018), in effect treating R as an unknown nucleotide that is either A or G whereas correctly it means a heterozygote (both A and G). The proportion of heterozygotes in each diploid genome should be informative about θ for that population, but such information is not used by HYDE. In BPP, heterozygote sites are resolved into their underlying nucleotides using an analytical integration algorithm (so that R means both A and G, say), with the uncertainty in the genotypic phase of multiple heterozygous sites in a diploid sequence accommodated by averaging over all possible heterozygote phase resolutions, weighting them according to their likelihoods based on the sequence alignment at the locus (Gronau *et al.*, 2011; Flouri *et al.*, 2018). Simulations suggest that this approach has nearly identical statistical performance to using fully phased haploid genomic sequences (Gronau *et al.*, 2011; Huang *et al.*, 2021).

In this paper we have focused on the heuristic method HYDE and the likelihood method BPP, as they have been used to analyze the *Tamias* data. By choosing parameter values to be representative of the *Tamias* data, our simulation has evaluated a tiny portion of the parameter space and does not constitute a systematic evaluation of the performance of HYDE. The strengths and weaknesses of heuristic and likelihood methods for inference under models of gene flow were discussed by Degnan (2018) and Jiao *et al.* (2021), but a comprehensive comparative study has not yet been conducted. For estimation of the species phylogeny under the MSC without gene flow, (Zhu and Yang, 2021, fig. 3) demonstrated a dramatic information loss resulting from pooling sites across loci in the site-pattern based methods (also known as coalescent-aware concatenation methods), and from the failure to use information in coalescent times or gene-tree branch lengths in the two-step methods (which infer the gene trees and then treat them as data to infer the species tree). Both the site pattern-based and the two-step methods are used to infer gene flow and to estimate the introgression probability (e.g., HYDE and the D-statistic in the first category and SNAQ in the second) and similar information loss may be expected. A detailed analysis of the performance of heuristic methods in comparison with likelihood methods will be interesting. Currently the gap between the heuristic and likelihood methods appears to be a large one. Heuristic methods are orders-of-magnitude more efficient computationally and can be applied to much larger datasets, whereas likelihood methods have far better statistical properties, being able to identify and estimate all parameters in the model. There are great opportunities for improving both the statistical performance of heuristic methods and the computational efficiency of likelihood methods (including the mixing efficiency of MCMC algorithms).

Introgression in T. quadrivittatus chipmunks

The joint introgression model for the *T. quadrivittatus* group (fig. 4**b**) was constructed using a stepwise approach that iteratively adds introgression events to the binary species tree. We note several limitations with this approach. First the approach assumes the availability of a stable binary species tree, and may not be feasible if the species tree is large and highly uncertain, possibly influenced by introgression events (Leaché *et al.*, 2014). The *Tamias* dataset analyzed here includes only six species, and the first stage of our procedure (i.e., the separate analysis) involved 16 possible introgression events, so that the computation was feasible. Second, the approach is not an exhaustive search in the space of introgression models and may miss certain introgression events. Note that introgression events not selected in the first stage of the procedure will not be incorporated in the final joint introgression model. In our analysis of the *Tamias* data, we considered introgressions between contemporary species, mostly based on phylogenetic analyses of the mitochondrial genome (Sarver *et al.*, 2017), and moved certain events to older ancestral branches when the estimated introgression time coincided with the species divergence time. We did not evaluate introgressions involving ancestral branches systematically. Furthermore, the criterion based on the Bayes factor used in our test is a stringent one, and the dataset of 1060 loci is relatively small. All those factors suggest that we cannot rule out the possibility that we may have missed some introgression events; in other words, our analysis may suffer from false-negative errors. In contrast, the three introgression events identified in our analysis (fig. 4**b**) appear to be robust and are unlikely to be false positives (figs. 5, table S2). We conclude that there is strong and robust evidence that gene flow has affected the nuclear genome in the *T. quadrivittatus* group of chipmunks.

Given the extensive mitochondrial introgression in the *Tamias* group (Sullivan *et al.*, 2014; Sarver *et al.*, 2017, 2021), introgression affecting the nuclear genome was expected, and the failure to detect any significant evidence for it in the HYDE analysis was surprising (Sarver *et al.*, 2021). Sarver *et al.* (2021) discussed the evidence for cytonuclear discordance in the pattern of introgression (Bonnet *et al.*, 2017; McElroy *et al.*, 2020; Sarver *et al.*, 2021), as well as possible roles of purifying selection affecting the coding genes or exons that make up the nuclear dataset being analyzed. Our results suggest a simpler explanation, that gene flow in the *Tamias* group is of a wrong type or in the wrong direction, undetectable by HYDE.

Our analyses suggest that species involved in excessive mitochondrial introgression tend to be those involved in nuclear introgression as well. *T. dorsalis* was noted to be a universal recipient of mtDNA from other species (Sullivan *et al.*, 2014; Sarver *et al.*, 2017). Consistent with this, our separate analysis (table 2) identified three introgression events into *T. dorsalis* with $\varphi > 5\%$ as well as one event with *T. dorsalis* to be the donor species, even though some of those events become non-significant after introgression involving older ancestors was incorporated in the model. It will be interesting to use expanded datasets to examine whether this is due to a lack of power to detect gene flow or a genuine lack of gene flow.

It will be very useful to generate more genomic data, especially the noncoding parts of the nuclear genome, including more species from the genus, to provide more power for detecting gene flow and estimating introgression rates. It will also be interesting to examine whether the noncoding and coding regions of the genome give consistent signals concerning species divergences and cross-species gene flow, and to examine how the effective rate of gene flow vary among chromosomes or across genomic regions. In a few genomic analyses, coding and noncoding parts of the genome were found to produce highly consistent results, with nearly proportional estimates of divergence times (τ) and population sizes (θ), and with very similar estimates of introgression rates (Shi and Yang, 2018; Thawornwattana *et al.*, 2018, 2022). One can also examine the posterior distribution of the gene trees to identify loci or genomic segments that are most likely to have been transferred across species boundaries, and to correlate with the functions of genes residing in or tightly linked to the segments.

Supplementary Material

Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.fxpnvx0t9.

Acknowledgements

We thank three reviewers (including Dr Bryan Carstens), the Associate Editor (Dr Robert Thompson), and the Editor (Dr Isabel Sanmartín) for many constructive comments.

Funding

This work was supported by Biotechnology and Biological Sciences Research Council grants (BB/T003502/1 and BB/R01356X/1) to Z.Y. and an NSF grant (NSF-SBS-2023723) to A.D.L.

References

Arnold, B. J., Lahner, B., DaCosta, J. M., Weisman, C. M., Hollister, J. D., Salt, D. E., Bomblies, K., and Yant, L. 2016. Borrowed alleles and convergence in serpentine adaptation. *Proc. Natl. Acad. Sci. USA*, 113(29): 8320–8325.

Barton, N. H. 2006. Evolutionary biology: how did the human species form? *Curr. Biol.*, 16: R647–R650.

Bi, K., Linderoth, T., Singhal, S., Vanderpool, D., Patton, J. L., Nielsen, R., Moritz, C., and Good, J. M. 2019. Temporal genomic contrasts reveal rapid evolutionary responses in an alpine mammal during recent climate change. *PLoS Genet.*, 15(5): e1008119.

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.

Bonnet, T., Leblois, R., Rousset, F., and Crochet, P.-A. 2017. A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. *Evolution*, 71(9): 2140–2158.

Brown, J. H. 1971. Mechanisms of competitive exclusion between two species of chipmunks. *Ecology*, 52(2): 305–311.

Burgess, R. and Yang, Z. 2008. Estimation of hominoid ancestral population sizes under Bayesian coalescent models incorporating mutation rate variation and sequencing errors. *Mol. Biol. Evol.*, 25: 1979–1994.

Chan, Y. C., Roos, C., Inoue-Murayama, M., Inoue, E., Shih, C. C., Pei, K. J., and Vigilant, L. 2013. Inferring the evolutionary histories of divergences in *Hylobates* and *Nomascus* gibbons through multilocus sequence data. *BMC Evol. Biol.*, 13: 82.

Chifman, J. and Kubatko, L. 2014. Quartet inference from snp data under the coalescent model. *Bioinformatics*, 30(23): 3317–3324.

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.

Dalquest, W. W., Baskin, J., and Schultz, G. 1996. Fossil mammals from a late miocene (clarendonian) site in beaver county, oklahoma. *Contributions in Mammalogy: A Memorial Volume Honoring Dr. J. Knox Jones, Jr. Museum of Texas Tech University*, pages 107–137.

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

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

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backstrom, N., Kawakami, T., Kunstner, A., Makinen, H., Nadachowska-Brzyska, K., Qvarnstrom, A., Uebbing, S., and Wolf, J. B. W. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, 491: 756–760.

Finger, N., Farleigh, K., Bracken, J., Leaché, A. D., Francois, O., Yang, Z., Flouri, T., Charran, T., Jezkova, T., Williams, D., and Blair, C. 2022. Genome-scale data reveal deep lineage divergence and a complex demographic history in the texas horned lizard (*Phrynosoma cornutum*) throughout the southwestern and central USA. *Genome Biol. Evol.*, 14(1): 10.1093/gbe/evab260.

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

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

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

Good, J. M. and Sullivan, J. 2001. Phylogeography of the red-tailed chipmunk (*Tamias ruficaudus*), a northern Rocky Mountain endemic. *Mol. Ecol.*, 10(11): 2683–2695.

Good, J. M., Demboski, J. R., Nagorsen, D. W., and Sullivan, J. 2003. Phylogeography and introgressive hybridization: chipmunks (genus *Tamias*) in the northern Rocky Mountains. *Evolution*, 57(8): 1900–1916.

Good, J. M., Hird, S., Reid, N., Demboski, J. R., Steppan, S. J., Martin-Nims, T. R., and Sullivan, J. 2008. Ancient hybridization and mitochondrial capture between two species of chipmunks. *Mol. Ecol.*, 17(5): 1313–1327.

Green, P. 1995. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82: 711–732.

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

Heled, J. and Drummond, A. J. 2010. Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol.*, 27: 570–580.

Heller, H. C. 1971. Altitudinal zonation of chipmunks (*Eutamias*): interspecific aggression. *Ecology*, 52(2): 312–319.

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.

Hird, S., Reid, N., Demboski, J., and Sullivan, J. 2010. Introgression at differentially aged hybrid zones in red-tailed chipmunks. *Genetica*, 138(8): 869–883.

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., Bennett, J., Flouri, T., and Yang, Z. 2021. Phase resolution of heterozygous sites in diploid genomes is important to phylogenomic analysis under the multispecies coalescent model. *Syst. Biol.*

Jeffreys, H. 1939. *Theory of Probability*. Clarendon Press, Oxford, England.

Jiao, X., Flouri, T., and Yang, Z. 2021. Multispecies coalescent and its applications to infer species phylogenies and cross-species gene flow. *Nat. Sci. Rev.*, 8(12): DOI: 10.1093/nsr/nwab127.

Kubatko, L. S. and Chifman, J. 2019. An invariants-based method for efficient identification of hybrid species from large-scale genomic data. *BMC Evol. Biol.*, 19(1): 112.

Kumar, V., Lammers, F., Bidon, T., Pfenninger, M., Kolter, L., Nilsson, M. A., and Janke, A. 2017. The evolutionary history of bears is characterized by gene flow across species. *Sci. Rep.*, 7: 46487.

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

Leaché, A. D., Harris, R. B., Rannala, B., and Yang, Z. 2014. The influence of gene flow on species tree estimation: a simulation study. *Syst. Biol.*, 63(1): 17–30.

Lindley, D. 1957. A statistical paradox. *Biometrika*, 44: 187–192.

Lohse, K. and Frantz, L. A. 2014. Neandertal admixture in Eurasia confirmed by maximum-likelihood analysis of three genomes. *Genetics*, 196(4): 1241–1251.

Mallet, J., Besansky, N., and Hahn, M. W. 2016. How reticulated are species? *BioEssays*, 38(2): 140–149.

Mao, Y., Economo, E. P., and Satoh, N. 2018. The roles of introgression and climate change in the rise to dominance of *Acropora* corals. *Curr. Biol.*, 28(21): 3373–3382 e5.

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

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

McElroy, K., Black, A., Dolman, G., Horton, P., Pedler, L., Campbell, C. D., Drew, A., and Joseph, L. 2020. Robbery in progress: Historical museum collections bring to light a mitochondrial capture within a bird species widespread across southern Australia, the copperback quail-thrush *Cinclosoma clarum*. *Ecol. Evol.*, 10(13): 6785–6793.

Mirarab, S. and Warnow, T. 2015. Astral-ii: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*, 31(12): i44–i52.

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

Ogilvie, H. A., Bouckaert, R. R., and Drummond, A. J. 2017. StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. *Mol. Biol. Evol.*, 34(8): 2101–2114.

Patterson, B. D. and Norris, R. W. 2016. Towards a uniform nomenclature for ground squirrels: the status of the Holarctic chipmunks. *Mammalia*, 80(3): 241–251.

Patterson, B. D. and Thaeler Jr, C. S. 1982. The mammalian baculum: hypotheses on the nature of bacular variability. *J. Mammal.*, 63(1): 1–15.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. 2012. Ancient admixture in human history. *Genetics*, 192(3): 1065–1093.

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

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

Reid, N., Demboski, J. R., and Sullivan, J. 2012. Phylogeny estimation of the radiation of western north american chipmunks (*Tamias*) in the face of introgression using reproductive protein genes. *Syst. Biol.*, 61(1): 44.

Root, J. J., Calisher, C. H., and Beaty, B. J. 2001. Microhabitat partitioning by two chipmunk species (*Tamias*) in western Colorado. *Western North American Naturalist*, pages 114–118.

Sarver, B. A., Demboski, J. R., Good, J. M., Forshee, N., Hunter, S. S., and Sullivan, J. 2017. Comparative phylogenomic assessment of mitochondrial introgression among several species of chipmunks (*Tamias*). *Genome Biol. Evol.*, 9(1): 7–19.

Sarver, B. A. J., Herrera, N. D., Sneddon, D., Hunter, S. S., Settles, M. L., Kronenberg, Z., Demboski, J. R., Good, J. M., and Sullivan, J. 2021. Diversification, introgression, and rampant cytonuclear discordance in Rocky Mountains chipmunks (Sciuridae: *Tamias*). *Syst. Biol.*, 70(5): 908–921.

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.

Shi, C. and Yang, Z. 2018. Coalescent-based analyses of genomic sequence data provide a robust resolution of phylogenetic relationships among major groups of gibbons. *Mol. Biol. Evol.*, 35: 159–179.

Silverman, B. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

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

Sullivan, J., Demboski, J., Bell, K., Hird, S., Sarver, B., Reid, N., and Good, J. 2014. Divergence with gene flow within the recent chipmunk radiation (*Tamias*). *Heredity*, 113(3): 185–194.

Swofford, D. L. 2003. *PAUP*: Phylogenetic Analysis by Parsimony (*and Other Methods), Version 4*. Sinauer Associates, Sanderland, Massachusetts.

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

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

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.

White, J. A. 2010. *The Baculum in the Chipmunks of Western North America*. Good Press.

Xu, B. and Yang, Z. 2016. Challenges in species tree estimation under the multispecies coalescent model. *Genetics*, 204(4): 1353–1368.

Yang, Z. 2006. *Computational Molecular Evolution*. Oxford University Press, Oxford, UK.

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

Yang, Z. 2015. The BPP program for species tree estimation and species delimitation. *Curr. Zool.*, 61(5): 854–865.

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

Zhang, C., Ogilvie, H. A., Drummond, A. J., and Stadler, T. 2018. Bayesian inference of species networks from multilocus sequence data. *Mol. Biol. Evol.*, 35(2): 504–517.

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.

Zhu, T. and Yang, Z. 2021. Complexity of the simplest species tree problem. *Mol. Biol. Evol.*, 39: 3993–4009.

Zhu, T., Flouri, T., and Yang, Z. 2022. A simulation study to examine the impact of recombination on phylogenomic inferences under the multispecies coalescent model. *Mol. Ecol.*, 31: 2814–2829.

Supplemental Information for
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

Jiayi Ji, Donavan J. Jackson, Adam D. Leaché, and Ziheng Yang

Table S1. Posterior means and 95% HPD CIs (in parentheses) of introgression probabilities (φ) and introgression times (τ) in the stepwise construction of the MSci model, applied to datasets of the two halves

		First half			Second half		
	Model	φ	τ ($\times 10^{-3}$)	B_{10}	φ	τ ($\times 10^{-3}$)	B_{10}
1	QIRCD → U	0.537 (0.247, 0.801)	0.841 (0.702, 0.980)	∞	0.632 (0.332, 0.869)	0.906 (0.740, 1.045)	∞
2	QIRCD → U	0.615 (0.427, 0.798)	0.869 (0.773, 0.966)	∞	0.695 (0.501, 0.876)	0.918 (0.802, 1.037)	∞
	QI ↔ D	0.138 (0.094, 0.185)	0.349 (0.311, 0.391)	∞	0.069 (0.038, 0.102)	0.322 (0.282, 0.363)	∞
		0.020 (0.000, 0.047)		0.03	0.018 (0.000, 0.037)		0.03
3	QIRCD → U	0.601 (0.369, 0.813)	0.863 (0.759, 0.971)	∞	0.696 (0.480, 0.892)	0.915 (0.782, 1.040)	∞
	QI → D	0.133 (0.080, 0.191)	0.331 (0.288, 0.372)	53.24	0.085 (0.038, 0.136)	0.331 (0.279, 0.388)	25.06
	Q → D	0.008 (0.000, 0.021)	0.153 (0.058, 0.261)	0.00	0.008 (0.001, 0.019)	0.100 (0.040, 0.210)	0.00
4	QIRCD → U	0.588 (0.360, 0.797)	0.854 (0.746, 0.958)	∞	0.681 (0.470, 0.869)	0.909 (0.787, 1.032)	∞
	QI → D	0.132 (0.080, 0.188)	0.330 (0.286, 0.373)	∞	0.088 (0.035, 0.149)	0.334 (0.270, 0.396)	35.99
	I → D	0.007 (0.000, 0.018)	0.120 (0.048, 0.197)	0.00	0.012 (0.001, 0.025)	0.160 (0.090, 0.227)	0.01
5	QIRCD → U	0.582 (0.326, 0.818)	0.852 (0.738, 0.961)	∞	0.689 (0.473, 0.883)	0.905 (0.778, 1.026)	∞
	QI → D	0.126 (0.084, 0.170)	0.307 (0.263, 0.352)	∞	0.099 (0.063, 0.139)	0.336 (0.291, 0.382)	∞
	Q ↔ I	0.036 (0.013, 0.065)	0.099 (0.062, 0.137)	2.90	0.048 (0.023, 0.075)	0.088 (0.051, 0.118)	∞
		0.030 (0.008, 0.055)		0.39	0.014 (0.000, 0.028)		0.02
6	QIRCD → U	0.589 (0.338, 0.827)	0.850 (0.738, 0.966)	∞	0.686 (0.491, 0.870)	0.902 (0.782, 1.022)	∞
	QI → D	0.118 (0.074, 0.165)	0.295 (0.246, 0.345)	∞	0.097 (0.060, 0.136)	0.334 (0.284, 0.381)	∞
	Q → I	0.041 (0.014, 0.074)	0.100 (0.066, 0.140)	7.90	0.055 (0.026, 0.087)	0.091 (0.053, 0.132)	∞

Note.— Introgression events are added sequentially onto the species tree of figure 4**a** and those that do not meet our cutoffs ($B_{10} \geq 20$) are grayed out. $B_{10} = \infty$ occurs when there are no MCMC samples with $\varphi < \varepsilon = 1\%$. A bidirectional introgression event, e.g., between Q and I has two introgression probabilities, e.g., $\varphi_{Q \to I}$ (above) and $\varphi_{I \to Q}$ (below). The final joint introgression model has three unidirectional introgression events.

Table S2. Bayes factors (B_{10}) for the three introgression probabilities (φ) obtained from BPP analyses of the full data of 1060 loci under the joint MSci model of figure 4b and different beta priors, φ ∼beta(α, β)

		B_{10}		
ε & Prior	$\mathbb{P}(\emptyset)$	QIRCD → U	QI → D	Q → I
$\varepsilon = 1\%$				
beta(0.2, 0.2)	0.210	∞	∞	∞
beta(0.2, 1)	0.398	∞	∞	∞
beta(0.2, 5)	0.585	∞	∞	∞
beta(1, 0.2)	0.002	∞	∞	∞
beta(1, 1)	0.010	∞	∞	∞
beta(1, 5)	0.049	∞	∞	∞
beta(5, 0.2)	6.0×10^{-12}	∞	∞	∞
beta(5, 1)	1.0×10^{-10}	∞	∞	∞
beta(5, 5)	1.2×10^{-8}	∞	∞	∞
$\varepsilon = 0.1\%$				
beta(0.2, 0.2)	0.132	∞	∞	∞
beta(0.2, 1)	0.251	∞	∞	∞
beta(0.2, 5)	0.371	∞	∞	∞
beta(1, 0.2)	2.0×10^{-4}	∞	∞	∞
beta(1, 1)	0.001	∞	∞	∞
beta(1, 5)	0.005	∞	∞	∞
beta(5, 0.2)	6.0×10^{-17}	∞	∞	∞
beta(5, 1)	1.0×10^{-15}	∞	∞	∞
beta(5, 5)	1.3×10^{-13}	∞	∞	∞

Note.— Bayes factor B_{10} is calculated using eq. 8, where the null region $\emptyset$ for φ is the interval $(0, \varepsilon)$ with $\varepsilon = 1\%$ or 0.1%. $B_{10} = \infty$ occurs when $\varphi > \varepsilon$ in all MCMC samples.

Table S3. Posterior means and 95% HPD CIs (in parentheses) of parameters under the MSci model of figure 4b obtained from BPP analyses of three real datasets (the two halves and the full dataset) and a simulated dataset

Parameters	First half, 530 loci	Second half, 530 loci	Full data, 1060 loci	Simulation, 1060 loci
Population sizes (θ, $\times 10^{-3}$)				
θ_Q	1.119 (0.817, 1.455)	1.032 (0.733, 1.370)	1.059 (0.844, 1.287)	1.098 (0.867, 1.347)
θ_I	1.556 (0.996, 2.191)	2.335 (1.474, 3.387)	1.923 (1.433, 2.473)	2.108 (1.574, 2.701)
θ_R	0.330 (0.266, 0.399)	0.377 (0.311, 0.442)	0.344 (0.295, 0.396)	0.366 (0.313, 0.421)
θ_C	0.478 (0.400, 0.556)	0.474 (0.407, 0.547)	0.478 (0.427, 0.534)	0.491 (0.436, 0.542)
θ_D	3.092 (2.580, 3.633)	3.460 (2.920, 4.016)	3.314 (2.915, 3.705)	3.386 (3.001, 3.781)
θ_U	0.953 (0.843, 1.063)	0.912 (0.814, 1.011)	0.932 (0.857, 1.004)	0.917 (0.844, 0.991)
θ_S	0.792 (0.680, 0.904)	0.934 (0.809, 1.052)	0.866 (0.782, 0.948)	0.817 (0.734, 0.900)
$\theta_{QIRCDUS}$	11.04 (9.516, 12.56)	10.83 (9.357, 12.30)	11.01 (9.924, 12.09)	10.38 (9.368, 11.40)
θ_{QIRCDU}	0.687 (0.332, 1.075)	0.601 (0.274, 0.955)	0.656 (0.367, 0.971)	0.475 (0.229, 0.734)
θ_{QIRCD}	1.963 (0.248, 4.652)	1.595 (0.336, 3.153)	2.203 (0.392, 4.533)	1.340 (0.236, 2.809)
θ_{QIRC}	3.212 (0.296, 6.828)	1.503 (0.232, 3.505)	2.222 (0.295, 4.800)	2.141 (0.192, 5.173)
θ_{QIR}	2.923 (0.724, 5.067)	1.990 (0.755, 3.258)	2.518 (1.266, 3.890)	2.792 (1.523, 4.167)
θ_{QI}	0.727 (0.198, 1.503)	1.033 (0.203, 2.427)	0.773 (0.177, 1.714)	1.157 (0.217, 2.714)
θ_J	1.017 (0.777, 1.272)	1.242 (0.954, 1.545)	1.107 (0.921, 1.298)	1.147 (0.934, 1.372)
θ_K	0.686 (0.177, 1.467)	0.799 (0.177, 1.808)	0.626 (0.170, 1.282)	0.959 (0.185, 2.241)
θ_L	1.911 (0.224, 4.493)	1.280 (0.399, 2.326)	1.568 (0.345, 2.936)	1.381 (0.315, 2.781)
θ_M	0.412 (0.245, 0.569)	0.430 (0.282, 0.578)	0.407 (0.275, 0.529)	0.415 (0.291, 0.543)
θ_N	0.439 (0.310, 0.574)	0.476 (0.350, 0.600)	0.440 (0.342, 0.543)	0.384 (0.301, 0.473)
θ_O	0.291 (0.190, 0.390)	0.422 (0.282, 0.552)	0.325 (0.239, 0.416)	0.350 (0.259, 0.443)
Speciation/introgression times (τ, $\times 10^{-3}$)				
$\tau_{QIRCDUS}$	3.297 (2.830, 3.851)	3.588 (3.062, 4.075)	3.423 (3.061, 3.783)	3.415 (3.066, 3.768)
τ_{QIRCDU}	1.872 (1.299, 2.456)	2.270 (1.703, 2.849)	2.029 (1.569, 2.489)	2.011 (1.673, 2.338)
τ_{QIRCD}	0.749 (0.634, 0.855)	0.750 (0.654, 0.837)	0.731 (0.642, 0.815)	0.753 (0.693, 0.815)
τ_{QIRC}	0.584 (0.469, 0.699)	0.673 (0.573, 0.765)	0.628 (0.542, 0.707)	0.697 (0.615, 0.766)
τ_{QIR}	0.363 (0.283, 0.452)	0.437 (0.360, 0.517)	0.389 (0.322, 0.452)	0.379 (0.312, 0.445)
τ_{QI}	0.267 (0.222, 0.312)	0.321 (0.272, 0.367)	0.290 (0.253, 0.327)	0.274 (0.238, 0.309)
$\tau_J = \tau_K = \tau_{QIRCD\to U}$	0.850 (0.738, 0.966)	0.902 (0.782, 1.022)	0.871 (0.778, 0.961)	0.832 (0.747, 0.917)
$\tau_L = \tau_M = \tau_{QI\to D}$	0.295 (0.246, 0.345)	0.334 (0.284, 0.381)	0.307 (0.268, 0.350)	0.298 (0.258, 0.336)
$\tau_N = \tau_O = \tau_{Q\to I}$	0.100 (0.066, 0.140)	0.091 (0.053, 0.132)	0.102 (0.074, 0.130)	0.094 (0.069, 0.118)
Introgression probabilities (φ)				
$\varphi_{QIRCD\to U}$	0.589 (0.338, 0.827) (∞)	0.686 (0.491, 0.870) (∞)	0.625 (0.442, 0.794) (∞)	0.587 (0.440, 0.733) (∞)
$\varphi_{QI\to D}$	0.118 (0.074, 0.165) (∞)	0.097 (0.060, 0.136) (∞)	0.106 (0.074, 0.139) (∞)	0.107 (0.077, 0.140) (∞)
$\varphi_{Q\to I}$	0.041 (0.014, 0.074) (8)	0.055 (0.026, 0.087) (∞)	0.050 (0.028, 0.074) (∞)	0.048 (0.028, 0.069) (∞)

Note.— Bayes factor B_{10} is given in parentheses, calculated using eq. 8: ∞ means that all sampled values of φ are $> \varepsilon = 1\%$.

Table S4. Power of BPP, HyDe and *D*-statistic tests of gene flow between sister species and average estimates of introgression probability in 100 simulated replicate datasets (each of 8000 loci) under the model of figure 8a

	Power			Proportion of
Methods	($\alpha = 1\%$)	($\alpha = 5\%$)	$\hat{\varphi}$±SD	invalid estimates
HyDe				
R→Ua	3%	8%	0.005 ± 0.003	48%
R→Ub	3%	5%	0.004 ± 0.004	51%
D-statistic				
R↔Ua	0%	1%	–	NA
R↔Ub	0%	2%	–	NA
BPP				
R→U	100%	100%	0.623 ± 0.066	0%

Note.— Bayesian test by BPP is considered significant at the 5% (or 1%) level if $B_{10} \geq 20$ (or 100). In the HyDe test, Ua and Ub were regarded as the 'hybrid' lineage to detect gene flow R→Ua and R→Ub, respectively, in figure 8**a**. In some datasets, the HyDe estimate of φ was outside the range (0, 1), and only the valid estimates were used to calculate the means.

Table S5. Power of BPP and HYDE tests of gene flow and average estimates of introgression probability in 100 simulated replicates under the three models of figure 8b-d

	BPP			HYDE			
# loci	Power ($\alpha = 1\%$)	Power ($\alpha = 5\%$)	$\hat{\varphi}$±SD	Power ($\alpha = 1\%$)	Power ($\alpha = 5\%$)	$\hat{\varphi}$±SD	Proportion of invalid estimates
Outflow asym (fig. 8**b**)							
500	39%	56%	0.096 ± 0.026	1%	4%	0.155 ± 0.111	44%
2000	100%	100%	0.104 ± 0.025	3%	9%	0.107 ± 0.057	33%
8000	100%	100%	0.105 ± 0.013	10%	24%	0.076 ± 0.042	20%
Inflow asym (fig. 8**c**)							
500	72%	84%	0.118 ± 0.030	23%	41%	0.331 ± 0.110	10%
2000	100%	100%	0.106 ± 0.014	87%	95%	0.321 ± 0.068	0%
8000	100%	100%	0.107 ± 0.009	100%	100%	0.325 ± 0.037	0%
inflow sym (fig. 8**b**, HYDE model)							
500	15%	27%	0.115 ± 0.037	2%	10%	0.124 ± 0.071	19%
2000	90%	95%	0.110 ± 0.022	14%	27%	0.101 ± 0.047	2%
8000	100%	100%	0.108 ± 0.010	83%	90%	0.108 ± 0.025	0%

Note.— The true introgression probability is $\varphi = 0.106$ (fig. 8**b**-**d**). See legend to table S4.