**Supplementary Material: Exact Bayesian Inference for Phylogenetic Birth-Death Models**

Kris V Parag and Oliver G Pybus

**S1. The Sampled Birth-Death Process**

Throughout the main text we assumed complete and isochronous sampling, which means that all extant lineages at observation time $T$ appear in the rBDP with certainty. If the probability of any extant lineage at $T$ being observed in the rBDP is $\nu < 1$, then sampling is still isochronous, but is now also incomplete. Other incomplete sampling protocols also exist and may include extant lineages from the full rBDP randomly with time [Stadler 2009], or may choose lineages according to non-random criteria [Hohna 2014]. Although solving the BDP inference problem under such sampling protocols is outside the scope of our work, we do show how incomplete, isochronous sampling can be easily handled using the SF.

The sampling fraction $\nu$ acts as an additional determinant of which lineages we will observe in the sampled rBDP (deaths are the other controlling factor). Consequently, the rate of producing observed rBDP lineages no longer depends on just $P(t, T)$ (see the Methods Section in the main text). Nee et al. [1994] noted that $P(t, T)$ simply needs to be replaced with a sampled version, $P_s(t, T)$. For the time-varying BDP this was derived by Morlon et al. [2011] in reverse time. We reproduce their function in forward time using the transformations described in the supporting information of their work. This is given in Eq. (S1) with $B(t, T) = \int_{t}^{T} \lambda(s)e^{\alpha(s, t)} ds$. $P_s(t, T) = \frac{e^{\alpha(t, T)}}{\nu - 1 + B(t, T)}$ (S1)

The sampled lineage survival function $P_s(t, T)$ falls with $\nu$, and at very small $\nu$, $P_s(t, T) \propto \nu$. Setting $\nu = 1$ recovers an equivalent definition for $P(t, T)$, given in Kubo and Iwasa [1995]. For the constant rate BDP, Eq. (S1) reduces to the well known expression $P_s(t, T) = e^{(\lambda(1 - \nu) - \mu)(T - t)}$, first given in Nee et al. [1994].

The sampled rBDP rate is then $\beta_s(t) = \lambda(t)P_s(t, T)\mathcal{F}(t)$. This rate, which controls the speed of producing informative events, still describes a self-exciting Markov process. The usual SF algorithm (Eq (1)-(3) in main text) is therefore still valid for inference under this sampling model. Further, if $\nu$ is known, then there is no methodological difference between the SF for the incomplete and completely sampled cases, as the parameter space is unchanged. Note that if $\nu$ is not known then it cannot be estimated simultaneously with $\tilde{x}$ from $\mathcal{F}$, and more information is required [Stadler 2009].

We now investigate how incomplete sampling reduces the rBDP rate. We can manipulate Eq. (S1) to show that $P(t, T) \geq P_s(t, T) \geq \nu P(t, T)$. This leads to a bound on the sampled rBDP rate: $\nu \beta(t) \leq \beta_s(t) \leq \beta(t)$. The sampled rBDP therefore grows at a rate that is between that of a completely sampled rBDP and a rBDP that includes lineages from the completely sampled one with probability $\nu$ at each event time (Bernoulli thinning). If we restrict ourselves to the constant rate BDP, we find that not only can incomplete sampling reduce $\beta_s(t)$, but it can also remove its time dependence. Setting $\nu = 1 - \rho$ in the $(\rho, \sigma)$ parametrisation and solving Eq. (S1) gives Eq. (S2).

$$\beta_s(t) = \frac{\nu \sigma}{\nu + (1 - \nu - \rho)} e^{-\sigma(T-t)} \mathcal{F}(t) \bigg|_{\nu = 1 - \rho} = \sigma \mathcal{F}(t)$$ (S2)

The lack of dependence on $T$ or $t$ is striking. This sampled BDP yields a reconstructed tree that matches a Yule process [Yule 1924] with birth rate $\lambda - \mu$. As a result, all of the analysis from Section S4 is valid with $\lambda$ replaced by $\sigma$ in those equations. Furthermore, for these parameter settings the BDP will present an analogous inference problem to that of the Kingman coalescent [Parag and Pybus 2017].

Gernhard (2008) showed that, for the constant rate BDP, the inference problem at any (known) $\nu$ could be transformed into one with complete sampling (thus the results in Section 3.1 of the main text apply at any $\nu$). The Yule behaviour noted here relates to this transformation and confounds the analysis of BDP trees by classical statistics, if $\nu$ is unknown. For example, the Pybus and Harvey (2000) gamma statistic will return 0 for any tree satisfying $\nu = 1 - \rho$. For any $\nu$, $\beta_s(t)$ is also equal to $\sigma \mathcal{F}(t)$ when $T \to \infty$. This is related to the asymptotic probability of non-extinction for a constant rate BDP, discussed in [Harvey et al. 1994].

As a final point, we note that while there are many other ways to deterministically or randomly sample extant lineages for the reconstructed process, these often do not yield simple relationships [Stadler 2009]. However, once the resulting sampled rBDP conforms to a Markov birth process then the SF can be applied. Investigating the application of the SF algorithm to various sampling measures will form part of our future work.

**S2. The Snyder Filter Algorithm**

The SF is a deterministic algorithm for directly computing the joint parameter posterior under a self-exciting DSPP (see Methods Section of main text). Below we provide some pseudo-code showing how to apply the SF to time-varying BDP inference problems. While we keep variable names as consistent as possible with the main text, we introduce an additional matrix $\mathcal{E}$. Here $\mathcal{E}$ describes all $m$ possible p-vectors of parameters in our grid. In the main text we used $\epsilon$ to denote a single, arbitrary p-vector and so $\mathcal{E}$ is the complete set of $\epsilon$. Also note that the actual parameters of the birth and death rates would be the subsets $\tilde{x}_k$ and $\tilde{x}_m$, but we have used $\tilde{x}$ to describe the most general case of using the complete parameter vector.

The following set of instructions underlies the implementation available at [https://github.com/kpzoo/snyder-birth-death-code](https://github.com/kpzoo/snyder-birth-death-code)

1. Input a parametric model with $p$ parameter vector $\tilde{x}$
   - Define birth and death rate functions: $\lambda(t, \tilde{x})$ and $\mu(t, \tilde{x})$
   - Create a function that calculates $P(t, T)$ from these rates
2. Construct a $p \times m$ parameter grid, $\mathcal{E}$, with prior probabilities
   - Set delimiting vectors for grid $\tilde{x}_{\text{min}}$ and $\tilde{x}_{\text{max}}$
   - Get $m$ point Cartesian grid, $m_i$ points for $i^{th}$ parameter
   - Define $1 \times m$ probability vector over grid $\mathcal{E}$ as $\mathbb{P}(\mathcal{E})$
3. Input observed phylogeny (timing data) for $n$ lineage
   - Extract branching times as vector $c = (c_0, c_1, \ldots, c_{n-1})$
   - If starting from MRCA then $c$ starts from $c_0$, else $c_0 = 0$
   - If tree is in reverse time, transform to forward time
   - Set $k = 2 \forall$ if starting from MRCA or not
4. Create a rate matrix function to calculate $\Lambda_k$ at $\mathcal{F}(t) = k$
   - Inputs of $P(t, T)$ and $\mathcal{E}$, returns $m \times m$ dimension $\Lambda_k(t)$
   - Calculate $\Lambda_k(t) = \lambda(t, \tilde{x}) P(t, T) k$ over $\mathcal{E}$
   - These are the rBDP grid rates for $k$ extant lineages
5. Define $1 \times m$ posterior vector at $k$ lineages, $q_k$
   - Initialise $q_k \leftarrow \mathbb{P}(\mathcal{E})$ (prior) and dummy $q_d \leftarrow q_d$
6) Solve ordinary differential equations between $c_{k-1} \leq t < c_k$
   • Integrate $m$ simultaneous equations: $\frac{dq_d}{dt} = -q_d \Lambda_k(t)$
   • Normalise $q_d = q_d \left( \sum q_d \right)^{-1}$ and update $q_d \leftarrow q_d$
   • This is the posterior just before the $k^{th}$ rBDP speciation
7) Discontinuously update posterior probabilities at $c_k$
   • Update $q_d \leftarrow q_d \Lambda_k(c_k) \left( \sum q_d \Lambda_k(c_k) \right)^{-1}$
   • Set $k \leftarrow k + 1$ and $q_d \leftarrow q_k$ (a new event has occurred)
8) Repeat above steps 5-7 while $k < n$ then stop
   • Resulting $q_d$ is equivalent to final posterior $q(T)$
   • Marginalise $q_d$ for dimension $i$ and multiply over $i^{th}$ grid
   • Gives $\hat{x}_i$, the MMSE estimator of $x_i$, repeat for $1 \leq i \leq p$

We used the above algorithm to generate the posteriors for the 3-parameter speciation-decay and 4-parameter logistic models of the main text. These were compared to the MCMC posteriors produced by TESS [Hohina et al., 2016]. To give an idea of relative computational complexity we list the mean, $E[t]$, and standard deviation, $\sigma[t]$, of the execution time for both methods (subscripts sf and mc for SF and MCMC respectively) in Table S1. These times are in minutes and obtained by pooling runs over 50 replicate trees for each model. The number of parameters, $p$, and grid size used in the SF, $m$, are also included. A 2-parameter sinusoidal model with $\lambda(t) = x_1 \sin(t) + 1.1$ and $\mu(t) = x_2$ is included for completeness.

For all the models considered, our implementation is faster than the TESS MCMC implementation, for comparable accuracy. Run times are also more consistent for our method, which is unsurprising for a deterministic inference approach. As the SF complexity is linearly dependent on $m$, it may not scale well on finer grids or as $p$ increases. However, as our current implementation is in Matlab, there remains much scope for improving on the SF computational times by porting the code to C/C++ or by employing adaptive grid strategies. Moreover, it is likely that when parameter dimensions become high, it will be wiser to use non-parametric skyline-like techniques.

<table>
<thead>
<tr>
<th>Model</th>
<th>$p$</th>
<th>$m$</th>
<th>$E[t_{sd}]$</th>
<th>$E[t_{mc}]$</th>
<th>$\sigma[t_{sd}]$</th>
<th>$\sigma[t_{mc}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sinusoidal</td>
<td>2</td>
<td>900</td>
<td>1.84</td>
<td>17.35</td>
<td>0.08</td>
<td>0.39</td>
</tr>
<tr>
<td>Speciation-decay</td>
<td>3</td>
<td>8000</td>
<td>4.55</td>
<td>23.74</td>
<td>0.24</td>
<td>3.17</td>
</tr>
<tr>
<td>Logistic</td>
<td>4</td>
<td>50625</td>
<td>25.08</td>
<td>25.62</td>
<td>1.04</td>
<td>1.45</td>
</tr>
</tbody>
</table>

S3. CONSTANT RATE BIRTH-DEATH ESTIMATOR COMPARISONS

We characterised the relative performance of the SF, for the constant rate BDP, by comparing our results with an alternative parametric, least squares method, developed by Paradis [2010]. This technique defines an empirical, reference cumulative distribution function (CDF) of the tree branching times, $F_c(t)$, obtained by rescaling the lineages through time plot of the observed rBDP. The Paradis [2010] method calculates, under a given BDP model with parameters $\mathbf{x}$, the theoretically expected branching time CDF, $F(t)$ and then identifies the optimal $\mathbf{x}_{\text{opt}}$ at which $F(t)$ is closest to $F_c(t)$, in a least squares sense. For the constant rate BDP, $\mathbf{x} = (\lambda, \mu)$ and $F(t)$ can be derived by taking the original formulae of Paradis [2010] and substituting expressions from Nee et al. [1994], to give Eq. (S3) with $D_{\mu} := \mu e^{(\lambda - \mu)T}$, where $\lambda$ and $\mu$ are the solutions to the optimisation-search problem of Eq. (S4).

$$F(t) = \frac{\int_0^t D_x^{-2} e^{(\lambda - \mu)s} \, ds}{\int_0^T D_x^{-2} e^{(\lambda - \mu)s} \, ds} = \frac{I(t)}{I(T)} \quad \text{(S3)}$$

$$\langle \lambda_{\text{opt}}, \mu_{\text{opt}} \rangle = \arg \min_{\lambda, \mu} \int_0^T (F_c(t) - F(t))^2 \, dt \quad \text{(S4)}$$

If the least squares search space for $(\lambda, \mu)$ is the same as that used in the SF grid then we can compare the results of both methods. Since the parameter space is small for the constant rate BDP we obtained the optimal least squares estimates $(\hat{\lambda}_{\text{opt}}, \hat{\mu}_{\text{opt}})$ using a brute force search. To maximise precision we used analytic solutions where possible. Setting $a = \lambda - \mu$, $b = \lambda$, and $c = \mu e^{(\lambda - \mu)T}$, we solved the integrals in equation Eq. (S3) to get

$$I(t) = \frac{1}{ab} \left[ e^{e^{at}} - e^{bt} \right] + \frac{c}{b} (e^{at} - 1) - e^{e^{at}}.$$ 

Fig. S1 compares the SF and Paradis [2010] approaches over 2724 replicate 200 tip rBDP trees. A uniform prior of $\frac{1}{ab}$ is used across the $m$ point SF grid in the $(\lambda, \mu)$ parameterisation to keep comparisons fair. The Paradis [2010] method achieves a slightly better estimate of $\lambda$ but a worse one for $\mu$. Overall the methods perform similarly, with the SF being computationally simpler. One key conceptual difference between these methods is that while both achieve least squares estimates (though according to different criteria), the SF has its optimisation built into the differential equations it solves and so does not require explicit optimisation algorithms. The Paradis [2010] method requires such algorithms and is susceptible to issues like local minima, especially when used on more complex models.

**Fig. S1. Comparison of constant rate BDP estimates.** The SF and Paradis [2010] inference methods are applied to 2724 replicate trees with $n = 200$ tips simulated under the constant rate BDP model. Both methods used a grid with $m = 30$ points for each parameter. Uniform priors over $[0, 0.1]$ for $\lambda$ and $[0.005, 0.5]$ for $\mu$ were used to ensure a fair comparison between methods. The estimates from each tree are summarised in the box plots against the true parameter value (horizontal dashed line).

S4. THE YULE MODEL AND THE KINGMAN COALESCENT

BDPs and coalescent models are alternative descriptions of the shared ancestry of the individuals within a population. Understanding the similarities and differences between these descriptions is important for choosing the appropriate null model for a given dataset. Volz and Frost [2014] and Studier et al. [2015]. Here we make some comparisons with the Kingman [1982] coalescent, which models the sample genealogy in a constant sized population. BDP and coalescent models have previously been compared in terms of density functions or inter-event intervals. Gernhard [2008] showed that constant rate BDPs with $\lambda = \mu = \frac{n}{2}$ have the same mean inter-event time as a
scaled Kingman (1982) coalescent. Hey (1992) found that the true birth-death equivalent to the coalescent (provided a rate correction is applied Volz (2012)) is achieved if birth and death events are synchronised and λ = μ.

Unlike the BDP, the coalescent is defined in backward time and assumes a population size that is large in comparison to the sample size n (Northore 2001). We first rewrite the coalescent equations in forward time to match the BDP and use \( F_c(t) \) to count its lineages at \( t \). The Kingman (1982) coalescent with population \( N(0) \) then has observable inter-event intervals exponentially distributed with rate \( \frac{\lambda}{2} F_c(t+1) N(0)^{-1} \). Given \( F_c(t) \), each interval derives from a homogeneous Poisson process. Consequently, the coalescent admits a simple and well-known analytic solution for the estimator of \( N(0) \).

We presented the analytic SF solution for this problem in Parag and Pybus (2017). The constant rate BDP with \( \lambda = \mu \), however, does not lead to a similar description. This is because the observable BDP is an inhomogeneous Poisson process given \( F(t) \). This can be seen by applying L’Hopital’s rule to the constant rate BDP intensity, \( \beta(t) \).

\[
\lim_{\mu \to \lambda} \beta(t) = \lim_{\mu \to \lambda} \frac{-\lambda F(t)}{\mu F_c(t)} = \frac{\lambda F(t)}{1 + \lambda T - t} \quad (S5)
\]

This results because the birth and death events are completely independent, which is in contrast to the Hey (1992) model. It also emerges because when the BDP is conditioned on survival (which is the case for an rBDP to be observed) then the expected number of lineages is not constant, despite the equality in birth and death rates (Hohna 2015). The impact of conditioning becomes clear if we note that, for all BDPs with \( \lambda(t) = \mu(t) \), \( \mathbb{P}(t, T) \) is the same as the inverse of the expected number of lineages at \( T \) conditioned on survival of the true BDP (Hohna 2015). Using the expressions of Hohna (2015), with \( l(t) \) as our true lineage count, we get:

\[
\beta(t) \bigg|_{\lambda(t) = \mu(t)} = \frac{\lambda(t) F(t)}{1 + \int_t^T \lambda(s) ds} = \frac{\lambda(t) F(t)}{\mathbb{E}[l(T) | l(t) \geq 1]} \quad (S6)
\]

However, the inference problem for the coalescent can be made identical to that for a BDP if the death rate is set to 0. This results in the Yule process (Yule 1924). Under this process the rBDP and BDP are equivalent \( (l(t) = F(t)) \). This allows us to analytically solve the SF equations in a manner similar to that presented in Parag and Pybus (2017) for the Kingman (1982) coalescent.

The Yule process has inter-event times that BDPs are independently and exponentially distributed with rate \( \kappa \lambda \) over the period \( c_k - c_{k-1} \). The known maximum likelihood estimate of \( \lambda \), \( \lambda_{\text{ML}} = \frac{(n - 1) T^{-1}}{T} \) (Moran 1951), \( T = \sum_{k=1}^{n-1} k(c_k - c_{k-1}) \) is the sufficient statistic for the Yule process, with \( c_0 := 0 \). If the observed intervals are exponentially scaled by the number of lineages to \( (c_k - c_{k-1}) \) \( k = d_k \) and a new observed process defined \( \tilde{G}(t) := \arg\max_{1 \leq k \leq n-1} \sum_{k} d_k \leq t \), then \( \tilde{G} \) is homogeneous and Poisson. The SF equations can then be solved analytically with \( \tilde{G} \) to yield the posterior of the Yule process (Eq. (S7), Snyder and Miller 1991), with \( \epsilon \) as an arbitrary value of \( \lambda \). This solution is only applicable to the Yule (and not the general BDP) because of the transformation from \( F_c \) to \( G_t \). Note that \( F_c(t) = \tilde{G}(t) \) and the MMSE estimator of the Yule model is \( \lambda_0 = e\mathbb{E}[G_t] \).

\[
\mathbb{P}(\epsilon | F_t) = \mathbb{P}(\epsilon | G_t) = \frac{e^{G_t(t)} e^{-e^{-G_t(t)} \mathbb{E}[\epsilon]} \mathbb{E}[\epsilon](G_t) \mathbb{E}[\epsilon] e^{-\mathbb{E}[\epsilon]} \mathbb{E}[\epsilon] d\epsilon} {\int_0^{\infty} e^{G_t(t)} e^{-e^{-G_t(t)} \mathbb{E}[\epsilon]} \mathbb{E}[\epsilon] \mathbb{E}[\epsilon] e^{-\mathbb{E}[\epsilon]} \mathbb{E}[\epsilon] d\epsilon} \quad (S7)
\]

Here \( \mathbb{P}(\epsilon) \) is the prior and \( G_t \) ends at \( T = T \). This equivalence to the Kingman (1982) coalescent also applies to constant rate BDPs with \( \mu \neq 0 \), when sampled isochronously at \( \nu = 1 - \frac{\lambda}{\mu} \) (see Section S1), since they are Yule processes with birth rate \( \lambda - \mu = \nu \lambda \).

S5. SF Likelihoods and other Time-varying Results

We previously showed that the SF implicitly solves the log-likelihood function \( H(\epsilon) \), with \( \epsilon \) as an arbitrary value of \( \tilde{r} \) (Eq. (8) of the main text). Here we present some additional results. We expand \( H(\epsilon) \) into its interval sums so that for an \( n \) tip BDP with \( F(T) = n: H(\epsilon) = \sum_{k=1}^{n-1} H_k \) and \( H_k = -\int_{c_k-1}^{c_k} \beta(s, \epsilon, k) ds + \log(\beta(c_k, \epsilon, k)) \). Solving this explicitly with the expressions from Section S1 admits the interval log-likelihood function for an incompletely, isochronously sampled, constant rate BDP (Eq. (S8)).

\[
H_k = -k\sigma(c_k - c_{k-1}) + \log \frac{k\nu}{\nu + (1 - \nu - \rho)e^{\sigma(c_k - T)}} - k\log \frac{\nu + (1 - \nu - \rho)e^{\sigma(c_k - T)}}{\nu + (1 - \nu - \rho)e^{\sigma(c_k - T)}} \quad (S8)
\]

We used this expression, at \( \nu = 1 \), in our comparisons with the most log-likelihoods from Stadler (2013) (Section 3.2 of the main text). These comparisons were made by numerically solving the relevant likelihood expressions. For the Yule model the correspondence between the SF and these 7 likelihoods can be shown exactly since \( H(\epsilon) \) can be analytically computed. The full likelihood calculated from the start of the tree \( F(0) = 1 \) is:

\[
H(\lambda) = (n - 1)\lambda^n - e^{-\sum_{k=1}^{n-1} k(c_k - c_{k-1})} = (n - 1)!\lambda^n - e^{-\sum_{k=1}^{n-1} \lambda(c_k - c_{k-1})}
\]

with \( \epsilon = \lambda \). This is equivalent to \((n - 1)!L(\lambda) \) where \( L(\lambda) \) is the tree likelihood for an rBDP conditioned on survival to T, with \( \mu = 0 \). This is the likelihood (2) from Stadler (2013) and further confirms the results in the Section 3.2 of the main text. A similar equality between the SF and the likelihood (5) can be obtained if the calculations are done from the tree MRCA \( F(c_1) = 2 \).

The interval log-likelihood, \( H_k \), also readily provides MMSE estimates that are analytic when the BDP intensity, \( \beta(t) \) has closed integrals. If we sequentially apply Bayes theorem along the branches of the BDP we can adapt a solution from the analysis of inhomogeneous Poisson process (Snyder and Miller 1991) to get Eq. (S9). Here we use the posterior from the last interval \( \mathbb{P}(\epsilon | F_t < c_{k-1}) \) as the prior for the subsequent one by the Markov property. Eq. (S9) can be seen as a generalisation of Eq. (S7).

\[
\int_{c_{k-1}}^{c_k} e^{H_k} \mathbb{P}(\epsilon | F_t < c_{k-1}) \int_{c_{k-1}}^{c_k} e^{H_k} \mathbb{P}(\epsilon | F_t < c_{k-1}) d\epsilon \quad (S9)
\]

For most time-varying BDPs these integrals are not closed so the SF should be used, as it is more numerically stable. However, for problems which do meet this special criteria, Eq. (S9) provides exact posteriors with no approximations.

S6. Further Model Selection Results

In Section 3.4 of the main text, we solved a time-varying BDP model selection problem for an Australian Agamid lizard dataset, and compared our results with the literature. Fig. S2 provides the empirical Agamid phylogeny that served as the observed rBDP here. We used a generalised form of Eq. (S4), which involves computing a squared error between a scaled Agamid LTT plot and the expected CDF from each BDP model, as our selection metric. The scaled LTT and CDFs for each model we examined are shown in Fig. S3. We applied this metric, as it is quick and easy to compute, similar to the Cramer-von Mises goodness of fit measure (Arnold and Emerson 2011), has precedence in the BDP literature (Paradis 2010), and transparently links our estimates to the observed LTT.
S7. The SF as a Practical Diagnostic Tool

The SF is a deterministic inference algorithm that leads to reproducible posteriors. In the main text, we suggested that these properties make the SF a good tool for debugging and validating MCMC based estimates. As an example of this, consider estimating the death parameter, $x_3$, of the speciation-decay model defined in the main text. Under settings that often produce valid runs, we find the two cases depicted in Fig. S4. Often users will rely on visual inspection of the trace (top panels) and the effective sampling sizes (ESS) for judging convergence (an ESS > 200 is usually seen as a rule of thumb for convergence). Without checking the autocorrelation function (acf), one may therefore think that the MCMC posterior (light grey) on the bottom left looks better than the one on the opposite right panel, since it seems symmetrical and the ESS is sufficient (it also happens to be closer to the true value (dashed lines)). The acfs (middle panels) reveal that the left trace has not actually converged.

The bottom right panel shows a consistency between the SF (dark grey) and MCMC and so we are confident in the results of the MCMC run here. One could argue that in the left case the MCMC simply needs to be run for more iterations or with different settings. However, this is precisely the issue we are highlighting. With no practical bound on the number of MCMC iterations above which convergence can be guaranteed, one cannot be sure that the left case would not occur. This demonstrates the importance of the SF as a deterministic reference. The SF could become even more valuable as a diagnostic tool when genealogical uncertainty is marginalised using MCMC since (i) there is an exponential increase in MCMC convergence time when such uncertainty is included and (ii) misleading or overconfident posteriors are a known issue in phylogenetic model selection problems.

Additionally, we present some more formal results on model selection. In Table S2 we use the Akaike information criterion (AIC) to discriminate between the proposed Agamid speciation and decay models. Smaller AICs indicate better fits. We utilise the equivalent SF likelihood to calculate its AIC (Section 3.2 of main text). We obtain AICs for the Rabosky [2006] method from the R package LASER. Table S2 reaffirms the conclusions previously made in Section 3.4.

Table S2

<table>
<thead>
<tr>
<th>Model</th>
<th>Rabosky</th>
<th>SF</th>
</tr>
</thead>
<tbody>
<tr>
<td>const</td>
<td>-562.9</td>
<td>-565.0</td>
</tr>
<tr>
<td>spvar</td>
<td>-580.5</td>
<td>-581.8</td>
</tr>
<tr>
<td>exvar</td>
<td>-560.9</td>
<td>-563.0</td>
</tr>
<tr>
<td>bothvar</td>
<td>-578.5</td>
<td>-579.5</td>
</tr>
</tbody>
</table>

Fig. S2. **Australian Agamid Phylogeny.** This $n = 69$ tip ultrametric empirical BDP tree was generated from the newick string given in Rabosky (2006), based on the data collected in Harmon et al. (2003). This was used as the observed rBDP for the SF model selection problem. The time units are normalised with $T = 0.322$ as the observation time of the tips.

Fig. S3. **Fitted CDFs for the Agamid model selection problem.** The SF expected CDFs, for each BDP model investigated, are shown as solid grey curves. These are compared to the empirical CDF (scaled LTT) (grey stepwise function). The const and exvar models fit poorly (light grey) while spvar and bothvar give equally good fits (dark grey). The Rabosky and Lovette (2008) CDFs are not shown as they are very close to the SF ones.

Fig. S4. **MCMC convergence issues under identical settings.** Two independent MCMC runs (left and right panel set) are given for the same settings and parameters. The top panels give the traces for the death parameter, $x_3$, of the speciation-decay model of the main text. Middle panels show the autocorrelation function (acf) for these traces while the bottom panels are the empirical CDFs (scaled LTT) (grey stepwise curves). These are compared to the empirical CDF (scaled LTT) (grey). The true values of $x_3$ indicate that the traces are valid (dashed vertical lines).
ACKNOWLEDGMENTS

This work was supported by the European Research Council under the European Commission Seventh Framework Programme (FP7/2007-2013)/European Research Council grant agreement 614725-PATHPHYLODYN.

REFERENCES


