EVIDENCE FOR CLIMATE-DRIVEN DIVERSIFICATION? A CAUTION FOR INTERPRETING ABC INFERENCES OF SIMULTANEOUS HISTORICAL EVENTS

Running head: CAUTION FOR ABC MODEL SELECTION

Jamie R. Oaks[1,2,5], Jeet Sukumaran[1,2], Jacob A. Esselstyn[3], Charles W. Linkem[1,2], Cameron D. Siler[4], Mark T. Holder[2] and Rafe M. Brown[1,2]

[1]*Biodiversity Institute*

[2]*Department of Ecology and Evolutionary Biology*

University of Kansas

Lawrence, KS 66045

USA

[3]*Biology Department*

McMaster University

Hamilton, Ontario L8S4K1

Canada

[4]*Department of Biology*

University of South Dakota

Vermillion, SD 57069

USA

[5]*Corresponding author* (joaks1@ku.edu)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/j.1558-5646.2012.01840.x

ABSTRACT

Approximate Bayesian computation (ABC) is rapidly gaining popularity in population genetics. One example, `msBayes`, infers the distribution of divergence times among pairs of taxa, allowing phylogeographers to test hypotheses about historical causes of diversification in co-distributed groups of organisms. Using `msBayes`, we infer the distribution of divergence times among 22 pairs of populations of vertebrates distributed across the Philippine Archipelago. Our objective was to test whether sea-level oscillations during the Pleistocene caused diversification across the islands. To guide interpretation of our results, we perform a suite of simulation-based power analyses. Our empirical results strongly support a recent simultaneous divergence event for all 22 taxon pairs, consistent with the prediction of the Pleistocene-driven diversification hypothesis. However, our empirical estimates are sensitive to changes in prior distributions, and our simulations reveal low power of the method to detect random variation in divergence times and bias toward supporting clustered divergences. Our results demonstrate that analyses exploring power and prior sensitivity should accompany ABC model-selection inferences. The problems we identify are potentially mitigable with uniform priors over divergence models (rather than classes of models) and more flexible prior distributions on demographic and divergence-time parameters.

KEY WORDS: Approximate Bayesian computation; biogeography; model choice; `msBayes`; Philippines; Pleistocene diversification; simultaneous divergence

Approximate Bayesian computation (ABC) is a statistical technique burgeoning in many subfields of biology due to its flexibility and ease of accommodating complex, parameter-rich models without the need of calculating a likelihood. (see Beaumont, 2010; Bertorelle et al., 2010; Csilléry et al., 2010, for reviews). The technique approximates the posterior of a model by accumulating samples of parameters from the prior that yield summary statistics similar to the values taken by these statistics on the observed data. The parameter estimates are often regression-adjusted to improve the approximation by accounting for variation in the probability of the data across the parameter space of the retained sample (Beaumont et al., 2002; Leuenberger and Wegmann, 2010; Blum and François, 2009).

One popular implementation of the ABC algorithm, `msBayes` (Huang et al., 2011), provides a statistical method for testing biogeographic hypotheses that predict temporally clustered divergences among co-distributed groups of organisms. Specifically, the `msBayes` model infers the distribution of divergence times among pairs of populations. Throughout this paper, we use "clustered", "simultaneous", and "co-divergence" interchangeably to describe the situation where `msBayes` infers the same time of divergence for any subset of population pairs.

In applications of `msBayes`, researchers have often found support for temporally clustered divergences among co-distributed pairs of taxa (Barber and Klicka, 2010; Bell et al., 2012; Carnaval et al., 2009; Chan et al., 2011; Daza et al., 2010; Hickerson et al., 2006; Huang et al., 2011; Lawson, 2010; Leaché et al., 2007; Plouviez et al., 2009; Stone et al., 2012; Voje et al., 2009). However, previous investigators have not performed power analyses to inform their interpretation of shared divergence times. Rather, support for co-divergence has been taken as support for a shared event, without determining how much variation in divergence times is permissible while still leading to an inference of "simultaneous" divergence. In this study, we use simulations based on an empirical dataset from the Philippines to determine the power of the ABC method implemented in `msBayes`

for detecting temporal variation among divergences.

PLEISTOCENE MODEL OF DIVERSIFICATION IN AN ISLAND ARCHIPELAGO

The 7100+ islands of the Philippines may harbor the highest concentration of biodiversity on Earth (Brown and Diesmos, 2009; Heaney and Regalado, 1998), and have a relatively well-understood geologic history (Dickerson, 1928; Hall, 1998; Heaney, 1985; Inger, 1954; Voris, 2000; Yumul et al., 2008). During Pleistocene glacial cycles, sea-level fluctuations caused groups of previously isolated islands in the Philippines to undergo repeated cycles of connectivity and isolation (Voris, 2000). During glacial periods, when sea levels dropped to 120 meters below current levels, neighboring islands coalesced into seven main landmasses known as Pleistocene Aggregate Island Complexes (PAICs; Brown and Diesmos, 2002). In interglacial periods, rising sea levels split the PAICs into the set of islands we see today. There have been at least six of these climate-driven cycles during the last 500,000 years (Rohling et al., 1998; Siddall et al., 2003), with additional cycles occurring in the late Pliocene and early Pleistocene (Haq et al., 1987; Miller et al., 2005).

The repeated formation and fragmentation of PAICs has been proposed as a mechanism of diversification across the Philippine Islands. (Heaney, 1986; Brown and Diesmos, 2002, 2009). The PAIC model makes a specific prediction: If repeated bouts of connectivity and isolation promoted diversification, divergence times between populations on islands connected during glacial lowstands should be clustered and correspond to when sea levels rose. If the PAIC cycles did not cause diversification, then divergences among island populations must be dispersal-mediated, and would not be temporally clustered across different groups. We test this prediction by inferring the temporal distribution of divergences among 22 population pairs from a diverse set of vertebrate taxa using mitochondrial sequence data from nearly one thousand individuals from across the Philippines.

Methods

OVERVIEW OF THE DATA

We gathered datasets for which there were mitochondrial sequence data for multiple individuals from populations of a species (or two closely related species) from two different islands within a PAIC. By maximizing population sample sizes and avoiding multiple use of data, we ended up with 470 individuals from 22 population pairs spanning five orders of terrestrial vertebrates. Thirteen pairs are from the Greater Mindanao PAIC and nine are from the Greater Negros-Panay PAIC (Table S1). We used these samples to infer the pattern of divergence times for the 22 population pairs with `msBayes`. We used 499 additional samples to estimate appropriate models of nucleotide substitution, gene trees, and the population mutation rate (θ). These samples are from other island populations of the same species and closely related species (Table S6).

All megachiropteran sequences are from Roberts (2006a,b). All *Crocidura* shrew sequences are from Esselstyn et al. (2009). We included sequence data from geckos of the genera *Cyrtodactylus* and *Gekko* (Siler et al., 2010, 2012), frogs of the genus *Limnonectes* (Evans et al., 2003), bats of the genus *Hipposideros* (Esselstyn et al., 2012), and *Sphenomorphus*-group scincid lizards (currently assigned to the genera *Pinoyscincus* and *Insulasaurus*; Linkem et al., 2010, 2011), and augmented these datasets using the protocols in these papers to collect additional sequences. We collected sequences from snakes of the genus *Dendrelaphis* based on the protocols in Linkem et al. (2010) using the primers from Burbrink et al. (2000). All samples and their corresponding Genbank accession numbers are in Table S6.

PHYLOGENETIC ANALYSES

For phylogenetic analyses, we grouped closely related taxa, resulting in a total of 11 datasets that were easily aligned using MUSCLE (v3.7; Edgar, 2004) with no or few gaps (Dryad doi: XXXXXX). We used RAxML (v7.0.4; Stamatakis, 2006) to estimate

maximum likelihood (ML) trees for each of the 11 alignments, using the rapid hill-climbing heuristic algorithm (Stamatakis et al., 2007). For each RAxML analysis, we ran 100 search replicates, applied the 'GTRMIX' model of nucleotide substitution, and used random starting trees. On each of the 11 ML trees inferred by RAxML, we estimated parameters of the HKY85 model (Hasegawa et al., 1985) using PAUP* (v4.0b10; Swofford, 2003). We used these HKY85 parameter estimates in subsequent `msBayes` analyses.

We used the Bayesian information criterion (BIC; Schwarz, 1978) to select the best-fit model of nucleotide substitution for each alignment using PAUP* and ModelTest (v3.7; Posada and Crandall, 1998). We inferred an ultrametric tree for each alignment in BEAST (v1.5.4; Drummond and Rambaut, 2007) using the BIC-selected model and a constant substitution rate of 2×10^{-8} per site per year. For each dataset, we ran two independent BEAST analyses for 20 million generations and, after a conservative burn-in period of 5 million generations, sampled parameter values from the chain every 5,000 generations. We assessed stationarity by plotting the sampled parameter values and likelihood scores of both independent chains over generations, and confirming congruence between consensus trees from both posterior samples.

ESTIMATES OF θ TO GUIDE PRIOR SPECIFICATION

Broad uniform priors may cause low marginal likelihoods for complex models, leading Bayesian model selection procedures to prefer overly simplistic models (Lindley, 1957; Jeffreys, 1961). We therefore estimated the population mutation rate ($\theta = N_e\mu$) using 499 samples from 30 populations (Table S6), and subsequently used these estimates to guide our choice of prior distributions of θ in the `msBayes` analyses. None of these 30 populations were used in the `msBayes` analyses, thus we avoided statistical problems associated with multiple use of data. All θ estimates were derived from the same gene fragment that was used for the corresponding taxon pair in the `msBayes` analyses. We used Dendropy (v3.2.1; Sukumaran and Holder, 2010) to calculate Watterson's θ (θ_W; Watterson, 1975) and the average per site nucleotide differences (π; Nei and Li, 1979).

THE msBayes MODEL

Let Y be the number of population pairs, k_i be the number of loci sampled for the i^{th} population pair, and K be the total number of unique sampled loci. Let $\mathbf{X} = \{X_{1,1}, \ldots, X_{Y,k_Y}\}$ represent the data, which is a vector of multiple sequence alignments, where $X_{i,j}$ is the alignment of the j^{th} locus sampled for the i^{th} population pair. The joint posterior distribution of the model implemented in `msBayes` is given by

$$
\begin{aligned}
&f(\mathbf{G}, \Psi, \boldsymbol{\tau}, \boldsymbol{\theta}_A, \boldsymbol{\theta}_{D\mathbf{1}}, \boldsymbol{\theta}_{D\mathbf{2}}, \alpha, \boldsymbol{\upsilon}, \boldsymbol{\tau}_B, \boldsymbol{\zeta}_{D\mathbf{1}}, \boldsymbol{\zeta}_{D\mathbf{2}}, \mathbf{m} \,|\, \mathbf{X}, \boldsymbol{\phi}, \boldsymbol{\rho}, \boldsymbol{\nu}) \\
&= \frac{1}{f(\mathbf{X})} f(\Psi) f(\boldsymbol{\tau} \,|\, \Psi) f(\alpha) \left[\prod_{i=1}^{Y} f(\theta_{A,i}) f(\theta_{D1,i}, \theta_{D2,i}) f(\tau_{B,i}) f(\zeta_{D1,i}) f(\zeta_{D2,i}) f(m_i) \right. \\
&\left. \prod_{j=1}^{k_i} f(X_{i,j} \,|\, G_{i,j}, \phi_{i,j}) f(G_{i,j} \,|\, \tau_i, \theta_{A,i}, \theta_{D1,i}, \theta_{D2,i}, \rho_{i,j}, \nu_{i,j}, \upsilon_j, \tau_{B,i}, \zeta_{D1,i}, \zeta_{D2,i}, m_i) \right] \left[\prod_{j=1}^{K} f(\upsilon_j \,|\, \alpha) \right],
\end{aligned}
\tag{1}
$$

where $\mathbf{G} = \{G_{1,1}, \ldots, G_{Y,k_Y}\}$ are the gene trees upon which each $X_{i,j}$ evolved according to the HKY85 substitution model parameters $\phi_{i,j}$. The HKY85 model parameters for each alignment reside in vector $\boldsymbol{\phi} = \{\phi_{1,1}, \ldots, \phi_{Y,k_Y}\}$, and are fixed constants provided by the user. $\boldsymbol{\tau}$ is the vector of times, in coalescent units, when the populations of each pair diverged, $\{\tau_1, \ldots, \tau_Y\}$. Ψ is the hyper-parameter controlling the number of unique τ within $\boldsymbol{\tau}$. $\boldsymbol{\theta}_A = \{\theta_{A,1}, \ldots, \theta_{A,Y}\}$ is the vector of θ parameters for the ancestral population of each population pair. $\boldsymbol{\theta}_{D\mathbf{1}} = \{\theta_{D1,1}, \ldots, \theta_{D1,Y}\}$ is the vector of θ parameters for the 1^{st} descendant population of each pair; $\boldsymbol{\theta}_{D\mathbf{2}} = \{\theta_{D2,1}, \ldots, \theta_{D2,Y}\}$ is the same for the 2^{nd} descendant population of each pair. $\boldsymbol{\rho} = \{\rho_{1,1}, \ldots, \rho_{Y,k_Y}\}$ and $\boldsymbol{\nu} = \{\nu_{1,1}, \ldots, \nu_{Y,k_Y}\}$ are vectors of θ-scaling constants provided by the user (Table 1). Furthermore, there are locus-specific θ-scaling parameters in the vector $\boldsymbol{\upsilon} = \{\upsilon_1, \ldots, \upsilon_K\}$. α is the shape parameter of the gamma prior distribution on each υ. $\boldsymbol{\zeta}_{D\mathbf{1}} = \{\zeta_{D1,1}, \ldots, \zeta_{D1,Y}\}$ is the vector of θ-scaling parameters that determine the magnitude of the bottleneck in the 1^{st} descendant population of each population pair, whereas $\boldsymbol{\zeta}_{D\mathbf{2}} = \{\zeta_{D2,1}, \ldots, \zeta_{D2,Y}\}$ is the same for the 2^{nd} descendant population of each pair; for both descendants of each

population pair, the bottleneck begins immediately after divergence in forward-time. $\boldsymbol{\tau}_B = \{\tau_{B,1}, \ldots, \tau_{B,Y}\}$ is the vector of the proportions of time between present and τ when the bottleneck ends for both populations in each pair; after which the populations grow exponentially to present. $\mathbf{m} = \{m_1, \ldots, m_Y\}$ is the vector of symmetric migration rates between the descendant populations of each pair.

Prior terms of Equation eq:fullmodel

The prior terms of Equation eq:fullmodel within the product over population pairs include $f(\theta_{A,i})$, $f(\tau_{B,i})$, $f(\zeta_{D1,i})$, $f(\zeta_{D2,i})$, and $f(m_i)$. For each pair, these are independently and identically distributed (*iid*) as $\theta_A \sim U(a_\theta, b_{\theta_A})$; $\tau_B \sim U(0, 0.95)$; $\zeta_{D1} \sim U(0.01, 1)$; $\zeta_{D2} \sim U(0.01, 1)$; and $m \sim U(0, b_m)$. Let us further denote the mean of the two descendant populations of the i^{th} pair as $\theta_{D,i}$, which, for each pair, is *iid* as $\theta_D \sim U(a_\theta, b_{\theta_D})$. The prior on θ of the 1^{st} and 2^{nd} descendant population of the i^{th} pair ($f(\theta_{D1,i}, \theta_{D2,i})$) is then $\theta_{D1,i}, \theta_{D2,i} \sim Dirichlet(1,1) \times 2\theta_{D,i}$.

The terms in Equation eq:fullmodel outside of the product over population pairs include the hyper-prior probability distributions $f(\Psi)$, $f(\alpha)$, and $f(\boldsymbol{\tau} \,|\, \Psi)$. The prior on Ψ is uniformly distributed on the integers 1 to Y. The prior on α is $\alpha \sim U(1, 20)$. If we let $\mathbf{T} = \{T_1, \ldots, T_\Psi\}$ be the vector of the Ψ unique divergence times, then $f(\boldsymbol{\tau} \,|\, \Psi) = f(\mathbf{T} \,|\, \Psi) f(\boldsymbol{\tau} \,|\, \mathbf{T})$. Each T within $\mathbf{T}$ is *iid* as $T \sim U(0, b_\tau)$. Each T is placed in $\boldsymbol{\tau}$ once, and the remaining $Y - \Psi$ slots within $\boldsymbol{\tau}$ are populated by randomly drawing from $\mathbf{T}$ with replacement. If we let $x_1, \ldots, x_\Psi$ denote the number of times each $T_1, \ldots, T_\Psi$ is selected for the $Y - \Psi$ slots, then the probability mass function is

$$f(\boldsymbol{\tau} \,|\, \mathbf{T}) = f(x_1, \ldots, x_\Psi; Y - \Psi, p_1, \ldots, p_\Psi) = \frac{Y - \Psi!}{x_1! \cdots x_\Psi!} p_1^{x_1} \cdots p_\Psi^{x_\Psi}, \tag{2}$$

where $p_1 = p_2 = \ldots = p_\Psi = 1/\Psi$ and $\sum_{i=1}^{\Psi} x_i = Y - \Psi$.

The τ parameters are in coalescent units relative to a constant reference population size, θ_C/μ, where $\theta_C = b_{\theta_D}/2$. We denote these coalescent units as $4N_C$ generations. Thus,

the τ within $\boldsymbol{\tau}$ are proportional to real time, and can be converted to the number of generations of the reference population, τ_G, by assuming a mutation rate, μ, and using

$$\tau_G = \tau \times \frac{\frac{b_{\theta_D}}{2}}{\mu}. \tag{3}$$

Note, for each τ within $\boldsymbol{\tau}$ to be on the same scale of $4N_C$ generations, and thus comparable, `msBayes` assumes the relative mutation rates among the populations are fixed and known. The relative rates are fixed according to the values in $\boldsymbol{\nu}$ provided by the user. To get the divergence times into units proportional to the realized population size, and thus the expected number of mutations, `msBayes` scales the divergence times for each locus of each population pair before simulating data, creating the vector $\mathbf{t} = \{t_{1,1}, \ldots, t_{Y,k_Y}\}$. For the j^{th} locus of the i^{th} pair, τ_i is scaled by

$$t_{i,j} = \tau_i \times \frac{\theta_C}{\theta_{D,i}\rho_{i,j}}. \tag{4}$$

The user-defined θ-scaling constants in $\boldsymbol{\rho}$ can thus be used to account for known differences in ploidy among the loci and/or differences in generation times among the taxa.

`msBayes` allows for intra-locus recombination, which, for simplicity, is not included in Equation eq:fullmodel. If the intra-locus recombination rate, r, is allowed to be non-zero, another prior $f(r)$ would be outside the product over population pairs in Equation eq:fullmodel, and it would be distributed as $r \sim U(0, b_r)$.

The term within the product over the K unique loci is the prior probability density of the θ-scaling parameter of the j^{th} locus, $f(\upsilon_j \,|\, \alpha)$. Each of these parameters is *iid* as $\upsilon \sim \Gamma(\alpha, 1/\alpha)$.

We use Θ to denote all of the parameters of the `msBayes` model, and $f(\Theta)$ to represent the joint prior probability distribution of the model (Table 1).

Likelihood terms of Equation eq:fullmodel

For the j^{th} locus of the i^{th} population pair, the term $f(X_{i,j} \mid G_{i,j}, \phi_{i,j})$ is the probability of the sequence alignment given a gene tree and HKY85 parameters, or the Felsenstein likelihood (Felsenstein, 1981). If the intra-locus recombination rate is allowed to be non-zero, Equation eq:fullmodel would require another product over the columns of each sequence alignment to allow sites to have different genealogies. The term $f(G_{i,j} \mid \ldots)$ is the probability of the gene tree under a multi-population coalescent model where the ancestral population of constant size $\theta_{A,i}\rho_{i,j}\nu_{i,j}\upsilon_j$ diverges at time τ_i into two descendant populations of constant size $\theta_{D1,i}\rho_{i,j}\nu_{i,j}\upsilon_j\zeta_{D1,i}$ and $\theta_{D2,i}\rho_{i,j}\nu_{i,j}\upsilon_j\zeta_{D2,i}$ that exchange migrants at symmetric rate m_i. After time $\tau_{B,i} \times \tau_i$ they grow exponentially to present size $\theta_{D1,i}\rho_{i,j}\nu_{i,j}\upsilon_j$ and $\theta_{D2,i}\rho_{i,j}\nu_{i,j}\upsilon_j$, respectively. Lastly, $f(\mathbf{X})$ is the marginal likelihood of the model, or the probability of the data.

THE ABC IMPLEMENTATION OF THE `msBayes` MODEL

`msBayes` does not estimate the posterior in Equation eq:fullmodel. Rather, it distills the alignments $\mathbf{X} = \{X_{1,1}, \ldots, X_{Y,k_Y}\}$ into vectors of summary statistics $\mathbf{S}^* = \{S^*_{1,1}, \ldots, S^*_{Y,k_Y}\}$ and estimates the approximate joint posterior distribution

$$
\begin{aligned}
&f(\mathbf{G}, \Psi, \boldsymbol{\tau}, \boldsymbol{\theta}_A, \boldsymbol{\theta}_{D\mathbf{1}}, \boldsymbol{\theta}_{D\mathbf{2}}, \alpha, \boldsymbol{\upsilon}, \boldsymbol{\tau}_B, \boldsymbol{\zeta}_{D\mathbf{1}}, \boldsymbol{\zeta}_{D\mathbf{2}}, \mathbf{m} \mid B_\epsilon(\mathbf{S}^*), \boldsymbol{\phi}, \boldsymbol{\rho}, \boldsymbol{\nu}) \\
&= \frac{1}{f(B_\epsilon(\mathbf{S}^*))} f(\Psi) f(\boldsymbol{\tau} \mid \Psi) f(\alpha) \left[\prod_{i=1}^{Y} f(\theta_{A,i}) f(\theta_{D1,i}, \theta_{D2,i}) f(\tau_{B,i}) f(\zeta_{D1,i}) f(\zeta_{D2,i}) f(m_i) \right. \\
&\left. \prod_{j=1}^{k_i} f(B_\epsilon(S^*_{i,j}) \mid G_{i,j}, \phi_{i,j}) f(G_{i,j} \mid \tau_i, \theta_{A,i}, \theta_{D1,i}, \theta_{D2,i}, \rho_{i,j}, \nu_{i,j}, \upsilon_j, \tau_{B,i}, \zeta_{D1,i}, \zeta_{D2,i}, m_i) \right] \left[\prod_{j=1}^{K} f(\upsilon_j \mid \alpha) \right],
\end{aligned}
\tag{5}
$$

where $B_\epsilon(\mathbf{S}^*)$ is the multi-dimensional Euclidean space around the vector of observed summary statistics, the radius of which is the tolerance ϵ.

To estimate this approximate posterior, `msBayes` uses an ABC rejection algorithm followed by regression adjustment. The first step of the algorithm is to draw n samples

from the joint prior, $f(\Theta)$. In the case of one locus per population pair, `msBayes` draws a sample from $f(\Theta)$ by (1) drawing Ψ from the integers 1 to Y; (2) drawing Ψ divergence times from $U(0, b_\tau)$ in units of $4N_C$ generations, to get $\mathbf{T} = \{T_1, \ldots, T_\Psi\}$; (3) randomly assigning each T to $\boldsymbol{\tau}$ once, and filling the remaining $Y - \Psi$ slots within $\boldsymbol{\tau}$ by randomly drawing from $\mathbf{T}$ with replacement, to get $\boldsymbol{\tau} = \{\tau_1, \ldots, \tau_Y\}$; (4) drawing values of the demographic parameters θ_A, θ_{D1}, θ_{D2}, τ_B, ζ_{D1}, ζ_{D2}, and m for each population pair from their respective prior distributions; (5) scaling each τ to get $\mathbf{t} = \{t_{1,1}, \ldots, t_{Y,k_Y}\}$ via Equation eq:timescaler; (6) simulating sequence data for each population pair according to the multi-population coalescent model described above; and (7) calculating population genetic summary statistics S for each pair from the simulated sequence matrix. The result of one draw from the prior is the parameter vector Λ and vector of summary statistics $\mathbf{S} = \{S_1, \ldots, S_Y\}$. For Λ, `msBayes` reports Ψ and three summary statistics calculated from $\boldsymbol{\tau}$: the mean ($E(\tau)$), variance ($Var(\tau)$), and dispersion index ($\Omega = Var(\tau)/E(\tau)$) (Hickerson et al., 2006; Huang et al., 2011). After repeating this process n times, we have a sample of parameter vectors, $\mathcal{P}_{f(\Theta)} = \{\Lambda_1, \ldots, \Lambda_n\}$, randomly drawn from $f(\Theta)$, and the associated vectors of summary statistics, $\mathbb{S} = \{\mathbf{S_1}, \ldots, \mathbf{S_n}\}$.

The vector $\mathbf{S}^*$ contains the same summary statistics calculated from the observed sequence data. During the rejection step, only the samples from $\mathcal{P}_{f(\Theta)}$ with $\mathbf{S}$ that fall within the Euclidean space $B_\epsilon(\mathbf{S}^*)$ are retained. We denote the set of retained samples as $\mathcal{P}_\epsilon$. Regression techniques are then used to adjust $\mathcal{P}_\epsilon$ for variation in the probability of the data across the retained parameter sample space (Beaumont et al., 2002; Leuenberger and Wegmann, 2010; Blum and François, 2009). The result is an estimate of the approximate posterior, which we denote $\mathcal{P}_{f(\Theta|B_\epsilon(\mathbf{S}^*))}$.

ESTIMATING THE PATTERN OF DIVERGENCE TIMES

To estimate the temporal pattern of divergences among the 22 population pairs, we used a modified version of `msBayes` v20100519 (Huang et al., 2011). In the midst of our work, we identified a bug in version 20100519 that mis-specified the prior on the θ_A

parameters, which we subsequently corrected. This error has also been corrected in version 20120222 of `msBayes` (full details in the Supporting Information).

Specifying and simulating the joint prior

To use `msBayes`, one must specify b_τ, a_θ, b_{θ_D}, b_{θ_A}, b_m, and b_r to control the prior distributions $\tau \sim U(0, b_\tau)$, $\theta_D \sim U(a_\theta, b_{\theta_D})$, $\theta_A \sim U(a_\theta, b_{\theta_A})$; $m \sim U(0, b_m)$; and $r \sim U(0, b_r)$. Given our island system and mitochondrial data, we assumed no migration ($b_m = 0$) and no recombination ($b_r = 0$). Currently, `msBayes` only offers the continuous uniform distribution to represent *a priori* knowledge about parameters of θ and τ. Thus, broad prior distributions must be used to avoid assigning zero probability density to plausible regions of parameter space. In our case, we must choose prior distributions that span the range of possible values for the 66 θ and 22 τ parameters. Specifically, we chose prior settings of $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$. We chose the θ settings to assure we spanned the 30 empirical estimates of θ calculated from independent data. We chose the prior on τ to represent the large *a priori* uncertainty about the divergence times for all 22 pairs of populations. Two of our population pairs (*Crocidura negrina-C. panayensis* and *Cyrtodactylus gubaot-C. sumuroi*) represent distinct species, and the species-level taxonomy of many vertebrate groups in the Philippines often masks deeply divergent cryptic lineages (Brown et al., 2008; Siler and Brown, 2010; Siler et al., 2011a,b, 2012; Welton et al., 2010; Linkem et al., 2010; Brown and Stuart, 2012; Esselstyn et al., 2012). Given the precedents for cryptic diversity in the Philippines, the long geological history of the archipelago (Brown and Diesmos, 2009; Yumul et al., 2008), and the large amount of uncertainty regarding the mutation rates and generation times of the taxa examined, we chose $b_\tau = 20$ to avoid giving plausible divergence times zero probability. Assuming $\mu = 2 \times 10^{-8}$ and applying Equation eq:realtime, this prior translates to $\tau_G \sim U(0, 5 \times 10^7)$ in generations.

We used the `msbayes.pl` script to generate $\mathcal{P}_{f(\Theta)}$ with 10^7 random samples from $f(\Theta)$, using the ML estimates of the HKY85 model parameters for each of the population

pairs and assuming μ is equal across all taxa. Each $\mathbf{S}$ contained π (Tajima, 1983), θ_W (Watterson, 1975), π_{net} (Takahata and Nei, 1985), and $SD(\pi - \theta_W)$ (Tajima, 1989) for each of the 22 population pairs.

Rejection sampling and regression adjustment

We used the `acceptRej.pl` script of the `msBayes` package and ABCtoolbox (v1.1; Wegmann et al., 2010) to estimate $\mathcal{P}_\epsilon$ by specifying ϵ so that 1000 samples from $\mathcal{P}_{f(\Theta)}$ were within $B_\epsilon(\mathbf{S}^*)$ and thus retained. ABCtoolbox and `msBayes` produced identical $\mathcal{P}_\epsilon$. We used two regression methods to adjust $\mathcal{P}_\epsilon$ and estimate the approximate posterior, $\mathcal{P}_{f(\Theta|B_\epsilon(\mathbf{S}^*))}$: (1) the weighted, local-linear regression (ABC_{LLR}) adjustment of Beaumont et al. (2002) as implemented in `msBayes` and (2) the general linear model (ABC_{GLM}) regression adjustment of Leuenberger and Wegmann (2010) as implemented in ABCtoolbox. To avoid additional notation, we include the multinomial logistic regression adjustment of Ψ used by `msBayes` under ABC_{LLR}. For both methods we used the same set of summary statistics as for the rejection step.

Vetting prior sample size and PAIC-specific analyses

To assess whether 10^7 samples from $f(\Theta)$ were sufficient, we split the prior samples into two sets of 5×10^6 and repeated the rejection sampling and ABC_{GLM}-regression adjustment using each subset. To compare the results between the two different PAICs, we repeated the ABC methods above on the nine population pairs from the Greater Negros-Panay PAIC and the 13 population pairs from the Greater Mindanao PAIC (Table S1) separately; we generated 5×10^6 prior samples for these PAIC-specific analyses.

Reducing summary statistics to PLS components

In practice, using too many summary statistics may introduce noise into ABC estimates (Joyce and Marjoram, 2008; Wegmann et al., 2009). In our case, each $\mathbf{S}$ contains 88 summary statistics (the four default `msBayes` summary statistics for each of the 22 population pairs). To reduce the dimensionality of $\mathbb{S}$, we transformed the 88 summary

statistics into 10 partial least squares (PLS) orthogonal components, following an initial Box-Cox transformation of the statistics (Wegmann et al., 2009). We used the `find_pls.r` script of ABCtoolbox to define the PLS components using all 10^7 samples within $\mathbb{S}$. Using these PLS definitions, we reduced $\mathbb{S}$ and $\mathbf{S}^*$ to $\mathbb{S}_{PLS}$ and $\mathbf{S}^*_{PLS}$ with 10 PLS components. Hereafter, we will refer to the prior sample containing the original 88 summary statistics as $\mathbb{S}_{stats}$ to distinguish it from $\mathbb{S}_{PLS}$. We repeated all the rejection-sampling and regression adjustment procedures discussed above using $\mathbb{S}_{PLS}$ and $\mathbf{S}^*_{PLS}$. Accommodating the PLS components during the rejection sampling and ABC_{LLR} adjustment in `msBayes` required modifications to the `acceptRej.pl` and `acceptRej.r` scripts provided with the package; the modified scripts are available at `https://github.com/joaks1/msbayes-hack`.

ASSESSING THE PERFORMANCE AND POWER OF `msBayes`

What does "simultaneous" mean?

The PAIC model predicts a pattern of recent and clustered divergences among co-distributed taxa (i.e., groups of divergence times associated with Pleistocene glacial cycles). Using `msBayes` to test this prediction assumes the method can reliably discriminate recent clustered divergences from random divergences. There have not been any simulation-based assessments of the power of `msBayes` to detect variation in divergence times. We simulated 1000 pseudo-replicate datasets (i.e., 1000 Λ with associated $\mathbf{S}$) with τ for each of the 22 population pairs (i.e., $\Psi = 22$) randomly drawn from a uniform distribution, $U(0, \tau_{max})$, where τ_{max} was set to: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 2.0, and 3.0, in $4N_C$ generations. We used the `msbayes.pl` script to generate the pseudo-replicates with all other settings the same as when we generated $\mathcal{P}_{f(\Theta)}$.

For each of the 17,000 simulated datasets, we repeated the same ABC inference procedures that we used for the empirical data. Specifically, we used $\mathbb{S}_{stats}$, ABC_{GLM}, and 5×10^6 samples from $f(\Theta)$. Using version 20100519 of `msBayes` (i.e., pre-bug fix), we explored all of the ABC methods used for the empirical data (ABC_{LLR} versus ABC_{GLM} and $\mathbb{S}_{stats}$ versus $\mathbb{S}_{PLS}$) as well as 2×10^6 versus 10^7 samples from the prior. All of these

methods performed similarly (see SI). The error in version 20100519 of `msBayes` enforces a lower bound of 0.01 on the prior for θ_A (see SI for specifics). This hard-coded bound was the same for simulations generating both the pseudo-observed datasets and the prior sample (i.e., the prior was correct). Thus, the bug should have little effect on the relative performance of the methods.

ASSESSING PRIOR SENSITIVITY OF `msBayes`

Our simulation-based analyses reveal a bias toward inferring clustered divergences (see Results), which could be caused by broad priors on θ and τ (Lindley, 1957). We studied this in two ways. First, we quantified the behavior of `msBayes` under the ideal conditions where the prior distributions are correct. To do this, we simulated 100,000 datasets by drawing parameter values from the prior ($\tau \sim U(0, 10)$, $\theta_D \sim U(0.0001, 0.05)$, and $\theta_A \sim U(0.0001, 0.025)$) and analyzed these datasets using 2×10^6 samples drawn from the same prior distributions. For all these simulations, we used the default distribution on Ψ (i.e., discrete uniform from 1 to Y). To greatly reduce computational time, these simulations used a single 1000 bp locus sampled from 10 individuals from 10 population pairs. We were particularly interested in the posterior probability of a single divergence time for each of these simulated datasets. So, we assigned the 100,000 estimates of the posterior probability of one divergence event (i.e., $p(\Psi = 1|B_\epsilon(\mathbf{S}^*))$) to 20 bins of width 0.05, and plotted the estimated $p(\Psi = 1|B_\epsilon(\mathbf{S}^*))$ of each bin against the proportion of replicates in that bin with a true value of $\Psi = 1$ (Huelsenbeck and Rannala, 2004). We repeated this exercise using the alternative criterion for one divergence event, $\hat{\Omega} < 0.01$ (Hickerson et al., 2006).

Our second evaluation of prior sensitivity attempted to assess the performance of `msBayes` under optimal real-world conditions. In this case, the priors used are not known to be correct, but represent the narrowest possible prior distributions (i.e., priors informed by the data). We repeated the simulation-based power analyses outlined above using two additional, highly informative prior settings: (1) $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and

$\theta_A \sim U(0.0005, 0.02)$; and (2) $\tau \sim U(0, 5)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$. For both prior settings, we generated 5×10^6 samples. Using the informed θ priors, we simulated an additional 17,000 pseudo-observed datasets with divergence times randomly drawn from the same series of uniform distributions (i.e., τ_{max}) described above. The priors on θ match the range of θ estimates from our empirical data (Tables S2 and S3). Assuming $\mu = 2 \times 10^{-8}$, the priors on τ translate to $\tau_G \sim U(0, 1 \times 10^7)$ and $\tau_G \sim U(0, 5 \times 10^6)$ in units of generations of the reference population, respectively. If we assume one generation per year, the latter prior is very similar to our range of gene divergence estimates (0.2–4 mybp; Fig. S1), and thus, given the uncertainty in generation times and mutation rates for the taxa we examined, is likely too narrow (i.e., excluding plausbile values).

We must emphasize that we only use such informed priors here to represent the narrowest possible priors for a real-world application of `msBayes`. Basing priors on estimates from the data is not a fully Bayesian statistical procedure, and we are not advocating such an approach. Whereas empirical Bayes, also known as maximum marginal likelihood, is a commonly used statistical framework for point estimation, empirical Bayesian estimates of the posterior distribution of parameters are too narrow and often inappropriately shaped and off-center (Morris, 1983; Laird and Louis, 1987; Carlin and Gelfand, 1990). This is because they fail to account for the uncertainty in estimating the prior. Many post-hoc correction methods have been proposed for estimating frequentist-like confidence intervals from empirical Bayesian estimates of the posterior distribution of parameters (Morris, 1983; Laird and Louis, 1987, 1989; Carlin and Gelfand, 1990; Hwang et al., 2009), but none are implemented in `msBayes`, and none would correct model choice estimates (e.g., posterior probabilities of models and Bayes factors).

A NOTE ON COALESCENT UNITS

Because our empirical data is comprised of mitochondrial sequences, our estimates of $E(\tau)$ and Ω are in coalescent units of N_C generations. However, the coalescent units of our simulated pseudo-observed data are arbitrary and, via Equation eq:realtime, can be

converted into generations by assuming μ (i.e., the conversion is independent of the inheritance constant of the coalescent units). When reporting our simulation results in coalescent units, we write "$4N_C$" generations, because most users of `msBayes` will be using diploid, biparentally inherited loci.

Results

ESTIMATES OF GENE DIVERGENCES AND θ

Our estimates of the gene divergence times range from approximately 0.2 to 4 mybp, with 16 of the 22 posterior mean estimates within the past million years (Fig. S1). Our estimates of θ from independent population samples range from 0.0011 to 0.0181 (Table S2); our estimates from the 44 populations we analyzed in `msBayes` had a greater range of 0.0003 to 0.0381 (Table S3).

EMPIRICAL `msBayes` ESTIMATES UNDER BROAD PRIORS

When using prior settings of $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$, our `msBayes` results strongly support one recent divergence event for all 22 population pairs, regardless of the method of post-sampling regression (ABC_{LLR} or ABC_{GLM}) or summary statistic matrix ($\mathbb{S}_{stats}$ or $\mathbb{S}_{PLS}$) used (Fig. S2 and Table 2). Consistent with one divergence event, all methods yield estimates of the dispersion index of population divergence times (Ω) of essentially zero (Table 2). $\hat{\Omega} < 0.01$ is commonly used as a criterion for one divergence time shared by all pairs of populations (Hickerson et al., 2006).

Estimates of the time of the divergence event ($\hat{E(\tau)}$) range from approximately 0.04–0.1 coalescent units ago (**??** 2). Assuming $\mu = 2 \times 10^{-8}$ and one generation per year, this translates to 100,000–250,000 years ago, consistent with Pleistocene-driven diversification. We used simulations to reject the possibility that these results can be explained by a model of no divergence (i.e., panmixia between the populations of each pair;

see SI).

ABC_{LLR} and ABC_{GLM} estimates of $E(\tau)$ are almost identical, regardless of whether using $\mathbb{S}_{stats}$ or $\mathbb{S}_{PLS}$. Both regression methods had larger confidence intervals when using $\mathbb{S}_{PLS}$ (**??** 2). Estimates based on 5×10^6 and 10^7 prior samples are very similar, regardless of whether $\mathbb{S}_{stats}$ or $\mathbb{S}_{PLS}$ are used (Table 2). Estimates from the Greater Mindanao and Negros-Panay PAICs were similar to the combined analyses (Table 2).

SIMULATION-BASED ASSESSMENT OF "SIMULTANEOUS"

Accuracy and precision of estimates with broad priors

The precision of the ABC estimates of $E(\tau)$ and Ω is low, especially when the true divergence times are more recent (Figs. 1 and S3). Also, `msBayes` is less accurate and precise in estimating Ω (Fig. 1) than $E(\tau)$ (Fig. S3). When τ_{max} is less than 0.9 coalescent units, `msBayes` tends to underestimate Ω, whereas it tends to overestimate Ω when τ_{max} is 1.0 or greater (Fig. 1). From our simulation results using version 20100519 of `msBayes` (pre-bug fix), all combinations of summary statistics and regression-adjustment methods are inaccurate for estimating Ω across the τ_{max} we simulated (Figs. $\text{Sfig}_a cc_{os} s_l lr_b ug, fig_a cc_{os} s_g lm_b ug, fig_a cc_{op} ls_l lr_b ug, fig_a cc_{op} ls_g lm_b ug$).

Power of `msBayes`

The power of `msBayes` to detect variation in divergence times is low at phylogeographic time scales (Figs. $\text{fig}_p ow_{os} s_g lm, fig_p ow_p si_m odes_s s_g lm$). $If we judge the procedure by conditions that lead to \leq$ 5% of the simulation replicates estimating one divergence event, then we find that `msBayes` is unable to reject one divergence event based on $\hat{\Psi}$ when τ_{max} is less than 1.3 coalescent units (i.e., $5.2N_C$ generations; Fig. 2). This translates to `msBayes` inferring a single divergence event more than 5% of the time when the 22 population pairs diverged randomly over the past 3 million generations (assuming $\mu = 2 \times 10^{-8}$).

`msBayes` does better based on $\hat{\Omega}$, but still cannot reject one divergence when τ_{max} is less than 0.8 coalescent units (i.e., $3.2N_C$ generations, or 2 million generations assuming $\mu = 2 \times 10^{-8}$; Fig. 3). This lack of power was not due to an overly stringent threshold (0.01). The true values of Ω from the simulations under all τ_{max} are consistent with multiple divergence times (i.e., greater than 0.01; Fig. S4). Thus, based on the criterion $\hat{\Omega} < 0.01$, `msBayes` would have rejected one divergence event for all the τ_{max} values we simulated if it was able to accurately estimate Ω. The lack of power is due to the bias of `msBayes` to underestimate Ω at recent divergence times (Fig. 1).

We can also assess power of the method by examining what conditions lead to $\leq 5\%$ of the simulation replicates estimating strong support for $\Psi = 1$. We use a Bayes factor of greater than 10 for the one divergence model compared to all other models ($BF_{\Psi=1,\Psi\neq1} > 10$) as a threshold for strong support (Jeffreys, 1935, 1961). According to this test, the method cannot reject one divergence when τ_{max} is less than 1.3 coalescent units (Fig. 5). Thus, assuming $\mu = 2 \times 10^{-8}$, `msBayes` strongly supports one divergence event more than 5% of the time when divergence times are random over the past 3 million generations. If we increase our threshold for strong support to a posterior probability greater than 0.95 ($BF_{\Psi=1,\Psi\neq1} > 399$), `msBayes` is still unable to reject one divergence event when τ_{max} is less than 0.8 coalescent units (2 million generations).

Using the criteria above based on the inference of one divergence event (or strong support for one divergence event) is quite generous, because an inference of $\hat{\Omega} > 0.01$ or $\hat{\Psi} > 1$ is not equivalent to success of the method. This is clearly illustrated by Figure 2. `msBayes` infers highly clustered divergence times across all of the τ_{max} we simulated. For example, even when divergence times are random over the past 3 coalescent units (7.5 million generations), the most probable inference of `msBayes` is still only two divergence events (i.e., $\hat{\Psi} = 2$).

Based on our simulation results using version 20100519 of `msBayes`, the power to detect variation in divergence times was low under all combinations of summary statistics,

regression-adjustment methods, and prior sample sizes we explored (Figs. Sfig$_p ow_p si_m odes_s s_g lm_b ug$, $fig_p ow_p si_m odes_p ls_l lr_b ug$, $fig_p ow_p si_m odes_p ls_g lm_b ug$, $fig_p ow_{os} s_l lr_b ug$, fi

PRIOR SENSITIVITY OF `msBayes`

Simulation results

Under the ideal conditions when the priors are correct, `msBayes` provides reasonable estimates of the posterior probability of one divergence event (Fig. 4). Based on Ψ and Ω results from $\mathcal{P}_\epsilon$ (i.e., unadjusted estimates), `msBayes` overestimates the posterior probability of one divergence event when the true probability is less than ≈ 0.6, and slightly underestimates it when the true probability is greater than ≈ 0.6 (Fig. 4A&B). However, regression adjustment of Ω using both ABC_{LLR} and ABC_{GLM} causes extreme underestimation of the posterior probability of one divergence event (Fig. 4D&F). Regression adjustment of Ψ has the same affect, but less extreme (Fig. 4C&E). Thus, Ψ is a better estimator of the posterior probability of one divergence event than Ω (Fig. 4C–F). Note that the adjustment of Ψ in `msBayes` (using multinomial logistic regression) failed for approximately 2% (2043) of the simulations, almost all of which had a true Ψ of one. Based on the unadjusted estimates, these failures would have been underestimates of the posterior probability of $\Psi = 1$. Thus, the plotted ABC_{LLR} results (Fig. 4C) are skewed due to these failed replicates, and should look more like the ABC_{GLM} results (Fig. 4E). Overall, when the priors are correct, regression-adjusted estimates of the posterior probability of one divergence event are downward biased. This is consistent with our observation of the switch in bias of Ω estimates; as τ_{max} becomes more similar to the prior, the bias switches from upward to downward (Fig. 1).

When the priors are informed by the data (i.e., narrower than possible under real-world applications of `msBayes`), the method's power is similar to using broad priors. Based on estimates of Ψ, `msBayes` cannot reject one divergence event when τ_{max} is less than 1.3 coalescent units (1.3 million generations, assuming $\mu = 2 \times 10^{-8}$) when prior settings are $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$, and 1.2

coalescent units when the prior settings are $\tau \sim U(0, 5)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$ (Figs. S9 and S10). Also, under both informed prior settings, `msBayes` still infers highly clustered divergences for all the τ_{max} we simulated (Figs. S9 and S10). The bias of `msBayes` to underestimate Ω at recent divergence times remains when using the informed priors (Figs. S7 and S8). As a result, the lack of power to detect variation in divergence times based on Ω estimates also remains (Figs. S11 and S12). Specifically, the method cannot reject a single divergence event when τ_{max} is less than 0.9 coalescent units (900,000 generations, assuming $\mu = 2 \times 10^{-8}$) when the prior settings are $\tau \sim U(0, 5)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$, and 0.8 coalescent units when the prior settings are $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$ (Figs. S11 and S12).

Empirical results

The results of `msBayes` analyses of the empirical data are very sensitive to the different prior settings we explored (Table 3). The estimates of Ω and $E(\tau)$ vary by over two and one order of magnitude, respectively, and the estimated 95% highest posterior density intervals do not overlap between some of the analyses. Our estimate of the posterior probability of one divergence event is essentially one when using appropriately broad priors, but is nearly zero with narrow priors (Table 3). Thus, we find strong support for different biogeographic scenarios (i.e., one divergence versus multiple divergences) under different prior distributions. However, under the informed priors, `msBayes` still infers highly clustered divergence times ($\hat{\Psi} = 2$; Table 3), suggesting the estimates under all three prior settings may suffer from the biases revealed by our simulation results.

Discussion

POWER AND BIAS ISSUES

Given the strong support for either one or two shared recent divergence(s) among the 22 taxon pairs (Tables 2 and 3), it would have been easy to accept the empirical results of `msBayes` as evidence for climate-driven diversification across the Philippine Archipelago. However, our simulations demonstrate a bias in the `msBayes` model toward inferring clustered divergences among population pairs when the divergences are random and relatively recent. We found `msBayes` always inferred temporally clustered divergences even when the taxon pairs diverged randomly over the past $12N_C$ generations (Fig. 2L). To put this in real time, assuming a mutation rate of 2×10^{-8} and a one year generation time, `msBayes` consistently infers an interesting biogeographical scenario when the taxon pairs diverged randomly over the past 7.5 million years. Also, `msBayes` will often ($> 5\%$ of the time) infer the extreme case of *one* divergence event with strong support when the taxon pairs diverged randomly over the past $3.2N_C$ generations (2 million years) (Fig. 3 and S5).

The bias did not improve when using empirically guided prior distributions (Fig. S7, S8, S11,and S12). Thus, our results show that even when using priors that are more informative than could be expected in a real-world application of `msBayes`, the method tends to infer simultaneous divergence too frequently.

`msBayes` is primarily used for comparing shallow divergences (Hickerson et al., 2006; Huang et al., 2011), which, according to our results, is when the method can be the most misleading. Due to the stochasticity of coalescent and mutational processes, a large variance in genetic divergence is expected among recently co-diverged taxa. Thus, any inference method for estimating divergence times is expected to struggle when applied to recently diverged taxa (and over much of the range of parameters we simulated here). Ideally, a method should express uncertainty in the face of such large expected variance. However, our simulations indicate that `msBayes` often returns strong support for the spurious conclusion of one divergence event (Fig. S5).

These findings are worrisome, because co-divergence is often a biologically interesting result that is interpreted as evidence for a shared historical event or barrier.

Our results suggest that any application of `msBayes` on recently diverged taxa is very likely to result in clustered divergences (Figs. 2, S9, and S10), and thus an "interesting" biogeographical interpretation.

PRIOR SENSITIVITY OF EMPIRICAL ESTIMATES

We find strong support for contradictory hypotheses (i.e., one versus multiple divergence events) and very different parameter estimates under the prior settings we explored (Table 3). This sensitivity is problematic, and introduces another challenge to interpreting the results of `msBayes` analyses. When different prior settings yield contrasting results, it is difficult to draw biologically meaningful conclusions.

The same sensitivity to the prior settings was not observed in our simulation-based results. This is not unexpected. All of the pseudo-observed data were generated by the `msBayes` model, with all of the parameters (except $\boldsymbol{\tau}$) drawn from the prior distributions (i.e., the priors were correct). As a result, changes in the priors for the parameters θ_D and θ_A were reflected in the simulated data. This, of course, is not true for our empirical data, where the parameter values are unknown. Thus, the empirical estimates are expected to be more variable under the different prior settings than the simulation results. Also, the empirical data likely violate assumptions of the `msBayes` model that are met by the simulated data, which could contribute to the prior sensitivity we observed with the empirical data.

The empirical estimates of Ψ switched from one to two between the broad and informed priors (Table 3). Such a switch occurred for some of the pseudo-replicate datasets simulated at larger τ_{max} values and subsequently analyzed under a prior of $\tau \sim U(0, 10)$ versus $\tau \sim U(0, 5)$ (Figs. S9 and S10). Thus, some of the prior sensitivity we observed in analyses of the empirical data is also evidenced in the simulation results.

Our correction of the error in version v20100519 of `msBayes` concerning the θ_A prior also had an impact on the results of the empirical analyses. Our Ω estimates vary as much as two-fold before and after this error was corrected (Table S5). Given the sensitivity of

our results to this error and prior settings, and the biases revealed by our simulations, we suggest that previously published results of `msBayes` be treated with some caution.

POSSIBLE CAUSES OF BIAS

`msBayes` implements a model selection procedure by approximating the posterior probabilities for different divergence models. The total number of unique divergence time models, or the number of possible ways to assign Y taxon pairs into Ψ divergence time categories, is calculated by the Bell number (Bell, 1934)

$$B_Y = \sum_{\Psi=1}^{Y} \left[\frac{1}{\Psi!} \sum_{j=0}^{\Psi-1} (-1)^j \binom{\Psi}{j} (\Psi - j)^Y \right]. \tag{6}$$

The number of unique models is enormous, even with a moderate number of taxon pairs (4,506,715,738,447,323 unique divergence models when $Y = 22$). The model implemented in `msBayes` takes advantage of the independence of the τ parameters and empirical sample sizes, allowing exchangeability of the summary statistics simulated under the various empirical sample sizes (Hickerson et al., 2006; Huang et al., 2011). This allows the identity/order of each τ within $\boldsymbol{\tau}$ to be ignored, which greatly reduces the divergence model space (1002 divergence models when $Y = 22$). The total number of unique divergence time models reduces to the number of partitions of Y (i.e., the integer partition), or the number of ways you can write Y as a sum of positive integers when the order of the addends does not matter (Sloan, 2011a,b). The closed-form expression for the integer partition is beyond the scope of this paper (but see Malenfant, 2011), but functions written in python to calculate it are provided in the appendix.

The prior on divergence models

In `msBayes`, the prior on Ψ is discrete uniform from 1 to Y, but it is not uniform over all the possible divergence models. For example, in our case $\Psi = 1$ and $\Psi = 22$ both represent a single divergence model, whereas $\Psi = 6$ comprises 136 models. Generally, we

can calculate the prior probability of the of the ith unique divergence model of class Ψ by

$$f(M_{\Psi,i}) = f(\Psi)\frac{1}{A(Y,\Psi)}, \tag{7}$$

where $f(\Psi)$ is the prior probability of Ψ, and $A(Y,\Psi)$ is the number of partitions of Y into Ψ divergence times (i.e., the number of unique divergence models for a given Ψ; Appendix 1; Sloan, 2011b). In the case of `msBayes`, the prior on Ψ is uniform discrete, so Equation eq:generalpriorprob simplifies to

$$f(M_{\Psi,i}) = \frac{1}{Y \times A(Y,\Psi)}. \tag{8}$$

The distribution of the number of divergence models across Ψ, and corresponding prior probability distribution of divergence time models is shown in **??** 5. In our case, the $M_{\Psi=1,i}$ model is 136 times more probable than a $M_{\Psi=6,i}$ model, *a priori*. The combination of the prior distribution over $M_{\Psi,i}$ (**??** 5B) and potentially small marginal likelihoods of models with large Ψ (discussed below) could create a strong bias towards models with small Ψ. An alternative prior that places a uniform probability across each possible divergence model, rather than across Ψ, would result in a lower posterior for the single divergence event scenario whenever $Y > 3$.

Decreasing marginal likelihoods with increasing Ψ

The preference for clustered divergence models (i.e., small Ψ) implies that the models with larger Ψ have lower marginal likelihoods. The marginal likelihood of a model is an integral over the entire parameter space of the likelihood weighted by the prior probability density. By using broad, uniform priors for each divergence time parameter, we force models with many distinct divergence times to integrate over a *much* larger parameter space. If most of the parameter space has low likelihood, the marginal likelihood will be small. For example, consider the comparison of the models with $\Psi = 1$ and $\Psi = 22$.

In our analyses, each of the 22 divergence time parameters of the $\Psi = 22$ model has a uniform prior from zero to 20 coalescent units. If most of the prior space of these divergence time parameters has low likelihood density, integrating over the vast 22-dimensional τ parameter space will result in a low marginal likelihood for the $\Psi = 22$ model compared to the model with only one τ parameter.

Bayesian parameter estimation is often relatively robust to the choice of prior, particularly if the prior is vague. But model selection in the Bayesian context can be strongly influenced by excessively broad priors on nuisance parameters (Lindley, 1957; Jeffreys, 1961). This sensitivity to priors should hold for fully Bayesian model selection and for ABC methods. It is not a weakness of the analysis paradigm, but merely an indication that the prior probability statements used for nuisance parameters must be carefully chosen.

If prior sensitivity in our analyses caused the bias toward small Ψ, then a simulation study in which all parameters are drawn from the prior distributions used in the analysis (Huelsenbeck and Rannala, 2004) should result in good performance with respect to the posterior probability of different values of Ψ. Our results demonstrate that when the priors are correct (i.e., there is no model misspecification) `msBayes` tends to be biased in the opposite direction and underestimates the posterior probability of one divergence event (Fig.4C–F). This suggests the bias toward clustered divergences when the prior on τ is broader than the true underlying distribution is caused, at least in part, by broad uniform priors reducing the marginal likelihoods of models with more τ parameters (i.e., larger Ψ).

However, our results suggest that, in practice, the uniform prior distributions may never be narrow enough to obviate the bias of `msBayes` toward clustered divergences. Even when we

tested uniform priors more informative than could be expected in practice, the bias remained (Figs. $\mathrm{fig}_a cc_{os} s_g lm, fig_p ow_p si_m odes_s s_g lm, fig_p ow_{os} s_g lm and S fig_a cc_{os} s_g lm_i nform 10, fig_p ow_p si_m odes_s s_g$ $normal) could better represent prior knowledge about \theta and \tau, while not placing too much prior probability den$

Insufficient sampling of parameter values

It is also possible that the estimates of the posterior probability for models with large Ψ will be inaccurate because of insufficient sampling of parameter values. If the prior over Ψ is uniform, models with larger prior parameter space (i.e., larger Ψ) will be less densely sampled. It is unclear if this phenomenon would bias the analysis toward models with fewer τ parameters or merely lead to higher variance in the estimates of the posterior probabilities. If this phenomenon is causing the bias, we expect analyses to be sensitive to the number of samples drawn from the prior. From our empirical analyses, we do not see such sensitivity when comparing prior sample sizes of 5×10^6 and 10^7. Also, our simulation results were unaffected by prior sample sizes of 2×10^6, 5×10^6, and 10^7 (see SI for full details). Thus, it seems unlikely that insufficient sampling of parameter space contributed to the bias.

Problems with ABC model choice

Recent work by Robert et al. (2011) has shown that ABC will often be biased in model choice. When summary statistics are insufficient for discriminating among competing models, which is the case for most empirical applications, ABC can be an inconsistent estimator of the models' posterior probabilities and can thus prefer the wrong model. This can occur even when the summary statistics are sufficient for each model under consideration Robert et al. (2011). The magnitude of the bias caused by the insufficiency of the statistics across models is unknown in most empirical settings, so there is no possibility of a correction factor (Robert et al., 2011). Thus, this problem does not have an obvious fix.

Whether the bias we observed is due to theoretical shortcomings of ABC model choice, broad uniform priors on nuisance parameters, the prior on divergence models, or some combination of these factors requires further investigation.

GENERAL RECOMMENDATIONS

Despite the limitations we identified, we conclude that `msBayes` can be a useful comparative phylogeographic tool. For example, `msBayes` can be used to test biogeographic hypotheses that require low temporal precision to be differentiated (i.e., on the scale of millions of generations). Alternatively, given the bias toward clustered divergences, an inference of no co-divergence (e.g., Topp and Winker, 2008) is likely robust. Also, there are several avenues to explore that might mitigate the problems we revealed, including (1) priors on $\boldsymbol{\tau}$ that are more uniform over divergence models (rather than classes of models), (2) more flexible priors on θ and τ parameters (e.g., gamma or log-normal distributions) that might increase the relative marginal likelihoods of models with more divergence times, and (3) adding more loci. The degree to which our results are contingent upon the number of taxon-pairs and sample sizes simulated here also needs to be explored. However, our results clearly demonstrate the need for (1) power analyses to accompany any inference of clustered divergences using `msBayes` and (2) analyses exploring the sensitivity of the results to prior distributions.

Conclusions

The hierarchical ABC model implemented in `msBayes` provides an appealing method of inferring the effects of historical events on diversification. Our goal was to use the model to test whether Pleistocene climate cycles and associated sea-level oscillations caused diversification across the islands of the Philippine Archipelago. Despite strong support for the recent simultaneous divergence of 22 pairs of populations across the Philippines, our simulation-based power analyses demonstrate that we were likely to infer such results. Our simulations show that `msBayes` will infer highly clustered divergence times when populations diverged randomly over the past $12N_C$ generations (7.5 million generations if $\mu = 2 \times 10^{-8}$), and will often ($> 5\%$ of replicates) infer the extreme case of one divergence event with high posterior probability (> 0.95) when divergences were random over the past $3.2N_C$ generations (2 million generations). For our empirical system,

this lack of power precludes us from ascribing biological processes to the results we obtained from `msBayes`. We also show that results of `msBayes` can be sensitive to prior distributions placed on parameters. We suggest any results of `msBayes` that are not shown to be robust to prior settings should be treated with caution. Furthermore, simulation-based power analyses should be used in cases in which clustered divergences are inferred. These analyses can provide useful guides for the range of divergence times which could have occurred while still being judged to be "simultaneous".

Acknowledgments

We thank L. Kubatko, M. Hickerson, and one anonymous reviewer for improving this article. We thank the Philippine Department of Environment and Natural Resources (DENR-Manila), in particular the Protected Areas and Wildlife Bureau (PAWB-Manila) for their steadfast support of our research program. Fieldwork followed approved University of Kansas Institutional Animal Care and Use Committee research protocols and administrative procedures established in a Memorandum of Agreement (MOA) between the University of Kansas and PAWB, as outlined in a Gratuitous Permit to Collect (GP) biological specimens, Nos. 185 and 201, administered by PAWB for 2007–2011. We thank T. M. Lim, J. de Leon, C. Custodio, A. Tagtag, M. Diesmos, and A. Diesmos for their logistical support of our fieldwork. Survey work was made possible by a U.S. National Science Foundation Biotic Surveys and Inventories grant (DEB 0743491) to RMB. JRO was supported by NSF Doctoral Dissertation Improvement grant DEB 1011423. Molecular data collected for this project was funded by NSF grants DEB 0640737 and DEB 0910341. Fieldwork for CDS was supported by NSF (DEB 0804115), Panorama Fund grant from The University of Kansas Biodiversity Institute, Madison and Lila Self Fellowship from the University of Kansas, and Fulbright and Fulbright-Hayes Fellowships. JAE was supported by NSF OISE 0965856. We also thank the Cincinnati Museum of Natural History, Field Museum of Natural History, National Museum of the Philippines, and Mindanao State

University, along with each institution's staff, for providing loans of specimens.

References

Barber, B. R. and J. Klicka, 2010. Two pulses of diversification across the Isthmus of Tehuantepec in a montane Mexican bird fauna. Proceedings Of The Royal Society B-Biological Sciences 277:2675–2681.

Beaumont, M., W. Zhang, and D. J. Balding, 2002. Approximate Bayesian computation in population genetics. Genetics 162:2025–2035.

Beaumont, M. A., 2010. Approximate Bayesian Computation in Evolution and Ecology. Annual Review of Ecology, Evolution, and Systematics 41:379–406.

Bell, E. T., 1934. Exponential numbers. Am Math Mon 41:411–419.

Bell, R. C., J. B. MacKenzie, M. J. Hickerson, K. L. Chavarria, M. Cunningham, S. Williams, and C. Moritz, 2012. Comparative multi-locus phylogeography confirms multiple vicariance events in co-distributed rainforest frogs. Proceedings Of The Royal Society B-Biological Sciences 279:991–999.

Bertorelle, G., A. Benazzo, and S. Mona, 2010. ABC as a flexible framework to estimate demography over space and time: some cons, many pros. Mol Ecol 19:2609–2625.

Blum, M. G. B. and O. François, 2009. Non-linear regression models for Approximate Bayesian Computation. Stat Comput 20:63–73.

Brown, R. M. and A. C. Diesmos, 2002. Application of lineage-based species concept to oceanic island frog populations: The effects of differing taxonomic philosophies on the estimation of Philippine biodiversity. Silliman Journal 42:133–162.

———, 2009. Philippines, biology. Pp. 723–732, *in* R. Gillespie and D. Clague, eds. Encyclopdia of Islands. University of California Press, Berkeley.

Brown, R. M., A. C. Diesmos, and A. C. Alcala, 2008. Philippine amphibian biodiversity is increasing in leaps and bounds. Pp. 82–83, *in* S. N. Stuart, M. Hoffmann, J. S. Chanson, N. A. Cox, R. Berridge, P. Ramani, and B. E. Young, eds. Threatened Amphibians of the World. Lynx Ediciones, Barcelona, Spain; IUCN—The World Conservation Union, Gland, Switzerland; and Conservation International, Arlington, Virginia, USA.

Brown, R. M. and B. L. Stuart, 2012. Patterns of biodiversity discovery through time: an historical analysis of amphibian species discoveries in the Southeast Asian mainland and island archipelagos. Pp. 348–389, *in* D. J. Gower, K. G. Johnson, J. E. Richardson, B. R. Rosen, L. Rüber, and S. T. Williams, eds. Biotic Evolution and Environmental Change in Southeast Asia. Cambridge University Press.

Burbrink, F. T., R. Lawson, and J. B. Slowinski, 2000. Mitochondrial DNA phylogeography of the polytypic North American Rat Snake (*Elaphe obsoleta*): A critique of the subspecies concept. Evolution 54:2107–2118.

Carlin, B. P. and A. E. Gelfand, 1990. Approaches for empirical Bayes confidence intervals. J Am Stat Assoc 85:105–114.

Carnaval, A. C., M. J. Hickerson, C. F. B. Haddad, M. T. Rodrigues, and C. Moritz, 2009. Stability Predicts Genetic Diversity in the Brazilian Atlantic Forest Hotspot. Science 323:785–789.

Chan, L. M., J. L. Brown, and A. D. Yoder, 2011. Integrating statistical genetic and geospatial methods brings new power to phylogeography. Mol Phylogenet Evol 59:523–537.

Csilléry, K., M. G. B. Blum, O. E. Gaggiotti, and O. François, 2010. Approximate Bayesian Computation (ABC) in practice. Trends In Ecology and Evolution 25:410–418.

Daza, J. M., T. A. Castoe, and C. L. Parkinson, 2010. Using regional comparative

phylogeographic data from snake lineages to infer historical processes in Middle America. Ecography 33:343–354.

Dickerson, R. E., 1928. Distribution of life in the Philippines. Philippine Bureau of Science, Manila, Philippines.

Drummond, A. J. and A. Rambaut, 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7:214.

Edgar, R. C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797.

Esselstyn, J. A., B. J. Evans, J. L. Sedlock, F. A. A. Khan, and L. R. Heaney, 2012. Single-locus species delimitation: A test of the mixed Yule-coalescent model, with an empirical application to Philippine round-leaf bats. Proceedings Of The Royal Society B-Biological Sciences 279:3678–3686.

Esselstyn, J. A., R. M. Timm, and R. M. Brown, 2009. Do geological or climatic processes drive speciation in dynamic archipelagos? the tempo and mode of diversification in Southeast Asian shrews. Evolution 63:2595–2610.

Evans, B., R. Brown, J. Mcguire, J. Supriatna, N. Andayani, A. Diesmos, D. Iskandar, D. Melnick, and D. Cannatella, 2003. Phylogenetics of fanged frogs: Testing biogeographical hypotheses at the interface of the Asian and Australian faunal zones. Syst Biol 52:794–819.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J Mol Evol 17:368–376.

Hall, R., 1998. The plate tectonics of Cenozoic SE Asia and the distribution of land and sea. Pp. 99–131, *in* R. Hall and J. D. Holloway, eds. Biogeography and geological evolution of SE Asia. Backhuys Publishers, Leiden.

Haq, B. U., J. Hardenbol, and P. R. Vail, 1987. Chronology of fluctuating sea levels since the Triassic. Science 235:1156–1167.

Hasegawa, M., H. Kishino, and T.-a. Yano, 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160–174.

Heaney, L. R., 1985. Zoogeographic evidence for middle and late pleistocene land bridges to the philippine islands. Mod Quatern Res SE Asia 9:127–144.

———, 1986. Biogeography of mammals in se asia: estimates of rates of colonization, extinction and speciation. Biol J Linn Soc 28:127–165.

Heaney, L. R. and J. C. Regalado, Jr., 1998. Vanishing treasures of the Philippine rain forest. Field Museum, Chicago, Illinois.

Hickerson, M. J., E. A. Stahl, and H. A. Lessios, 2006. Test for simultaneous divergence using approximate Bayesian computation. Evolution 60:2435–2453.

Huang, W., N. Takebayashi, Y. Qi, and M. J. Hickerson, 2011. MTML-msBayes: Approximate Bayesian comparative phylogeographic inference from multiple taxa and multiple loci with rate heterogeneity. BMC Bioinformatics 12:1.

Huelsenbeck, J. P. and B. Rannala, 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. Syst Biol 53:904–913.

Hwang, J. T. G., J. Qiu, and Z. Zhao, 2009. Empirical Bayes confidence intervals shrinking both means and variances. J Roy Stat Soc B 71:265–285.

Inger, R. F., 1954. Systematics and zoogeography of Philippine Amphibia. Fieldiana 33:182–531.

Jeffreys, H., 1935. Some tests of significance, treated by the theory of probability. Proceedings of the Cambridge Philosophical Society 31:203–222.

———, 1961. Theory of Probability. 3rd ed. Oxford University Press, Oxford, U.K.

Joyce, P. and P. Marjoram, 2008. Approximately sufficient statistics and Bayesian computation. Statistical Applications in Genetics and Molecular Biology 7:1–16.

Laird, N. M. and T. A. Louis, 1987. Empirical Bayes confidence intervals based on bootstrap samples. J Am Stat Assoc 82:739–750.

———, 1989. Empirical Bayes confidence intervals for a series of related experiments. Biometrics 45:481–495.

Lawson, L. P., 2010. The discordance of diversification: evolution in the tropical-montane frogs of the Eastern Arc Mountains of Tanzania. Mol Ecol 19:4046–4060.

Leaché, A. D., S. C. Crews, and M. J. Hickerson, 2007. Two waves of diversification in mammals and reptiles of Baja California revealed by hierarchical Bayesian analysis. Biology Letters 3:646–650.

Leuenberger, C. and D. Wegmann, 2010. Bayesian computation and model selection without likelihoods. Genetics 184:243–252.

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

Linkem, C. W., A. C. Diesmos, and R. M. Brown, 2011. Molecular systematics of the Philippine forest skinks (Squamata: Scincidae: *Sphenomorphus*): testing morphological hypotheses of interspecific relationships. Zool J Linn Soc-lond 163:1217–1243.

Linkem, C. W., K. M. Hesed, A. C. Diesmos, and R. M. Brown, 2010. Species boundaries and cryptic lineage diversity in a Philippine forest skink complex (Reptilia; Squamata; Scincidae: Lygosominae). Mol Phylogenet Evol 56:572–585.

Malenfant, J., 2011. Finite, closed-form expressions for the partition function and for Euler, Bernoulli, and Stirling numbers. arXiv.org arXiv:1103.1585v6 [math.NT].

Miller, K. G., M. A. Kominz, J. V. Browning, J. D. Wright, G. S. Mountain, M. E. Katz, P. J. Sugarman, B. S. Cramer, N. Christie-Blick, and S. F. Pekar, 2005. The Phanerozoic record of global sea-level change. Science 310:1293–1298.

Morris, C. N., 1983. Parametric empirical bayes inference: Theory and applications. J Am Stat Assoc 78:47–55.

Nei, M. and W.-H. Li, 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. P Natl Acad Sci Usa 76:5269–5273.

Plouviez, S., T. M. Shank, B. Faure, C. Daguin-Thiebaut, F. Viard, F. H. Lallier, and D. Jollivet, 2009. Comparative phylogeography among hydrothermal vent species along the East Pacific Rise reveals vicariant processes and population expansion in the South. Mol Ecol 18:3903–3917.

Posada, D. and K. A. Crandall, 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818.

Robert, C. P., J.-M. Cornuet, J.-M. Marin, and N. S. Pillai, 2011. Lack of confidence in approximate Bayesian computation model choice. P Natl Acad Sci Usa 108:15112–15117.

Roberts, T. E., 2006a. History, ocean channels, and distance determine phylogeographic patterns in three widespread Philippine fruit bats (Pteropodidae). Mol Ecol 15:2183–2199.

———, 2006b. Multiple levels of allopatric divergence in the endemic Philippine fruit bat *Haplonycteris fischeri* (Pteropodidae). Biol J Linn Soc 88:329–349.

Rohling, E. J., M. Fenton, F. J. Jorissen, P. Bertrand, G. Ganssen, and J. P. Caulet, 1998. Magnitudes of sea-level lowstands of the past 500,000 years. Nature 394:162–165.

Schwarz, G., 1978. Estimating the dimension of a model. Ann Stat 6:461–464.

Siddall, M., E. J. Rohling, A. Almogi-Labin, C. Hemleben, D. Meischner, I. Schmelzer, and D. A. Smeed, 2003. Sea-level fluctuations during the last glacial cycle. Nature 423:853–858.

Siler, C. D. and R. M. Brown, 2010. Phylogeny-based species delimitation in Philippine slender skinks (Reptilia: Squamata: Scincidae: *Brachymeles*): Taxonomic revision of pentadactyl species groups and description of three new species. Herpetol Monogr 24:1–54.

Siler, C. D., A. C. Diesmos, A. C. Alcala, and R. M. Brown, 2011a. Phylogeny of Philippine slender skinks (Scincidae: *Brachymeles*) reveals underestimated species diversity, complex biogeographical relationships, and cryptic patterns of lineage diversification. Mol Phylogenet Evol 59:53–65.

Siler, C. D., A. M. Fuiten, R. M. Jones, A. C. Alcala, and R. M. Brown, 2011b. Phylogeny-based species delimitation in Philippine slender skinks (Reptilia: Squamata: Scincidae) II: Taxonomic revision of *Brachymeles samarensis* and description of five new species. Herpetol Monogr 25:76–112.

Siler, C. D., J. R. Oaks, J. A. Esselstyn, A. C. Diesmos, and R. M. Brown, 2010. Phylogeny and biogeography of Philippine bent-toed geckos (Gekkonidae: *Cyrtodactylus*) contradict a prevailing model of Pleistocene diversification. Mol Phylogenet Evol 55:699–710.

Siler, C. D., J. R. Oaks, L. J. Welton, C. W. Linkem, J. Swab, A. C. Diesmos, and R. M. Brown, 2012. Did geckos ride the Palawan raft to the Philippines? J Biogeogr .

Sloan, N. J. A., 2011a. The On-Line Encyclopedia of Integer Sequences, Sequence A000041. Published electronically at http://oeis.org. URL `http://oeis.org/A000041`.

———, 2011b. The On-Line Encyclopedia of Integer Sequences, Sequence A008284. Published electronically at http://oeis.org. URL `http://oeis.org/A008284`.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Stamatakis, A., F. Blagojevic, D. Nikolopoulos, and C. Antonopoulos, 2007. Exploring new search algorithms and hardware for phylogenetics: RAxML meets the IBM cell. J Vlsi Signal Proc 48:271–286.

Stone, G. N., K. Lohse, J. A. Nicholls, P. Fuentes-Utrilla, F. Sinclair, K. Schönrogge, G. Csóka, G. Melika, J.-L. Nieves-Aldrey, J. Pujade-Villar, M. Tavakoli, R. R. Askew, and M. J. Hickerson, 2012. Reconstructing community assembly in time and space reveals enemy escape in a Western Palearctic insect community. Curr Biol 22:532–537.

Sukumaran, J. and M. T. Holder, 2010. DendroPy: A Python library for phylogenetic computing. Bioinformatics 26:1569–1571.

Swofford, D. L., 2003. PAUP*, phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, MA.

Tajima, F., 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437–460.

———, 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.

Takahata, N. and M. Nei, 1985. Gene genealogy and variance of interpopulational nucleotide differences. Genetics 110:325–344.

Topp, C. M. and K. Winker, 2008. Genetic patterns of differentiation among five landbird species from the Queen Charlotte Islands, British Columbia. Auk 125:461–472.

Voje, K. L., C. Hemp, Ø. Flagstad, G.-P. Saetre, and N. C. Stenseth, 2009. Climatic change as an engine for speciation in flightless Orthoptera species inhabiting African mountains. Mol Ecol 18:93–108.

Voris, H. K., 2000. Maps of the Pleistocene sea levels in South East Asia: Shorelines, river systems, time durations. J Biogeogr 27:1153–1167.

Watterson, G. A., 1975. On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–276.

Wegmann, D., C. Leuenberger, and L. Excoffier, 2009. Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. Genetics 182:1207–1218.

Wegmann, D., C. Leuenberger, S. Neuenschwander, and L. Excoffier, 2010. ABCtoolbox: a versatile toolkit for approximate bayesian computations. BMC Bioinformatics 11:116.

Welton, L. J., C. D. Siler, C. W. Linkem, A. C. Diesmos, and R. M. Brown, 2010. Philippine bent-toed geckos of the *Cyrtodactylus agusanensis* complex: Multilocus phylogeny, morphological diversity, and descriptions of three new species. Herpetol Monogr 24:55–85.

Yumul, G., C. Dimalanta, V. Maglambayan, and E. Marquez, 2008. Tectonic setting of a composite terrane: A review of the Philippine island arc system. Geosciences Journal 12:7–17.

Appendix 1

PYTHON FUNCTIONS FOR CALCULATING INTEGER PARTITION

```
def integer_partition(n):
    """
    Takes an integer (n) and returns the total number of partitions of n, and
    a dictionary of the number of partitions of n for each number of categories
    (k).
    """
    part_nums = {}
    for k in range(1, n + 1):
        part_nums[k] = A(n, k)
    return sum(part_nums.values()), part_nums

def A(n, k):
    """
    Takes two integers, (n) and the number of categories (k) to partition n,
    and returns the number of partitions of n into k categories.
    """
    if k == 1 or n == k:
        return 1
    if k > n or k < 1:
        return 0
    return A(n - 1, k - 1) + A(n - k, k)
```

Figure Captions

Figure 1. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and S_{stats}. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$.

Figure 2. Histograms of the estimated number of divergence events ($\hat{\Psi}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of inferring one divergence event, $p(\hat{\Psi} = 1)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and S_{stats}. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$.

Figure 3. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and S_{stats}. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$.

Figure 4. The relationship between the posterior and true probability of (A, C & E) $\Psi = 1$ and (B, D & F) $\Omega < 0.01$ based on 100,000 simulations. The results are based on the (A & B) unadjusted ($\mathcal{P}_\epsilon$), (C & D) ABC_{LLR}-adjusted, and (E & F) ABC_{GLM}-adjusted posterior estimate from each simulation replicate. All simulated replicates were generated under the model prior (i.e., the ideal situation where the prior model is correct). Prior settings were $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0001, 0.05)$, and $\theta_A \sim U(0.0001, 0.025)$, and the number of samples from the prior was 2×10^6. The simulated data structure was 10 population pairs, with a single 1000 bp locus sampled from 10 individuals from each population. The 100,000 estimates of the posterior probability of one divergence event were assigned to 20 bins of width 0.05. The estimated posterior probability of each bin is plotted against the proportion of replicates in that bin with a true value consistent with one divergence event (i.e., $\Psi = 1$ or $\Omega < 0.01$).

Figure 5. The (A) number of unique divergence models across all possible numbers of divergence time categories (Ψ) when $Y = 22$, and (B) the prior probability of any one of the unique divergence time models within each Ψ.

Table 1. Summary of the notation used throughout this work.

Symbol	Description
Y	Number of population pairs.
k_i	Number of loci sampled from population pair i.
K	Total number of unique loci sampled.
$X_{i,j}$	Sequence alignment of locus j sampled from population pair i.
$S^*_{i,j}$	Population genetic summary statistics calculated from $X_{i,j}$.
$\mathbf{X}$	Vector containing the sequence alignments of each locus from each population pair: $\{X_{1,1}, \ldots, X_{Y,k_Y}\}$.
$\mathbf{S}^*$	Vector containing the summary statistics of each locus from each population pair: $\{S^*_{1,1}, \ldots, S^*_{Y,k_Y}\}$.
$B_\epsilon(\mathbf{S}^*)$	Multi-dimensional Euclidean space around the observed summary statistics, $\mathbf{S}^*$.
ϵ	Radius of $B_\epsilon(\mathbf{S}^*)$, i.e., the tolerance of the ABC estimation.
$G_{i,j}$	Gene tree of the sequences in $X_{i,j}$.
$\mathbf{G}$	Vector containing the gene trees of each locus from each population pair: $\{G_{1,1}, \ldots, G_{Y,k_Y}\}$.
Ψ	Number of unique population divergence times among the Y population pairs.
T	One of the Ψ unique divergence times in $4N_C$ generations.
$\mathbf{T}$	Vector of Ψ unique divergence times: $\{T_1, \ldots, T_\Psi\}$.
τ_i	Time of divergence in $4N_C$ generations between the populations of pair i.
$t_{i,j}$	Scaled time of divergence between the populations of pair i for locus j.
τ	Vector containing the divergence times for each population pair: $\{\tau_1, \ldots, \tau_Y\}$.
$\mathbf{t}$	Vector containing the scaled divergence times of each locus from each population pair: $\{t_{1,1}, \ldots, t_{Y,k_Y}\}$.
$\theta_{D1,i}$, $\theta_{D2,i}$	θ of the 1^{st} and 2^{nd} descendent population, respectively, of pair i.
$\theta_{A,i}$	θ of the population ancestral to pair i.
$\theta_{D\mathbf{1}}$, $\theta_{D\mathbf{2}}$	Vectors $\{\theta_{D1,1}, \ldots, \theta_{D1,Y}\}$ and $\{\theta_{D2,1}, \ldots, \theta_{D2,Y}\}$, respectively.
θ_A	Vector containing the θ_A parameters for each population pair: $\{\theta_{A,1}, \ldots, \theta_{A,Y}\}$.
υ_j	θ-scaling parameter for locus j.
υ	Vector containing the θ-scaling parameters for each locus: $\{\upsilon_1, \ldots, \upsilon_K\}$.
α	Hyper-parameter for the shape of the gamma prior distribution on each υ.
$\zeta_{D1,i}$, $\zeta_{D2,i}$	θ-scaling parameters that determine the magnitude of the population bottleneck in the 1^{st} and 2^{nd} descendant population of pair i, respectively. The bottleneck in each descendant population begins immediately after divergence.
$\zeta_{D\mathbf{1}}$, $\zeta_{D\mathbf{2}}$	Vectors $\{\zeta_{D1,1}, \ldots, \zeta_{D1,Y}\}$ and $\{\zeta_{D2,1}, \ldots, \zeta_{D2,Y}\}$, respectively.
$\tau_{B,i}$	Proportion of time between present and τ_i when the bottleneck ends for the descendant populations of pair i.
τ_B	Vector containing the τ_B parameters for each population pair: $\{\tau_{B,1}, \ldots, \tau_{B,Y}\}$.
m_i	Symmetric migration rate between the descendant populations of pair i.
$\mathbf{m}$	Vector containing the migration rates for each population pair: $\{m_1, \ldots, m_Y\}$.
$\rho_{i,j}$	θ-scaling constant provided by the user for locus j of pair i. This constant is intended to allow the user to scale θ for differences in ploidy among loci or differences in generation times among taxa.
$\nu_{i,j}$	θ-scaling constant provided by the user for locus j of pair i. This constant is intended to allow the user to scale θ for differences in mutation rates among loci or among taxa.
ρ	Vector of ploidy and/or generation time scaling constants: $\{\rho_{1,1}, \ldots, \rho_{Y,k_Y}\}$.
ν	Vector of mutation-rate scaling constants: $\{\nu_{1,1}, \ldots, \nu_{Y,k_Y}\}$.
$E(\tau)$	Mean of τ.
$Var(\tau)$	Variance of τ.
Ω	Dispersion index of τ ($Var(\tau)/E(\tau)$).
Θ	Vector containing all the parameters of the model implemented in `msBayes`.
$f(\Theta)$	Joint prior probability distribution of the `msBayes` model.
n	Number of samples from the joint prior, $f(\Theta)$.
Λ	Vector containing the following summary of Θ: $\{\Psi, E(\tau), Var(\tau), \Omega\}$.
$\mathbf{S}$	Vector containing the summary statistics calculated from data simulated under Θ.
$\mathcal{P}_{f(\Theta)}$	Random sample of $\Lambda_1, \ldots, \Lambda_n$ from $f(\Theta)$.
$\mathbb{S}$	Summary statistic vectors $\mathbf{S_1}, \ldots, \mathbf{S_n}$ for each $\Lambda_1, \ldots, \Lambda_n$ within $\mathcal{P}_{f(\Theta)}$.
$\mathcal{P}_\epsilon$	Samples retained from $\mathcal{P}_{f(\Theta)}$ after rejection sampling. I.e., the Λ's from $\mathcal{P}_{f(\Theta)}$ with $\mathbf{S}$'s that fall within $B_\epsilon(\mathbf{S}^*)$.
$\mathcal{P}_{f(\Theta\|B_\epsilon(\mathbf{S}^*))}$	The estimate of the approximate posterior, $f(\Theta\|B_\epsilon(\mathbf{S}^*))$. I.e., the regression-adjusted $\mathcal{P}_\epsilon$.

Table 2. Summaries of the posterior estimates from all msBayes analyses run with prior settings of $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$. Mode estimates of the dispersion index (Ω) and mean ($E(\tau)$) of divergence time vector, $\boldsymbol{\tau}$, and the number of unique divergence times (Ψ) are provided, followed in square brackets by the 95% highest posterior density for ABC_{GLM} analyses and the 2.5% and 97.5% quantiles for the ABC_{LLR} analyses. The estimated posterior probability for the one divergence model ($p(\Psi = 1|B_\epsilon(\mathbf{S}^*))$) is given, followed in parentheses by the Bayes factor of the $\Psi = 1$ model compared to all other models. Ω and $E(\tau)$ are in units of N_C generations.

Prior sample	Analysis	$\hat{\Omega}$	$\hat{E(\tau)}$	$\hat{\Psi}$	$p(\Psi = 1\|B_\epsilon(\mathbf{S}^*))$
$\mathbb{S}_{stats,N=1\times10^7}$	ABC_{LLR}	0.0 [0.0–1.84×10^{-4}]	0.064 [0.043–0.092]	1.0 [1.00–1.00]	1.000 (∞)[1]
$\mathbb{S}_{stats,N=1\times10^7}$	ABC_{GLM}	1.29×10^{-4} [-1.08×10^{-6}–1.18×10^{-3}]	0.063 [0.039–0.092]	1.0 [1.00–1.06]	1.000 (∞)[1]
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{GLM}	-8.36×10^{-17} [-5.49×10^{-6}–5.89×10^{-3}]	0.062 [0.035–0.092]	1.0 [1.00–1.06]	1.000 (∞)[1]
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{GLM}	6.53×10^{-4} [-2.18×10^{-5}–1.89×10^{-3}]	0.071 [0.042–0.103]	1.0 [1.00–1.06]	0.999 (2.1×10^4)
$\mathbb{S}_{PLS,N=1\times10^7}$	ABC_{LLR}	0.0 [0.0–6.38×10^{-4}]	0.084 [0.036–0.136]	1.0 [1.00–1.00]	0.993 (2979)
$\mathbb{S}_{PLS,N=1\times10^7}$	ABC_{GLM}	-9.80×10^{-17} [-7.56×10^{-6}–7.59×10^{-3}]	0.082 [0.024–0.127]	1.0 [1.00–1.14]	0.977 (892)
$\mathbb{S}_{PLS,N=5\times10^6}$	ABC_{GLM}	-9.09×10^{-17} [-1.11×10^{-5}–1.31×10^{-2}]	0.064 [0.013–0.126]	1.0 [1.00–1.13]	0.992 (2604)
$\mathbb{S}_{PLS,N=5\times10^6}$	ABC_{GLM}	-9.80×10^{-17} [-7.64×10^{-6}–8.26×10^{-3}]	0.086 [0.029–0.143]	1.0 [1.00–1.08]	0.966 (597)
13 Mindanao population-pairs					
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{LLR}	2.56×10^{-3} [0.0–6.47×10^{-3}]	0.079 [0.034–0.122]	1.0 [1.00–1.00]	0.987 (1594)
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{GLM}	-8.03×10^{-17} [-6.00×10^{-6}–1.36×10^{-2}]	0.070 [0.027–0.115]	1.0 [1.00–1.08]	0.962 (532)
9 Negros-Panay population-pairs					
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{LLR}	0.0 [0.0–1.08×10^{-2}]	0.060 [0.012–0.099]	1.0 [1.00–1.00]	1.000 (∞)[1]
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{GLM}	-5.96×10^{-17} [-5.14×10^{-5}–4.53×10^{-2}]	0.055 [0.014–0.095]	1.0 [1.00–1.12]	0.999 (2.1×10^4)

[1] An estimate of 1.0 for a posterior probability is an artifact of sampling error.

Table 3. Posterior estimates of `msBayes` from the empirical data under three different model priors. Mode estimates of the dispersion index (Ω) and mean ($E(\tau)$) of divergence time vector, $\boldsymbol{\tau}$, and the number of unique divergence times (Ψ) are provided, followed in square brackets by the 95% highest posterior density for ABC_{GLM} analyses and the 2.5% and 97.5% quantiles for the ABC_{LLR} analyses. The estimated posterior probability for the inferred divergence model ($p(\Psi = \hat{\Psi}|B_\epsilon(\mathbf{S}^*))$) is given, followed in parentheses by the Bayes factor of the inferred model compared to all other models. The estimated posterior probability that Ψ is not one ($p(\Psi \neq 1|B_\epsilon(\mathbf{S}^*))$) is also given. Ω and $E(\tau)$ are in units of N_C generations.

Prior sample	**Analysis**	$\hat{\Omega}$	$\hat{E(\tau)}$	$\hat{\Psi}$	$p(\Psi = \hat{\Psi}\|B_\epsilon(\mathbf{S}^*))$	$p(\Psi \neq 1\|B_\epsilon(\mathbf{S}^*))$
$f(\Theta)$: $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$						
$\mathbb{S}_{stats,N=1\times10^7}$	ABC_{LLR}	0.0 [0.0–1.84$\times10^{-4}$]	0.064 [0.043–0.092]	1.00 [1.00–1.00]	0.992 (2604)	0.008
$\mathbb{S}_{stats,N=1\times10^7}$	ABC_{GLM}	1.29$\times10^{-4}$ [-1.08$\times10^{-6}$–1.18$\times10^{-3}$]	0.063 [0.039–0.092]	1.00 [1.00–1.06]	1.000 (∞)[1]	0.000
$f(\Theta)$: $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$						
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{LLR}	0.217 [0.112–0.381]	0.391 [0.282–0.485]	2.00 [2.00–2.00]	0.986 (1479)	0.986
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{GLM}	0.249 [0.030–0.475]	0.399 [0.271–0.513]	2.01 [1.39–3.40]	0.785 (77)	0.970
$f(\Theta)$: $\tau \sim U(0, 5)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$						
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{LLR}	0.421 [0.200–0.639]	0.465 [0.336–0.566]	2.00 [2.00–4.00]	0.784 (76)	0.999
$\mathbb{S}_{stats,N=5\times10^6}$	ABC_{GLM}	0.441 [0.190–0.711]	0.513 [0.358–0.601]	2.04 [1.51–5.16]	0.343 (11)	0.986

[1] An estimate of 1.0 for a posterior probability is an artifact of sampling error.

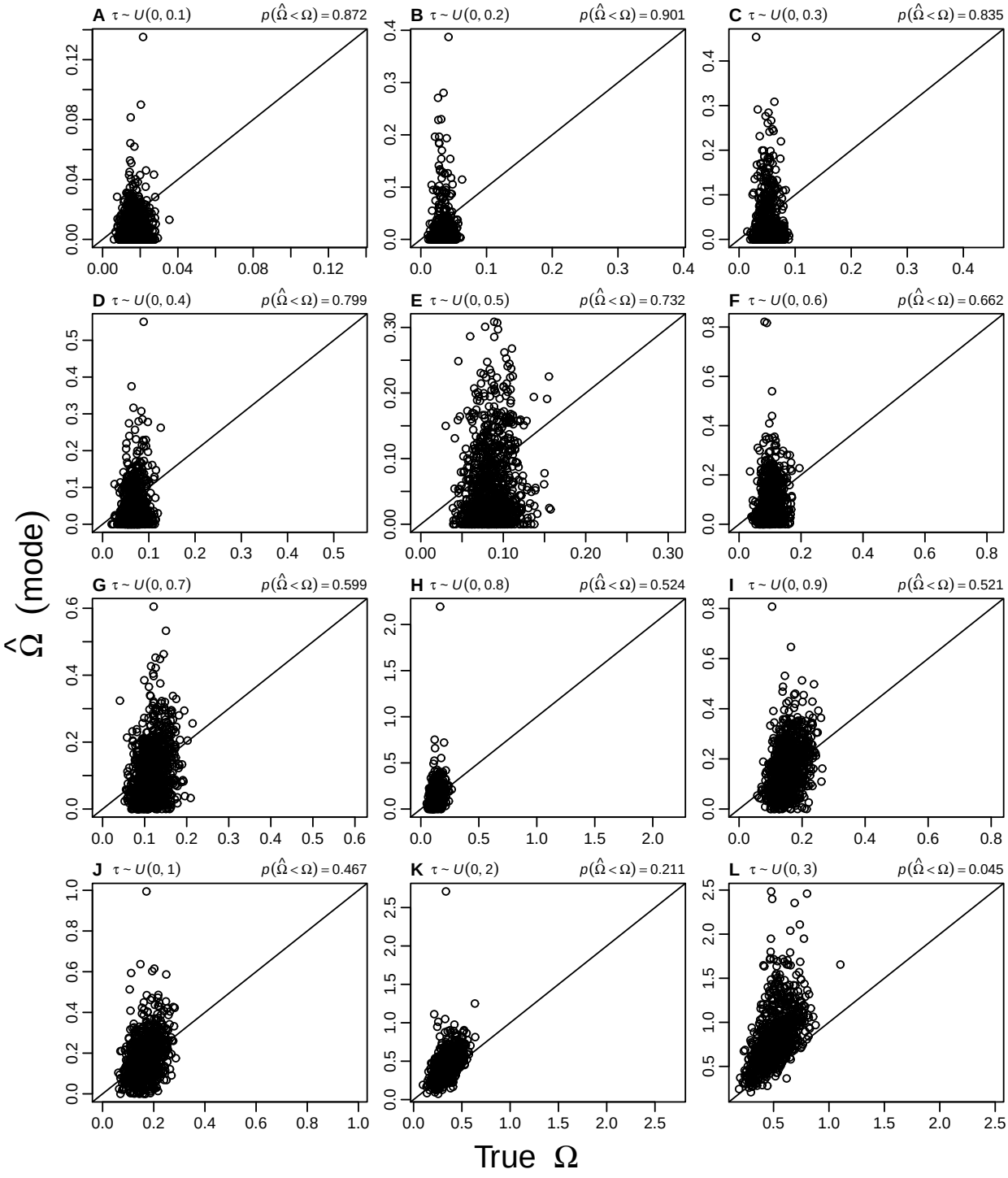

Figure 1. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$.

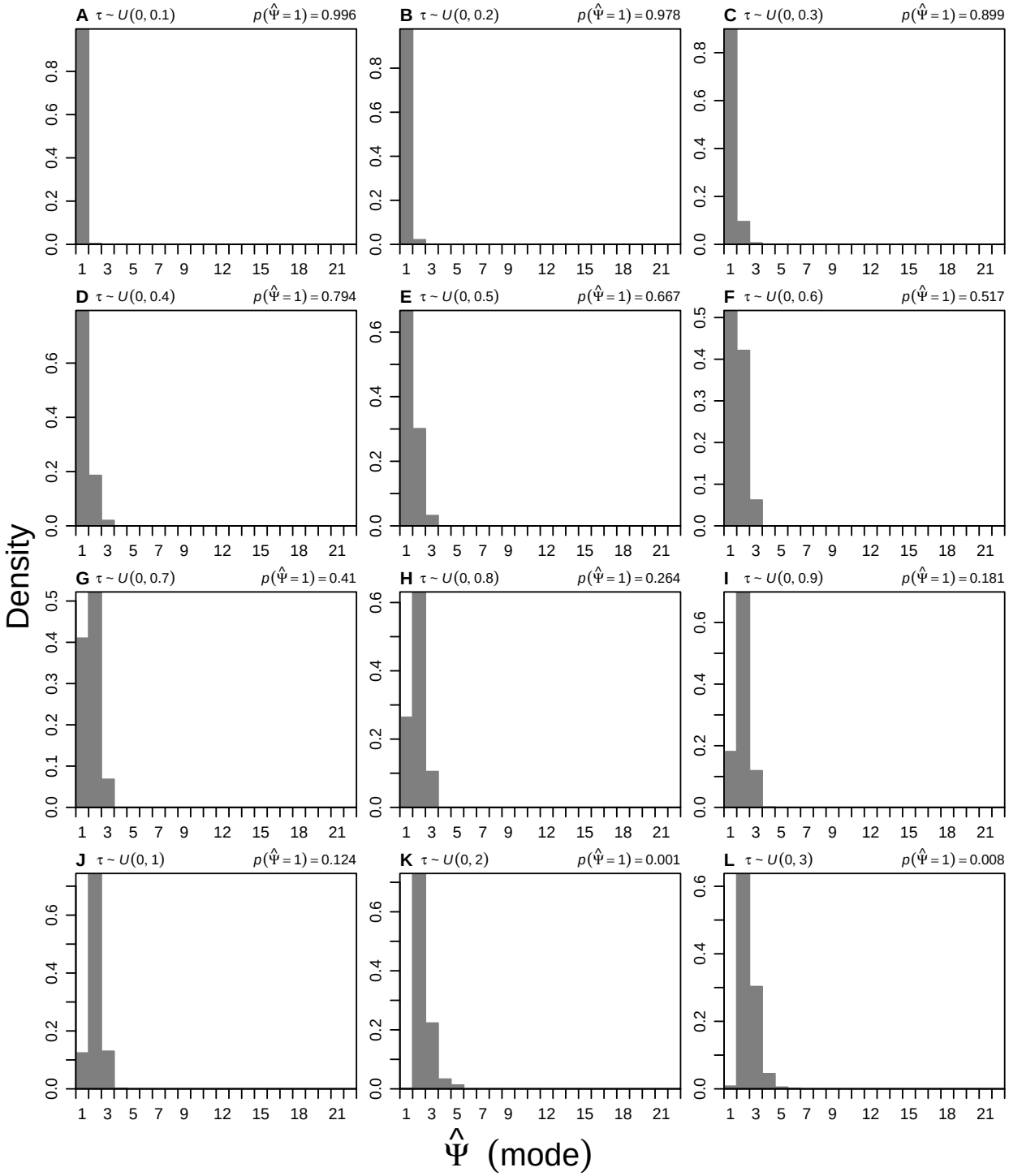

Figure 2. Histograms of the estimated number of divergence events ($\hat{\Psi}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of inferring one divergence event, $p(\hat{\Psi} = 1)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$.

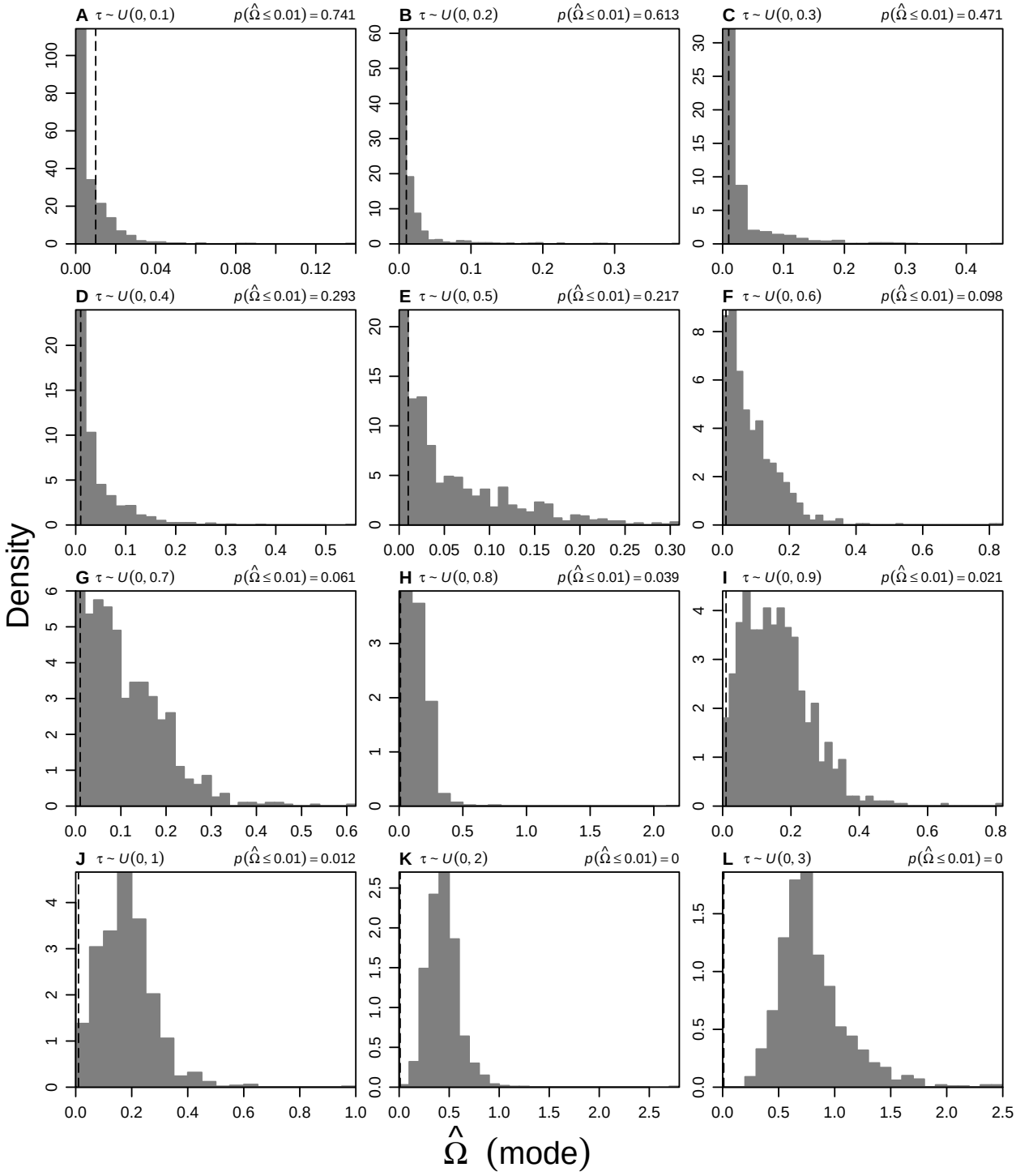

Figure 3. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$.

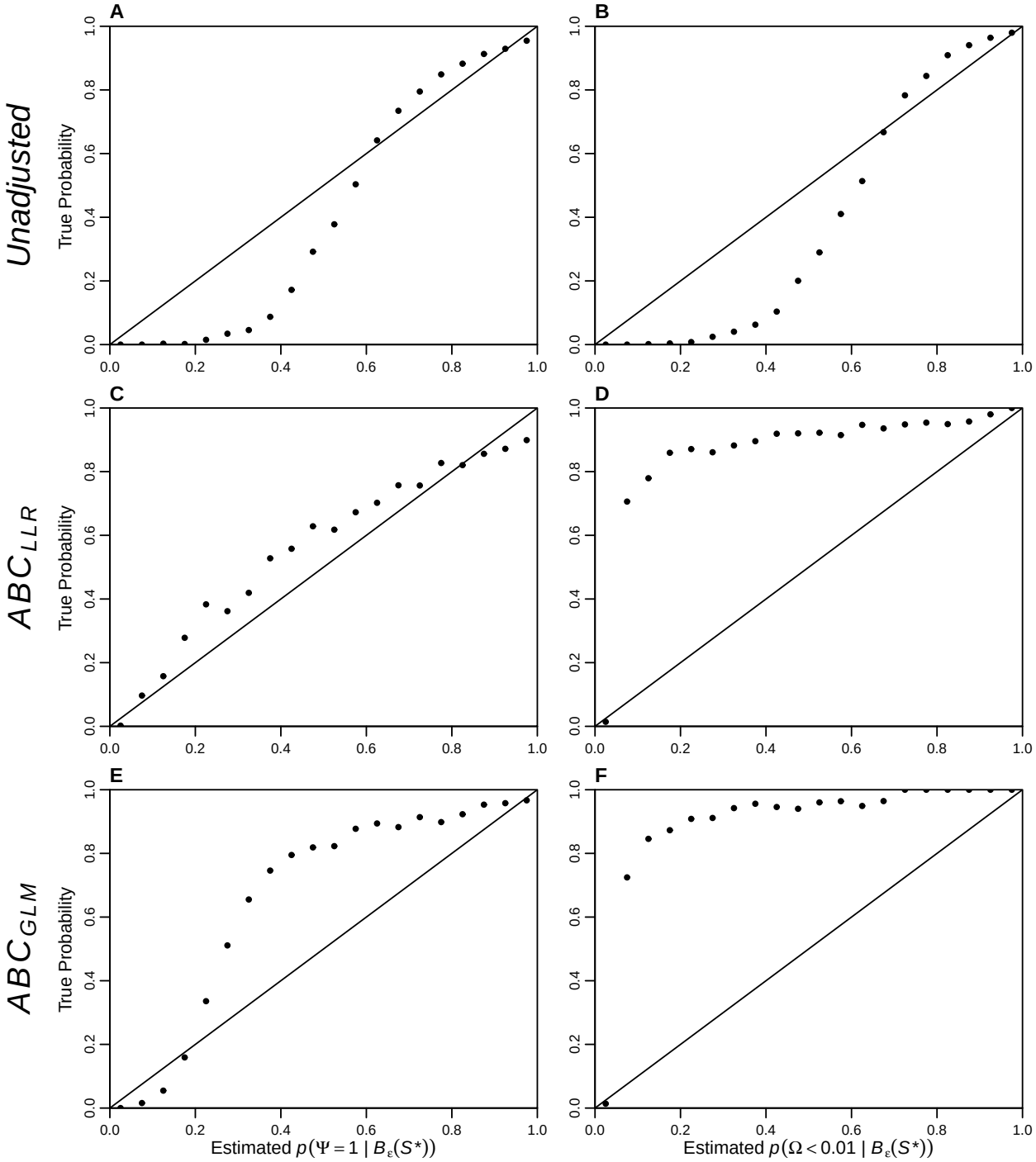

Figure 4. The relationship between the posterior and true probability of (A, C & E) $\Psi = 1$ and (B, D & F) $\Omega < 0.01$ based on 100,000 simulations. The results are based on the (A & B) unadjusted ($\mathcal{P}_\epsilon$), (C & D) ABC_{LLR}-adjusted, and (E & F) ABC_{GLM}-adjusted posterior estimate from each simulation replicate. All simulated replicates were generated under the model prior (i.e., the ideal situation where the prior model is correct). Prior settings were $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0001, 0.05)$, and $\theta_A \sim U(0.0001, 0.025)$, and the number of samples from the prior was 2×10^6. The simulated data structure was 10 population pairs, with a single 1000 bp locus sampled from 10 individuals from each population. The 100,000 estimates of the posterior probability of one divergence event were assigned to 20 bins of width 0.05. The estimated posterior probability of each bin is plotted against the proportion of replicates in that bin with a true value consistent with one divergence event (i.e., $\Psi = 1$ or $\Omega < 0.01$).

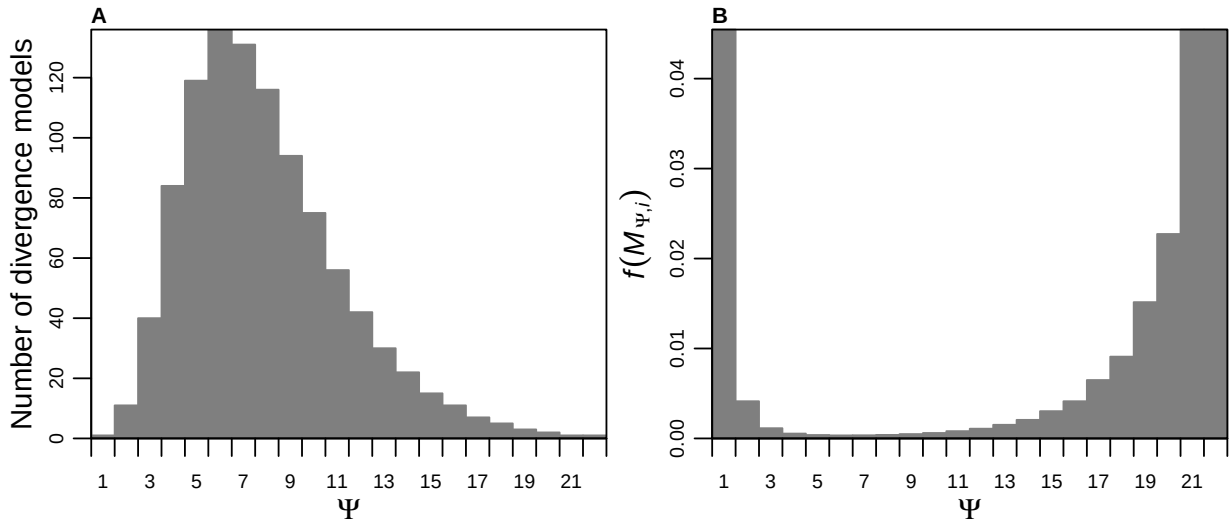

Figure 5. The (A) number of unique divergence models across all possible numbers of divergence time categories (Ψ) when $Y = 22$, and (B) the prior probability of any one of the unique divergence time models within each Ψ.

SI Table Captions

Table S1. Summary of the data collected from the 22 population pairs from the Greater Mindanao and Negros-Panay Pleistocene Aggregate Island Complexes (PAICs). On the map, dark gray represents the extent of land at current sea-level, and light gray is the extent when sea-level is 100m lower (data from ETOPO2; `http://www.ngdc.noaa.gov/mgg/global/etopo2.html`).

Table S2. Estimates of the population mutation rate (θ) from different island populations of the same species used in the `msBayes` analyses and/or closely related species.

Table S3. Estimates of the population mutation rate (θ) for the island populations used in the `msBayes` analyses.

Table S4. Summaries of the posterior estimates from all `msBayes` analyses run with prior settings of $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$. Mode estimates of the dispersion index (Ω) and mean ($E(\tau)$) of divergence time vector, $\boldsymbol{\tau}$, and the number of unique divergence times (Ψ) are provided, followed in square brackets by the 95% highest posterior density for ABC_{GLM} analyses and the 2.5% and 97.5% quantiles for the ABC_{LLR} analyses. The estimated posterior probability for the one divergence model ($p(\Psi = 1|B_\epsilon(\mathbf{S}^*))$) is given, followed in parentheses by the Bayes factor of the $\Psi = 1$ model compared to all other models. Ω and $E(\tau)$ are in units of N_C generations.

Table S5. Posterior estimates of `msBayes` from the empirical data under three different model priors, before and after the θ_A error in the source code was corrected. All results are from analyses using $\mathbb{S}_{stats}$ and ABC_{GLM}. Analyses with broad priors used 10^7 samples from the prior, whereas the analyses with the informed prior settings used 5×10^6 samples. Mode posterior estimates of Ω, $E(\tau)$, and Ψ are provided, followed in square brackets by the 95% highest posterior density. The estimated posterior probability for the inferred divergence model ($p(\Psi = \hat{\Psi}|B_\epsilon(\mathbf{S}^*))$) is given, followed in parentheses by the Bayes factor of the inferred model compared to all other models. The estimated posterior probability that Ψ is not one ($p(\Psi \neq 1|B_\epsilon(\mathbf{S}^*))$) is also given. Ω and $E(\tau)$ are in units of N_C generations.

Table 6. List of all data used in this study.

SI Figure Captions

Figure S1. Plot of the mean and 95% highest posterior density of gene divergence times estimated for each of the 22 population pairs using a fixed rate of 2×10^{-8} substitutions per site per year in BEAST.

Figure S2. The estimated joint posterior densities of the dispersion index (Ω) and mean ($E(\tau)$) of divergence time vector, $\boldsymbol{\tau}$, using (A & B) ABC_{LLR} and (C & D) ABC_{GLM} regression methods, and (A & C) the `msBayes` summary statistics and (B & D) partial least squares (PLS) components of the `msBayes` summary statistics. Ω and $E(\tau)$ are in units of N_C generations. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$.

Figure S3. Accuracy and precision of $E(\tau)$ estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{E(\tau)} < E(\tau))$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$.

Figure S4. Histograms of the true dispersion index of divergence times (Ω) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the proportion of true Ω values consistent with one divergence event, $p(\Omega \leq 0.01)$, is given for each τ_{max}.

Figure S5. Histograms of the estimated posterior probability of one divergence event, $p(\Psi = 1|B_\epsilon(\mathbf{S}^*))$, from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of inferring one divergence event with a Bayes factor greater than 10 (dashed black line), $p(BF_{\Psi=1,\Psi\neq1} > 10)$, is given for each τ_{max}. The red line indicates $p(\Psi = 1|B_\epsilon(\mathbf{S}^*)) = 0.95$, and the estimated probability of inferring a posterior probability greater than 0.95 is given to the right of the line.. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.0001, 0.05)$.

Figure S6. Results of 1000 simulation replicates of no divergence between population pairs, using $\mathbb{S}_{stats}$, ABC_{GLM}, and 5×10^6 samples from the prior. The posterior mode estimate of the mean of the divergence time vector, $E(\tau)$ (in $4N_C$ generations), is plotted for each replicate, and the observed estimate from the empirical data is indicated by the dotted line. The proportion of the simulation replicates that exceeded the observed empirical estimate is also given.

Figure S7. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$.

Figure S8. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 5)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$.

Figure S9. Histograms of the estimated number of divergence events ($\hat{\Psi}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of inferring one divergence event, $p(\hat{\Psi} = 1)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$.

Figure S10. Histograms of the estimated number of divergence events ($\hat{\Psi}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of inferring one divergence event, $p(\hat{\Psi} = 1)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 5)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$.

Figure S11. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$.

Figure S12. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 5)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.0005, 0.02)$.

Figure S13. Accuracy and precision of $E(\tau)$ estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{E(\tau)} < E(\tau))$) is given for each τ_{max}. All estimates were obtained using ABC_{LLR} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 2×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S14. Accuracy and precision of $E(\tau)$ estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{E(\tau)} < E(\tau))$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 2×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S15. Accuracy and precision of $E(\tau)$ estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{E(\tau)} < E(\tau))$) is given for each τ_{max}. All estimates were obtained using ABC_{LLR} and $\mathbb{S}_{PLS}$. Each plot represents 1000 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S16. Accuracy and precision of $E(\tau)$ estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{E(\tau)} < E(\tau))$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{PLS}$. Each plot represents 1000 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S17. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{LLR} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 2×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S18. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 2×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S19. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{LLR} and $\mathbb{S}_{PLS}$. Each plot represents 1000 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S20. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{PLS}$. Each plot represents 1000 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S21. Histograms of the estimated number of divergence events ($\hat{\Psi}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of inferring one divergence event, $p(\hat{\Psi} = 1)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 2×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S22. Histograms of the estimated number of divergence events ($\hat{\Psi}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of inferring one divergence event, $p(\hat{\Psi} = 1)$, is given for each τ_{max}. All estimates were obtained using ABC_{LLR} and $\mathbb{S}_{PLS}$. Note, the logistic regression adjustment provided with `msBayes` failed for approximately 2-30% of the simulations. The failure rate was positively correlated with the number of posterior samples with $\Psi = 1$. Thus, these plots are likely biased towards larger Ψ. Each plot represents 1000 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S23. Histograms of the estimated number of divergence events ($\hat{\Psi}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of inferring one divergence event, $p(\hat{\Psi} = 1)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{PLS}$. Each plot represents 1000 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S24. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{LLR} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 2×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S25. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 1000 simulation replicates using the same 2×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S26. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{LLR} and $\mathbb{S}_{PLS}$. Each plot represents 1000 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S27. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{PLS}$. Each plot represents 1000 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S28. Accuracy and precision of $E(\tau)$ estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{E(\tau)} < E(\tau))$) is given for each τ_{max}. All estimates were obtained using ABC_{LLR} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S29. Accuracy and precision of $E(\tau)$ estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{E(\tau)} < E(\tau))$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S30. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{LLR} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S31. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S32. Histograms of the estimated number of divergence events ($\hat{\Psi}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of inferring one divergence event, $p(\hat{\Psi} = 1)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S33. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{LLR} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S34. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 10^7 samples from the prior. Prior settings were $\tau \sim U(0, 20)$, $\theta_D \sim U(0.0001, 0.1)$, and $\theta_A \sim U(0.01, 0.05)$.

Figure S35. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.01, 0.02)$.

Figure S36. Accuracy and precision of Ω estimates from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The proportion of estimates less than the true value ($p(\hat{\Omega} < \Omega)$) is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 5)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.01, 0.02)$.

Figure S37. Histograms of the estimated number of divergence events ($\hat{\Psi}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of inferring one divergence event, $p(\hat{\Psi} = 1)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.01, 0.02)$.

Figure S38. Histograms of the estimated number of divergence events ($\hat{\Psi}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The estimated probability of inferring one divergence event, $p(\hat{\Psi} = 1)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 5)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.01, 0.02)$.

Figure S39. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.01, 0.02)$.

Figure S40. Histograms of the estimated dispersion index of divergence times ($\hat{\Omega}$) from simulations where τ (in $4N_C$ generations) for 22 population pairs is drawn from a series of uniform distributions, $\tau \sim U(0, \tau_{max})$. The threshold for one divergence event (Hickerson et al., 2006) is indicated by the dashed line, and the estimated probability of inferring one divergence event, $p(\hat{\Omega} \leq 0.01)$, is given for each τ_{max}. All estimates were obtained using ABC_{GLM} and $\mathbb{S}_{stats}$. Each plot represents 500 simulation replicates using the same 5×10^6 samples from the prior. Prior settings were $\tau \sim U(0, 5)$, $\theta_D \sim U(0.0005, 0.04)$, and $\theta_A \sim U(0.01, 0.02)$.

Figure S41. The relationship between the posterior and true probability of (A) $\Psi = 1$ and (B) $\Omega < 0.01$ based on 100,000 simulations. The results are based on the unadjusted posterior sample ($\mathcal{P}_\epsilon$) from each simulation replicate. All simulated replicates were generated under the model prior (i.e., the ideal situation where the prior model is correct). Prior settings were $\tau \sim U(0, 10)$, $\theta_D \sim U(0.0001, 0.05)$, and $\theta_A \sim U(0.01, 0.025)$, and the number of samples from the prior was 2×10^6. The simulated data structure was 10 population pairs, with a single 1000 bp locus sampled from 10 individuals from each population. The 100,000 estimates of the posterior probability of one divergence event were assigned to 20 bins of width 0.05. The estimated $p(\Psi = 1|B_\epsilon(\mathbf{S}^*))$ of each bin is plotted against the proportion of replicates in that bin with a true value of $\Psi = 1$.