# Statistical and structural bias in birth-death models

Jeremy M. Beaulieu[1*†] and Brian C. O'Meara[2†]

[1*]Department of Biological Sciences, University of Arkansas, Fayetteville, 72701, AR, USA.
[2]Department of Ecology and Evolutionary Biology & National Institute for Modeling Biological Systems, University of Tennessee, Knoxville, 37996-1610, TN, USA.

*Corresponding author(s). E-mail(s): jmbeauli@uark.edu;
Contributing authors: bomeara@utk.edu;
†The authors contributed equally to this work.

**Abstract**

Accurate estimation of speciation ($\boldsymbol{\lambda}$) and extinction ($\boldsymbol{\mu}$) rates from phylogenetic trees is central to studies of diversification, yet it remains unclear whether commonly used estimators are unbiased. Here we examine two sources of error: (1) statistical bias in the estimators themselves, and the (2) structural bias introduced by how small trees are handled in likelihood calculations. For the Yule process, we re-derive the expected bias of the standard estimator, showing that $\hat{\boldsymbol{\lambda}}$ underestimates $\boldsymbol{\lambda}$ by a factor of $\boldsymbol{(n-2)/(n-1)}$. Extending to the general birth–death model, we use symbolic regression to find functional forms that minimize the bias in both $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$. The best-performing correction for $\boldsymbol{\lambda}$ is similar to the Yule result, while the bias in $\boldsymbol{\mu}$ depends on both turnover ($\boldsymbol{\lambda+\mu}$) and the extinction fraction ($\boldsymbol{\mu/\lambda}$). Applying these corrections substantially improves the fit between the estimated and generating values. When these corrected estimators are used to derive other diversification-related parameters, turnover is nearly unbiased, but net diversification ($\boldsymbol{\lambda-\mu}$) remains systematically underestimated, particularly at larger values, due to the slight overestimation of $\boldsymbol{\mu}$. On the whole, these results clarify the statistical and structural sources of bias in diversification rate estimation and provide a general framework for improving inference under birth–death models.

**Keywords:** birth-death, Yule, diversification, bias, cherry trees, conditioning, phylogenies

# 1 Introduction

Understanding and uncovering the dynamics of speciation and extinction is a central question in evolutionary biology. Birth–death models provide a mathematical framework for estimating these rates from phylogenetic trees and are widely used to study the tempo of diversification. Reliable estimates of speciation and extinction rates form the basis for downstream analyses, such as testing for shifts in diversification across lineages or through time (e.g., Nee et al. 1992; Rabosky 2006, 2009; Bokma 2008; Rabosky and Lovette 2008; Morlon et al. 2011; Etienne et al. 2012), identifying lineage-specific heterogeneity (e.g., Alfaro et al. 2009; Rabosky 2014; Maliet et al. 2014; Vasconcelos et al. 2022a), and linking diversification patterns to ecological and/or trait data (e.g., Slowinski and Guyer 1993; Maddison et al. 2007; Beaulieu and O'Meara 2016).

While there are various ways to estimate rates from these models, the most common is maximum likelihood (or related Bayesian approaches). Maximum likelihood has many advantages, such as consistency with increasing sample size and efficient use of information in the data. However, likelihood estimators are often biased at small sample sizes. For many questions, sample sizes are necessarily limited: there are only so many species of living whales, so any approach for understanding modern whale diversificaton has a small number of data points. Moreover, many popular approaches (e.g., Medusa, Alfaro et al. 2009; BAMM, Rabosky 2014; ClaDS, Maliet et al. 2014; MiSSE, Vasconcelos et al. 2022a) subdivide a tree into smaller parts, each with its own estimates of rates.

We have become concerned with several potential sources of bias that can influence inference. One concern is *statistical bias* – that is, systematic deviations in the expectations of birth and death rate estimators from their true generating values. Such biases produce apparent trends in diversification that are artifacts of the estimation procedure rather than features of the underlying process, especially with small sample sizes. For instance, the commonly used Yule estimator (Yule 1924; Nee 2001)

systematically underestimates speciation rates (O'Meara and Beaulieu 2024). A second concern is *structural bias*, arising from how likelihoods are conditioned and how trees of particular sizes are treated. Many formulations assume survival of the crown clade (e.g., Stadler 2013), but in practice trees of particular sizes are often excluded from analysis. Specifically, two-taxon trees ($n = 2$) are often excluded, either because the likelihood is not defined or because such clades are rarely, if ever, subject to rate estimation. This implicit filtering introduces an additional layer of conditioning that could alter the distribution of observed clades and, in turn, affect parameter estimates. A third potential source is *ascertainment bias*, which is when researchers preferentially select unusual or noteworthy clades. While this does not bias estimates within those clades, it can limit how representative they are of broader diversification patterns. Finally, ignoring uncertainty in measurements introduces its own biases; for more on that, see O'Meara and Beaulieu (2024).

These issues are not just academic. Because many diversification rate frameworks have been derived to estimate heterogeneity in diversification rates across the tree of life (e.g., Medusa, Alfaro et al. 2009; BAMM, Rabosky 2014; ClaDS, Maliet et al. 2014; MiSSE, Vasconcelos et al. 2022a), statistical and structural biases in the estimation of $\lambda$ and $\mu$ can propagate into spurious conclusions about rate variation. For example, upwardly biased estimates for young clades can be misinterpreted as evidence for declining diversification rates over time or lineage-specific bursts of speciation (O'Meara and Beaulieu 2024).

In this paper, we examine the mathematical underpinnings of the first two potential biases and propose solutions. We begin by analyzing the behavior of rate estimators on the smallest possible clades, demonstrating why $n = 2$ trees lack sufficient information to identify both speciation and extinction separately. Throughout, we refer to complete two-taxon trees as cherry trees – that is, rooted trees with only two terminal edges. This usage distinguishes them from internal "cherries" (pairs of sister taxa within

larger trees; see McKenzie and Steel 2000) and emphasizes their role as minimal cases in diversification models. We then derive the appropriate conditioning for likelihoods when cherry trees are excluded from empirical datasets. Building on this, we then revisit the Yule estimator and derive an analytical bias correction, clarifying how an earlier empirical correction from our previous work relate to theory. Finally, we extend these insights to the general birth–death model, where extinction complicates analytical results, and use symbolic regression to obtain practical bias corrections. Together, these results highlight the importance of careful statistical treatment when estimating diversification rates and provide a framework for reducing biases in diversification rate analyses when dealing with small trees or subtrees.

## 2 Likelihood functions and the problem of cherry trees

With the Yule model of phylogenetic tree growth, there is only a birth rate, $\lambda$, and no extinction ($\mu$=0). This means that once a lineage arises, it persists indefinitely, without the possibility of going extinct. Consequently, every branch of the tree represents a successful speciation event without subsequent loss. The Yule model is often considered a framework that captures unbounded species diversification dynamics. The maximum likelihood estimate of $\lambda$ can be calculated directly using the formula derived in Nee (2001):

$$\hat{\lambda} = \frac{n-2}{s}, \tag{1}$$

where $n$ is the number of tips in the tree and $s$ is the total sum of all the lengths of the branches. Notice that when $n = 2$ estimates of $\hat{\lambda}$ become zero. This is intuitive: in the absence of extinction, initializing a tree at $n = 2$ that remains unchanged after some time interval implies that no additional speciation events occurred, which is a speciation rate of zero.

With a standard birth-death model, where we assume that extinction is $\mu > 0$, when estimating rates on a cherry tree, we have more parameters than observed speciation events: there is simply not enough information in them to uniquely estimate both $\lambda$ and $\mu$ (see Appendix A). However, there are more immediate and practical considerations. As discussed by Stadler (2013), most implementations of the constant birth-death model rely on the same underlying likelihood function. Specifically, many use a form of,

$$f(T|t_1 = t_{crown}) = \left(\frac{p_1(t_1)}{1 - p_0(t_1)}\right)^2 \prod_{i=2}^{n-1} \lambda p_1(t_i), \tag{2}$$

which assumes that the process is conditioned on the survival of the crown (that at least one lineage persists from each of the daughter lineages of the root). Without delving into the specifics of how to calculate $p_1$ and $p_0$ in the above expression, we direct attention to the product term on the right-hand side. This term iterates over a vector of $n - 1$ speciation times from a phylogeny with $n$ species, but the indexing begins at $i = 2$. In other words, the product excludes the crown node, which is handled separately by the term on the left. When the tree is a cherry tree ($n = 2$), the vector of speciation times is empty and the term of the product is undefined. In practice, this means that the likelihood cannot be computed for two-taxon trees using this formulation (or any variant thereof; see Stadler 2013), and therefore implicitly exclude such trees by assuming that tree sets contain only $n \geq 3$ trees.

The exclusion of two-taxon trees has important consequences, especially with young clades. To illustrate, we simulated 100,000 trees under a constant-rate birth-death process ($\lambda$=0.1, $\mu$=0.05), estimating both $\lambda$ and $\mu$ using the likelihood function of Stadler (2013) across a range of clade ages. As expected, estimates based on all surviving trees (but excluding those with only two tips) show a clear trend: younger clades exhibit upwardly biased diversification rates (Table 1).

We note that this issue does not arise when using alternative likelihood frameworks that handle small trees more directly. For example, the approach developed by

**Table 1** Summary of estimates of speciation and extinction rates using the likelihood formula of Stadler (2013).

| | Conditioned on $n > 0$ | | Conditioned on $n > 2$ | |
|---|---|---|---|---|
| Clade age | $\widehat{\lambda}$ | $\widehat{\mu}$ | $\widehat{\lambda}$ | $\widehat{\mu}$ |
| 5 | 0.165 | 0.096 | 0.093 | 0.041 |
| 10 | 0.121 | 0.071 | 0.095 | 0.049 |
| 15 | 0.109 | 0.064 | 0.097 | 0.054 |
| 20 | 0.104 | 0.061 | 0.098 | 0.056 |
| 25 | 0.102 | 0.060 | 0.099 | 0.057 |
| 30 | 0.101 | 0.060 | 0.099 | 0.058 |

Generating rates were $\lambda = 0.1$ and $\mu = 0.05$.

Maddison et al. (2007) that is implemented in SSE models does not rely on the product over internal nodes and thus can accommodate trees with as few as two taxa. The likelihood under this formulation is,

$$L(\lambda, \mu) = \left[ \prod_{i=1}^{2m} e^{-(\lambda-\mu)(t_{i,r}-t_{i,t})} \frac{(\lambda - e^{-(\lambda-\mu)t_{i,t}}\mu)^2}{(\lambda - e^{-(\lambda-\mu)t_{i,r}}\mu)^2} \right] \frac{\lambda^m}{\left( \frac{\lambda(\lambda-\mu)^2}{(\lambda - e^{-(\lambda-\mu)T}\mu)^2} \right)} \tag{3}$$

where $m$ is the number of internal nodes, and thus $2m$ is the number of branches in the tree, $T$ is the total height of the tree, and finally $t_{i,t}$ and $t_{i,r}$ reflect the tipward and rootward ages of node $i$.

However, this raises an important question: even if a likelihood can be evaluated for $n = 2$, does such a tree contain enough information to estimate both speciation and extinction rates? Analytical results show that it does not – cherry trees simply lack the complexity needed to disentangle these two processes, even when the likelihood is defined (see Appendix A). We therefore recommend excluding two-taxon trees (or subtrees) from analyses, but conditioning the likelihood accordingly to account for their absence.

# 3 Conditioned likelihood for $n > 2$ censored tree sets

As shown above, two-taxon trees contain little information for estimating speciation and extinction rates, and in some likelihood formulations their probability cannot be computed directly. In practice, such trees are often excluded – sometimes explicitly by imposing a minimum clade size, and sometimes implicitly because likelihood terms are undefined. This exclusion introduces an additional form of conditioning beyond the usual requirement that the crown lineage survives to the present (i.e., Stadler 2013). In other words, inference is restricted to clades with more than two extant taxa. This second layer of conditioning is nontrivial because the probability of observing a two-taxon tree depends on both the underlying model and the clade's age, and therefore varies under varying birth-death dynamics. To account for this, we derive the appropriate conditioning term when datasets are censored to exclude $n \leq 2$ trees, under a general birth–death process and its two special cases: the Yule process ($\mu = 0$) and the critical branching process ($\lambda = \mu$).

## 3.1 Birth-death model ($\lambda \neq \mu$)

We begin with the unconditioned likelihood equation of Maddison et al. (2007),

$$L(\lambda, \mu) = \left[ \prod_{i=1}^{2m} e^{-(\lambda-\mu)(t_{i,r}-t_{i,t})} \frac{(\lambda - e^{-(\lambda-\mu)t_{i,t}}\mu)^2}{(\lambda - e^{-(\lambda-\mu)t_{i,r}}\mu)^2} \right] (\lambda)^m,$$

where $m$ is the number nodes, and $t_{i,t}$ is the time at the tipward end, and $t_{i,r}$ is the time at the rootward end, of the the $i$th edge. Conditioning on the existence of the crown node with two surviving lineages requires scaling the likelihood by the probability of survival. The probability that a single lineage goes extinct after time, $t$ is:

$$P_{n=0}(t) = 1 - \frac{\lambda - \mu}{(\lambda - e^{-(\lambda-\mu)t}\mu)}.$$

Thus, the probability of survival of the two descendants of the root is, $[1 - P_{n=0}(t)]^2$. The factor $\lambda$ accounts for the initial speciation event at the crown. Together, the likelihood scaled by crown survival becomes,

$$L(\lambda, \mu) = \left[\prod_{i=1}^{2m} e^{-(\lambda-\mu)(t_{i,r}-t_{i,t})} \cdot \frac{\left(\lambda - \mu e^{-(\lambda-\mu)t_{i,t}}\right)^2}{\left(\lambda - \mu e^{-(\lambda-\mu)t_{i,r}}\right)^2}\right] \cdot \lambda^{m-1} \times \frac{1}{[1 - P_{n=0}(t)]^2} \quad (4)$$

We are interested in further conditioning on observing trees with more than two tips. The probability that a single lineage produces exactly one descendant after time, $t$, is given by Stadler (2013) as:

$$P_{n=1}(t) = \frac{(\lambda - \mu)^2 e^{-(\lambda-\mu)t}}{(\lambda + (\lambda - \mu)e^{-(\lambda-\mu)t})^2}$$

For a crown with two descendants, the probability that both produce exactly one tip is:

$$P_{(n=2|n>0)}(t) = P_{n=1}(t)^2$$

Therefore, that the crown survives *and* produces more than two tips is:

$$P_{(n>2)}(t) = \lambda[1 - P_{n=0}(t)]^2[1 - P_{n=1}(t)^2]$$

The new conditioned likelihood then becomes:

$$L(\lambda, \mu \mid n > 2) = \left[\prod_{i=1}^{2m} e^{-(\lambda-\mu)(t_{i,r}-t_{i,t})} \cdot \frac{\left(\lambda - \mu e^{-(\lambda-\mu)t_{i,t}}\right)^2}{\left(\lambda - \mu e^{-(\lambda-\mu)t_{i,r}}\right)^2}\right] \cdot \lambda^{m-1} \times \frac{1}{[1 - P_{n=0}(t)]^2\,[1 - P_{n=1}(t)^2]} \quad (5)$$

Having derived the appropriate conditioning for datasets where $n = 2$ trees are excluded, we next evaluated whether incorporating this adjustment corrects the systematic bias observed in simulations (Table 1). When we return to the simulated datasets, the effect of applying this additional conditioning is clear. If we modify the likelihood so that it reflects not only the probability that the crown lineage at time $t$ survives to the present, but also that it gives rise to a clade with $n > 2$, the upward trend in rate estimates disappears. In fact, estimate becomes slightly downward biased compared to the true generating parameters (Table 1).

## 3.2 Special case 1: Yule model ($\mu = 0$)

Following Nee (2001, Eq. 3), the likelihood of an observed tree under a Yule model is defined as,

$$L(\lambda) = (n-1)!\lambda^{n-2}e^{-\lambda s}$$

where the term $s$ represents the sum of all $x_i$ edge lengths in the tree (i.e., $\Sigma(x_i)$). To condition the log-likelihood on the observed tree by the probability of producing any tree with more than two taxa. In other words, $P(n > 2)$ represents the probability that a tree of the exact same age and speciation rate would have more than two taxa. To condition on $n > 2$, we start with the probability of $n$ lineages, given $\lambda$ and $t$, from Nee (2001),

$$P_n = (n-1)e^{-2\lambda t}(1-e^{-\lambda t})^{n-2}$$

where $t$ represents the length from root to tip. Note that the use of $t$ is to distinguish it from $s$ in the log-likelihood. If we assume that $n = 2$, this equation simplifies to:

$$P_{n=2} = e^{-2\lambda t}$$

Therefore, the probability that a tree has $n > 2$, given $\lambda$ and $t$ is

$$P_{n>2} = 1 - e^{-2\lambda t}$$

and the conditioned likelihood is calculated as:

$$L(\lambda | n > 2) = \frac{(n-1)!\lambda^{n-2}e^{-\lambda s}}{1 - e^{-2\lambda t}} \tag{6}$$

### 3.3 Special case 2: Critical branching process ($\lambda = \mu$)

With the general birth-death model, the probability of a phylogeny depends on both $\lambda$ and $\mu$. However, when $\lambda = \mu$ the contribution of $\mu$ to the likelihood becomes indistinguishable from that of $\lambda$. Specifically, both speciation and extinction events contribute equally to the branching patterns in the tree, and their exact roles in determining the timing and structure of the tree are indistinguishable from each other. As a result, the likelihood function is simplified and depends only on a single parameter, $\lambda$. Following the derivation provided by Maddison et al. (2007), the unconditioned likelihood of an observed tree under the critical branching process model:

$$L(\lambda) = \left[\prod_{i=1}^{2m} \frac{(1 + \lambda t_{i,t})^2}{(1 + \lambda t_{i,r})^2}\right] (\lambda)^m$$

Bailey (1964, Eq. 8.53) derived the probability of observing $n$ species at time, $t$, under a critical branching process (i.e., when $\lambda = \mu$):

$$P_n(t) = \frac{(\lambda t)^{n-1}}{(1 + \lambda t)^{n+1}}$$

As with the general birth-death model, we are interested in the probability that a crown group of age, $t$, both survives (i.e., both initial lineages leave descendants) and

produces more than two tips. The probability of extinction of a single lineage after time, $t$, is calculated as (Maddison et al. 2007):

$$P_{n=0}(t) = \frac{\lambda t}{(1 + \lambda t)}$$

Thus, the probability of survival of the two descendants of the root is,

$$P_{n>0}(t) = \lambda[1 - P_{(n=0)}(t)]^2.$$

The probability of generating exactly two tips (i.e., both lineages yield exactly one descendant) is:

$$P_{(n=2|n>0)}(t) = P_{n=1}(t)^2$$

Therefore, the joint probability of crown survival and tree size $n > 2$ is:

$$P_{(n>2)}(t) = \lambda(1 - P_{n=0}(t))^2(1 - P_{n=1}(t)^2)$$

The new conditioned likelihood then becomes:

$$L(\lambda|n > 2) = \left[\prod_{i=1}^{2m} \frac{(1 + \lambda t_{i,t})^2}{(1 + \lambda t_{i,r})^2}\right] \lambda^{m-1} \cdot \frac{1}{[1 - P_{n=0}(t)]^2\,[1 - P_{n=1}(t)^2]}. \tag{7}$$

# 4 Correcting estimator bias

In a recent publication (O'Meara and Beaulieu 2024), we showed that the Yule estimator is actually biased *downwards* as a function of time – that is, as time approaches the present, or across short time intervals, rate estimates increasingly fall below the generating value. We suspect that this phenomenon has gone largely unnoticed because trees

that fail to speciate (trees of $n = 2$) after time $t$ are almost always discarded. Excluding these trees make it seem like rates increase dramatically towards the present or across shorter time intervals. In our study, we used a bias "correction" of $n/(n-1)$ that produced a remarkably constant trend through time when the zero rates are included. What was not clear from the paper is that this correction was actually derived empirically from simulation results (we simply showed that the naive estimator was indeed biased). As we detail below, the proper correction term is $(n-1)/(n-2)$.

By contrast, analytically deriving bias corrections for birth-death models is considerably more difficult, even though both $\lambda$ and $\mu$ appear downwardly biased across short timescales (Table 1). The presence of extinction introduces nonlinear dependencies between parameters, making it intractable to isolate closed-form expectations of the estimators. This complexity arises because the likelihood under a birth-death process depends jointly on speciation ($\lambda$) and extinction ($\mu$) and becomes even more complex when it is conditioned on survival and a minimum number of extant taxa. As a result, while exact analytical corrections are not readily available, we used symbolic regression to search for *approximate* transformations of the estimated parameters that systematically reduce bias. To validate this approach, we first applied it to trees simulated under a Yule process and confirmed that symbolic regression recovered the known analytical correction. We then extend this method to estimates from trees simulated under a wide range of birth-death processes.

## 4.1 Bias correction for the Yule model

A key insight from Steel and Mooers (2010) is that, under a Yule process, the expected average edge length is $1/(2\lambda)$, as opposed to $1/\lambda$, because each speciation creates two lineages. We begin by restating the standard Yule likelihood formula (Nee 2001) from

Eq. 1, where the MLE estimator of $\lambda$ is computed as

$$\hat{\lambda} = \frac{n-2}{\sum x_i}.$$

If we view $s$ as the total branch length in a bifurcating tree with $2n-2$ edges and use the edge-length expectation $E[x_i] = 1/(2\lambda)$, then $E[s] = (2n-2) * \frac{1}{2\lambda} = (n-1)/\lambda$. Plugging this into the expectation for $\hat{\lambda}$ yields

$$E(\hat{\lambda}) = \frac{n-2}{E[s]} = \frac{n-2}{(n-1)/\lambda} = \lambda\frac{n-2}{n-1},$$

so $\hat{\lambda}$ is downward biased. To derive the bias of the MLE estimate of $\lambda$, we then subtracted $\lambda$ from $E(\hat{\lambda})$:

$$Bias(\hat{\lambda}) = E(\hat{\lambda}) - \lambda = \frac{\lambda(n-2)}{n-1} - \lambda = \frac{-\lambda}{n-1}.$$

To construct an unbiased estimator from this, we found a multiplier $C$ that satisfies

$$E(C\hat{\lambda}) = \lambda.$$

Substituting $E(\hat{\lambda})$,

$$C\frac{\lambda(n-2)}{n-1} = \lambda$$

and then solving for $C$ we obtain the unbiased estimator as,

$$\hat{\lambda}_{cor} = \hat{\lambda}\frac{n-1}{n-2}.$$

Empirically, O'Meara and Beaulieu (2024) showed that applying a $n/(n-1)$ correction to the Yule estimator worked well in removing the bias and we can show why

this might be case here. If instead of conditioning on the number of tips we consider a Yule process that grows for a fixed time, $t$, and the tip number $n$ varies, Kendall (1949) derived the variance of the estimator as

$$Var(\hat{\lambda}) = \frac{\lambda^2}{2(e^{\lambda t} - 1)}$$

This expression describes variation among replicate trees of equal age $t$ but differing sizes, rather than among trees of fixed $n$. Since the bias-corrected estimator is obtained by multiplying the original by a constant factor, its variance scales by the square of that factor. Thus,

$$Var(\hat{\lambda}_{cor}) = c^2 \frac{\lambda^2}{2(e^{\lambda t} - 1)}.$$

Although this variance pertains to the fixed-time case, it illustrates how different small-sample corrections (e.g., $c = \frac{n}{n-1}$ versus $c = \frac{n-1}{n-2}$) affect the bias–variance trade-off: the former yields slightly lower variance but introduces bias under the fixed $n$ likelihood.

## 4.2 Bias correction for general birth–death models

While analytical bias corrections are tractable for the Yule process, they are not for the general birth–death process where extinction is present. To address this, we developed a symbolic regression approach to search for approximate transformations of the estimated parameters that systematically reduce bias. Before applying this method to general birth–death models, we first validated it under the Yule process to ensure that it could recover known analytical corrections (see above). We then extended the approach to simulations under a broad range of birth–death scenarios, where closed-form expectations are unavailable.

### 4.2.1 Validating symbolic regression with the Yule model

To validate the symbolic regression approach, we simulated 250,000 Yule trees that span a wide range of clade ages and speciation rates. To do so, we first sampled 250,000 pairs of points for clade age and speciation rate ($\lambda$) using a Latin hypercube sampling design, drawing each variable independently from a uniform distribution over the interval [0,5] and [0,0.5], respectively. For each combination of clade age and speciation rate, trees were simulated and those with $n = 2$ were discarded. This yielded 140,449 trees, for which rates were estimated using the standard Yule likelihood conditioned on $n > 2$.

Using the R package `gramEvol` (Noorian et al. 2016) we defined a grammar that allows transformations and combinations of relevant variables using basic algebraic operations and mathematical functions. We included the following variables: $\hat{\lambda}$ is the likelihood estimate for speciation conditioning only on crown survival; $n$ is the number of taxa in a given tree; $(n-1)/(n-2)$ and $n/(n-1)$ are potential sample size bias correction functions; $s$ is the total tree length ($s$), the sum of all edge lengths; $T$ is clade age; and a placeholder for constants. Each symbolic expression generated by the grammar was evaluated by comparing its predicted output across simulated trees to the known generating speciation rates. Specifically, the true rates were treated as predictors and the outputs of the symbolic expressions as responses in a robust linear regression. We settled on using a robust linear model with the Huber loss function to downweight the influence of outliers when evaluating how closely each expression recovered the true generating values: various other ways of creating loss functions, like using RMSE, were very sensitive to some of the extremely bad estimates from small trees. Transformations were selected based on minimizing squared deviations from the 1:1 relationship between predicted and true rates, identifying those that systematically reduced bias in the estimated parameters.

To explore the expression space, we implemented a custom exhaustive search function instead of using the default evolutionary search within `gramEvol`, allowing us to systematically evaluate all possible expressions up to a user-defined maximum expression depth, which we set to max.depth = 4.

To penalize for model complexity, we added the term $\Omega = \alpha \cdot \text{depth}$ to each cost score, where depth is the depth of the expression tree and $\alpha$ is a regularization parameter controlling the strength of the penalty. We tested a range of $\alpha$ values from 0 to 0.02 to determine how strongly one must penalize complexity before a simpler expression overtakes the true model. Smaller values of $\alpha$ indicate that only mild penalization is needed to favor simpler expressions, whereas larger values imply that even under strong penalization the more complex model remains the better fit. This approach allowed us to assess the robustness of the analytically derived correction under increasing levels of regularization.

Table 2 summarizes the raw scores across the top five expressions returned (out of 38,771). For most of the $\alpha$ range ($\alpha \in [0, 0.0169]$, covering 84.6% of the grid), the analytically derived correction $\hat{\lambda} \cdot \frac{n-1}{n-2}$ achieved the lowest penalized cost score, with a raw fit error nearly two orders of magnitude smaller than that of the uncorrected $\hat{\lambda}$. Only when $\alpha$ exceeded 0.0169, where the complexity penalty was nearly as large as the fit score itself, did the uncorrected $\hat{\lambda}$ briefly become favored (15.4% of the grid). However, the advantage was driven purely by penalization, not fit. Thus, the analytically derived correction remained the best-supported expression across a wide range of penalty weights, demonstrating that it is robust to regularization and not simply favored under weak complexity penalties.

Finally, as shown in Figure 1, the uncorrected estimates systematically underestimated the true value, with the fitted slopes falling below one. In contrast, the corrected estimates aligned with the 1:1 line throughout the entire range of simulated values,

**Table 2** Ranking of symbolic expressions for corrections to the Yule estimator.

| Rank | Expression | Depth | Raw Score |
|---|---|---|---|
| 1 | $\hat{\lambda} \cdot \frac{n-1}{n-2}$ | 2 | 0.00013 |
| 2 | $\hat{\lambda} \cdot \frac{n}{n-1}$ | 2 | 0.00066 |
| 3 | $\hat{\lambda} \cdot \frac{s}{n+T}$ | 4 | 0.00144 |
| 4 | $\sqrt{\hat{\lambda}} / \frac{n-1}{n-2}$ | 3 | 0.00197 |
| 5 | $\hat{\lambda} \cdot n \cdot \log(\frac{n-1}{n-2})$ | 3 | 0.00243 |

confirming that the analytically derived correction of the naive estimator not only eliminates bias but also produces robust and accurate speciation rate estimates.

### 4.2.2 Extending to general birth–death processes

Having confirmed that symbolic regression recovers the known Yule correction, we next applied the method to trees simulated under a wide range of birth–death processes. We sampled 500,000 sets of points for clade age, speciation rate ($\lambda$), and extinction fraction ($\varepsilon = \mu/\lambda$), again, using a Latin hypercube sampling design, drawing each variable independently from a uniform distribution over the interval [0,10], [0,1], and [0,1], respectively. Note that the simulated value for $\mu$ was based on multiplying $\lambda$ by $\varepsilon$. Trees were then simulated for each set of clade age, speciation rate, and extinction rate, discarding any tree where $n \leq 2$. Speciation ($\lambda$) and extinction ($\mu$) rates were then estimated using the standard birth–death likelihood conditioned on $n > 2$. However, rather than estimating $\lambda$ and $\mu$ directly, we first estimated two composite parameters – turnover ($\tau = \lambda + \mu$) and extinction fraction ($\varepsilon = \mu/\lambda$) – and then back-transformed these values to recover estimates of $\lambda$ and $\mu$. To ensure that maximum likelihood estimates reflected genuine maxima, we excluded simulation replicates in which the variance–covariance matrix derived from the Hessian returned negative variance for either parameter. In total, this resulted in 339,416 trees from which we were able to obtain robust estimates of $\lambda$ and $\mu$.

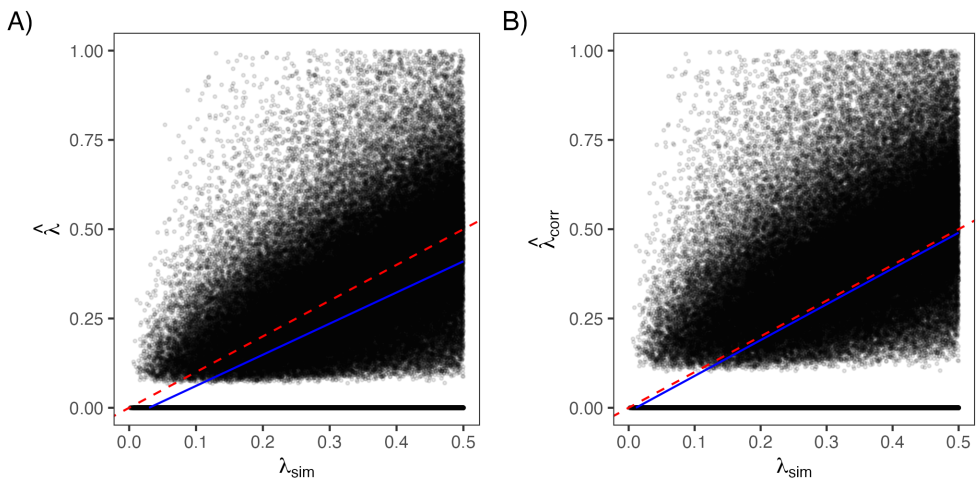


**Fig. 1** Effect of bias correction on estimates of the speciation rate $\lambda$ in Yule-simulated trees. Each panel shows estimated values of $\lambda$ plotted against the true simulated value $\lambda_{\text{sim}}$ across nearly 150,000 trees. (A) Uncorrected estimates ($\hat{\lambda}$) systematically underestimate the true value. (B) Applying the correction factor $\frac{n-1}{n-2}$ to $\hat{\lambda}$ improves accuracy, bringing estimates closer to the 1:1 line (dashed red). Blue lines show robust linear model fits.

The symbolic regression was then used to identify transformations that removed the downward biases exhibited by $\lambda$ and $\mu$ with respect to their known generating values. However, unlike with the Yule simulations, we did not have an analytically known "best" expression for the birth–death case. Instead, we identified the leading expressions by systematically varying the regularization weight $\alpha$ across a dense grid of 5,000 points. For each $\alpha$, we computed the total penalized score and recorded the expression with the lowest value. Whenever the highest-ranked expression changed, we defined an interval of $\alpha$ values over which the previous expression remained optimal, along with summary metrics including the minimum and maximum $\alpha$ and $\Omega$ at the endpoints. This procedure allowed us to trace how expression rankings shifted as the penalty for complexity increased, distinguishing corrections that were robust across broad $\alpha$ intervals from those that appeared only under extreme regularization. The search range for $\lambda$ was $\alpha \in [0, 10^{-3}]$ and $\alpha \in [0, 10^{-1}]$ for $\mu$. These bounds ensured that all regions with

**Table 3** Candidate expressions for speciation ($\lambda$) and extinction ($\mu$) rates under a birth–death model, ranked by their raw score.

| Parameter | Expression | Depth | Raw Score | $\alpha_{\min}$ | $\alpha_{\max}$ | $\Omega_{\min}$ | $\Omega_{\max}$ |
|---|---|---|---|---|---|---|---|
| $\lambda$ | $\hat{\lambda} \cdot \frac{n}{n-1}$ | 2 | 0.00004 | 0.00000 | 0.00078 | 0.00000 | 0.00157 |
| $\lambda$ | $\hat{\lambda}$ | 1 | 0.00083 | 0.00079 | 0.00100 | 0.00079 | 0.00100 |
| $\mu$ | $(\hat{\mu} + \hat{\lambda}) \cdot \varepsilon$ | 4 | 0.00175 | 0.00000 | 0.01156 | 0.00000 | 0.04625 |
| $\mu$ | $\hat{\mu}/\frac{n-1}{n-2} * 2$ | 3 | 0.01329 | 0.01156 | 0.02044 | 0.03469 | 0.06133 |
| $\mu$ | $\hat{\mu} * 2$ | 2 | 0.03373 | 0.02044 | 0.04767 | 0.04089 | 0.09534 |
| $\mu$ | $\hat{\varepsilon}$ | 1 | 0.08139 | 0.04767 | 0.10000 | 0.04767 | 0.10000 |

non-negligible penalized score were explored while avoiding oversampling in irrelevant ranges.

Table 3 summarizes the main results of symbolic regression for $\lambda$ and $\mu$ under the birth–death model. For $\lambda$, the expression $\hat{\lambda} \cdot \frac{n}{n-1}$ consistently achieved the lowest penalized cost over nearly the entire range of regularization weights ($\alpha \in [0, 0.0008]$). This correction reduced the raw fitting error by more than an order of magnitude relative to the uncorrected $\hat{\lambda}$, and remained favored until the complexity penalty became comparable in magnitude to the fit score itself ($\alpha \approx 0.001$). Beyond this point, the uncorrected $\hat{\lambda}$ was marginally preferred, though the improvement was driven solely by penalization rather than by fit. Thus, a correction similar to one found for $\lambda$ under the Yule process also emerged as the most stable and parsimonious adjustment in the birth–death case.

For $\mu$, the symbolic regression yielded a more complex pattern of potential corrections. The best-performing expression across most regularization weights ($\alpha \in [0, 0.0116]$) was $(\hat{\mu} + \hat{\lambda}) \cdot \hat{\varepsilon}$, indicating that extinction rate bias depends jointly on both the turnover rate and the extinction fraction. This relationship suggests that uncertainty in $\hat{\lambda}$ affects $\hat{\mu}$, reflecting their joint dependence in the likelihood rather than a sample-size effect as in seen with $\lambda$. As the regularization penalty increased, simpler multiplicative adjustments such as $\hat{\mu}/\frac{n-1}{n-2} * 2$ or $2\hat{\mu}$ were briefly favored. At the strongest penalties ($\alpha > 0.05$), the regression collapsed to $\hat{\varepsilon}$, demonstrating that

extinction bias could not be captured by a single-term correction. Together, these results highlight, perhaps unsurprisingly, that $\mu$ bias is structurally more complex than $\lambda$ bias.

### 4.2.3 Bias corrections and derived diversification parameters

Having derived bias corrections for $\lambda$ and $\mu$, we next examined how these adjustments affect other diversification parameters — specifically, turnover $(\lambda + \mu)$ and net diversification $(\lambda - \mu)$ — that are commonly used to summarize macroevolutionary dynamics. To evaluate whether the symbolic regression–based corrections for $\lambda$ and $\mu$ fully removed systematic bias, we also derived *expectation-based corrections* by regressing the corrected rate estimates on their generating values across simulations. These expectation-based corrections serve as an empirical calibration step: while the symbolic regression addresses structural bias, the regression-based mapping removes the small residual bias that remains when corrected estimates are compared directly to their true values. This provided linear mappings of the form

$$\lambda' = \frac{\lambda_{\text{corr}} - a_\lambda}{b_\lambda}, \quad \mu' = \frac{\mu_{\text{corr}} - a_\mu}{b_\mu},$$

where $a$ and $b$ are the intercept and slope from the corresponding regression fits. These transformations effectively invert any small residual bias remaining after the symbolic regression corrections, yielding expected values centered on the true parameters. Turnover and net diversification were then recomputed using both the directly corrected and expectation-based rates to assess how these refinements influence composite diversification parameters.

When turnover was calculated directly from the corrected $\lambda$ and $\mu$, the relationship with simulated values was nearly unbiased, with an intercept near zero (-0.0006) and a slope slightly above one (1.056). This reflects the partial cancellation of small residual biases in $\lambda$ and $\mu$, where both corrected estimators slightly under predict at higher

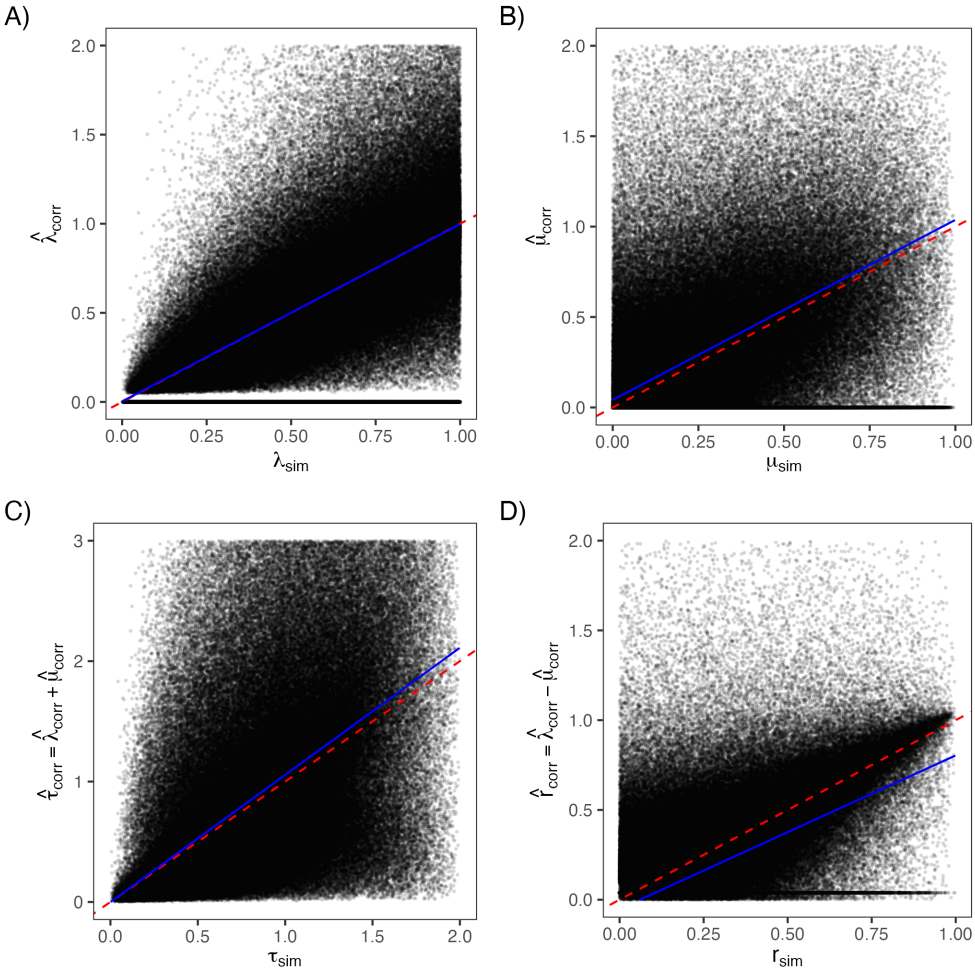


**Fig. 2** Relationships between simulated ("sim") and corrected estimates of (A) speciation rate, (B) extinction rate, (C) turnover rate, and (D) net diversification. Each panel shows the empirical regression line (dark blue) and the 1:1 line (dashed red). The close alignment between the two lines in panels (A-C) indicates that bias corrections for $\lambda$ and $\mu$ largely remove systematic deviation, while panel (D) shows that net diversification remains slightly underestimated at high true values due to small asymmetries between $\lambda$ and $\mu$ corrections.

values but in similar proportion. Using the expectation-based corrections yielded a comparable fit (slope = 1.061, intercept = -0.045), indicating that turnover is largely robust to modest differences in the individual rate corrections (Fig. 2).

In contrast, net diversification ($\lambda - \mu$) remained more sensitive to differential bias between $\lambda$ and $\mu$. When computed directly from the corrected rates, the slope was

notably below one (0.856), and the intercept negative (-0.052), reflecting persistent underestimation at higher true net-diversification values. Applying the expectation-based correction slightly improved this relationship (slope = 0.861, intercept = -0.013) but did not fully remove the bias (Fig. 2).

On the whole, these results show that turnover behaves as a stable composite of $\lambda$ and $\mu$, whereas net diversification is more strongly affected by small asymmetries in their respective biases, resulting in a consistent flattening of its regression slope. This is primarily due to $\mu$ being slightly overestimated relative to $\lambda$. These patterns are consistent with the algebraic expectations derived from the corrected estimators, where net diversification inherits asymmetric bias from the difference between $\hat{\lambda}$ and $\hat{\mu}$ expectations. Because the magnitude of this difference increases with true net diversification, the resulting slope of estimated versus simulated values remains systematically below one, even when both underlying rate corrections are nearly unbiased.

# 5 Discussion

The results presented here highlight additional statistical and structural sources of bias in the estimation of diversification rates under the birth-death framework. With respect to estimator bias, we showed analytically that for the Yule process ($\mu = 0$) the maximum likelihood estimator for $\lambda$ is biased downward. Our empirically driven approach to approximate the downward bias for $\lambda$ and $\mu$ (when $\mu > 0$) in the birth-death model identified functional expressions that minimized the error between the generating and estimated values. The best-performing correction for $\lambda$ in a birth-death model was similar to the correction in the Yule model. In contrast, $\mu$ exhibited a more complex pattern, where the bias depended jointly on turnover ($\lambda + \mu$) and extinction fraction ($\mu/\lambda$).

Both $\lambda$ and $\mu$ are underestimated by their maximum likelihood estimators, but the bias is not symmetrical. After applying the correction, $\lambda$ closely follows the 1:1

relationship with its generating values, albeit slightly too low, whereas $\mu$ tends to remain slightly overestimated. Because turnover is the sum of $\lambda$ and $\mu$ it seems that the opposing biases in the two parameters largely offset each other, producing nearly unbiased estimates. In contrast, net diversification, defined by the difference in $\lambda$ and $\mu$ is even less well estimated than either alone: a too large number is subtracted from a too small number. As a result, empirical results that summarize diversification using turnover are likely to be less biased than those relying on net diversification, especially in small clades and/or when the underlying extinction fraction is high. In studies of modern taxa net diversification rate differences remain a perennial favorite estimator, despite some encouragement to use other parameters for biological realism (Beaulieu and O'Meara 2016; Vasconcelos et al. 2022b). In paleontology, there is considerably more enthusiasm for using net turnover, or its reciprocal, average species lifespan (Vrba 1993).

With respect to structural bias, conditioning the likelihood on a censored tree set changes the probability space on which the model is defined. Because two-taxon "cherry trees" lack information to jointly identify distinct speciation and extinction rates, conditioning on both crown survival and $n > 2$ restricts the likelihood to trees that contain information about both parameters. We found that this conditioning scheme reduced inflation of rate estimates when small clades were censored. The analytical basis for the limited information in cherry trees is reflected in their likelihood surface.

These findings have direct implications for birth-death models that allow for heterogeneity in rates across a given tree. To estimate birth, death, or net diversification rates with even a hope of similarity to the truth one must analyze clades of at least three taxa; this holds true even if the clade of interest is inside a much larger tree. Methods that try different discrete regime paintings on the tree, like BAMM (Rabosky 2014), ClaDS (Maliet et al. 2014), and MEDUSA (Alfaro et al. 2009) could in theory

be configured such that any examined regimes have the requisite amount of information. This can also be used for models that associate states with rates. For example, BiSSE could be limited to only trying distinct rates for states present in at least three tips. Difficulties come with methods that do not allow such discrete limitations. For example, with MiSSE (Vasconcelos et al. 2022a), rates are associated with hidden states that are reconstructed at tips and at nodes: there is no fixed painting of regimes, only statistical inferences about likelihoods.

Note that Bayesian methods are not an automatic fix to these issues. If a likelihood estimator is biased, weighting it by a Bayesian prior does not eliminate this bias for usual priors, though one could attempt to create various priors with biases or constraints to "correct" the bias. For both likelihood and Bayesian approaches, the best approach might be to estimate parameters as normal (subject to constraints about tree size) but then use the corrections above to get more unbiased estimates.

Likelihood estimators remain a powerful tool for inferring diversification dynamics, but their biases are most apparent in small trees or small subclades within larger trees. Recognizing and correcting these biases is essential for accurate rate estimation and for the broader application of birth–death models in comparative analyses.

**Acknowledgments.** We thank James Boyko, Nathanael Walker-Hale, and Andrew Alverson for discussions and edits that have improved the content and presentation.

**Funding.** We acknowledge the National Science Foundation grants DEB-1916558 (JMB) and DEB-1916539 (BCO).

**Data Availability.** All scripts and substantial outputs are available at https://github.com/thej022214/BeaulieuOMeara_BDBiasCode.

# Declarations

**Conflict of Interest.** The authors declare that they have no conflict of interest.

# A Identifiability limits in cherry trees

To explore whether cherry trees contain sufficient information to separately estimate speciation and extinction rates, we focus on the constant-rate birth-death likelihood in Eq. 3, which conditions on the survival of the crown and relies on a product over the edges as opposed to internal nodes. By reducing this expression to the special case of a two-taxon tree, we can isolate the contribution of a single observed bifurcation event and ask whether the data it contains can, in principle, identify both parameters. This simplified setting serves as a useful conceptual test case: even with a likelihood that can be evaluated, do cherry trees carry enough information to distinguish between speciation and extinction?

First, since we are examining a cherry tree, we can set $m = 1$, $t_{i,r} = T$ and $t_{i,t} = 0$ in Eq. 3. Focusing on the product term on the left-hand side, we can simplify this expression based on what each edge of the two edges in a cherry tree contributes:

$$e^{-(\lambda-\mu)(T-0)} \frac{(\lambda - e^{-(\lambda-\mu)\cdot 0}\mu)^2}{(\lambda - e^{-(\lambda-\mu)T}\mu)^2}$$

which simplifies further to,

$$e^{-(\lambda-\mu)T} \frac{(\lambda-\mu)^2}{(\lambda - e^{-(\lambda-\mu)T}\mu)^2}.$$

There are two such terms, one for each edge, so we square this

$$\left[e^{-(\lambda-\mu)T} \frac{(\lambda-\mu)^2}{(\lambda - e^{-(\lambda-\mu)T}\mu)^2}\right]^2$$

and this simplifies to,

$$e^{-2(\lambda-\mu)T} \frac{(\lambda-\mu)^4}{(\lambda - e^{-(\lambda-\mu)T}\mu)^4}.$$

Since $m = 1$, the right hand term in Eq. 3 that includes the conditioning scheme simplifies to:

$$\frac{(\lambda - e^{-(\lambda-\mu)T}\mu)^2}{(\lambda-\mu)^2}.$$

The full likelihood therefore is,

$$L(\lambda,\mu) = e^{-2(\lambda-\mu)T}\frac{(\lambda-\mu)^4}{(\lambda - e^{-(\lambda-\mu)T}\mu)^4}\cdot\frac{(\lambda - e^{-(\lambda-\mu)T}\mu)^2}{(\lambda-\mu)^2}.$$

If we multiply the product term and the conditioning term together and further simplify, we obtain a final likelihood expression for a cherry tree as

$$L(\lambda,\mu) = e^{-2(\lambda-\mu)T}\cdot\frac{(\lambda-\mu)^2}{(\lambda - e^{-(\lambda-\mu)T}\mu)^2}$$

Finally, we convert this function to a log-likelihood:

$$\ell(\lambda,\mu) = -2(\lambda-\mu)T + 2\log(\lambda-\mu) - 2\log(\lambda - e^{-(\lambda-\mu)T}\mu). \tag{8}$$

To assess identifiability, we examined the shape of the log-likelihood surface by taking the partial derivative of Eq. 8 with respect to $\lambda$, holding $\mu$ constant. The derivative is:

$$\frac{\partial\ell}{\partial\lambda} = -2\left(\frac{\mu T\mathrm{e}^{-T(\lambda-\mu)}+1}{\lambda-\mu\mathrm{e}^{-T(\lambda-\mu)}} - \frac{1}{\lambda-\mu} + T\right)$$

This expression describes how the likelihood changes with $\lambda$, and allowed us to determine whether the function has a maximum, is flat, or decreases monotonically.

We evaluated the first derivative of Eq. 8 on a grid of ($\lambda$,$\mu$) values at fixed time depths, $T$=1 and $T$=10. For both depths examined, the surface of $\partial\ell/\partial\lambda$ is strictly nonzero across the parameter space, except at the singularity $\lambda = \mu$ (Fig. 3). As $T$ increases, the magnitude of the derivative increases, indicating that the log-likelihood becomes steeper with respect to $\lambda$. This confirms that, for any fixed value of $\mu$, the

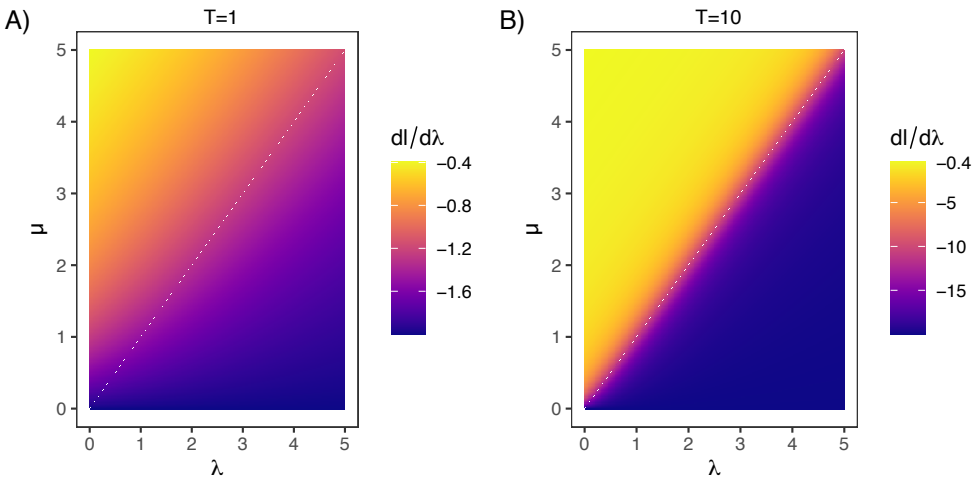


**Fig. 3** Heatmap plots of the partial derivative of the two-taxon log-likelihood function with respect to the speciation rate, $\lambda$, across a grid of $\lambda$ and $\mu$ values. Colors represent the value of the derivative, with darker shades indicating steeper gradients. The left panel shows results at time depth $T$=1, and the right panel at $T$=10. A green contour line marks where the derivative equals zero. In both cases, the derivative is strictly negative throughout the parameter space except along the singularity at $\lambda$=$\mu$, suggesting the likelihood estimate for $\lambda$ must be at the lower bound, zero.

log-likelihood is a strictly decreasing function of $\lambda$, with no stationary point in the $\lambda$ direction. The absence of a maximum implies that $\lambda$ and $\mu$ are not jointly identifiable from a single bifurcation in a two-taxon tree. Although a likelihood surface can be computed, it contains insufficient information to distinguish between alternative combinations of speciation and extinction rates.

The derivative with respect to $\mu$ is more nuanced. It is given by:

$$\frac{\partial \ell}{\partial \mu} = \frac{2\lambda \left(\mathrm{e}^{T(\mu-\lambda)} - T\left(\mu-\lambda\right) - 1\right)}{\left(\mu-\lambda\right)\left(\mu \mathrm{e}^{T(\mu-\lambda)} - \lambda\right)}$$

Unlike $\partial\ell/\partial\lambda$, which is strictly nonzero across the parameter space, $\partial\ell/\partial\mu$ exhibits regions where it approaches zero. These near-zero values occur when extinction far exceeds speciation (i.e., $\mu \gg \lambda$), particularly at deeper time depths ($T$=10)(Fig. 4). In these regions, the log-likelihood surface is relatively flat in the $\mu$ direction, indicating weak sensitivity to changes in $\mu$ and thus weak identifiability. Importantly, the

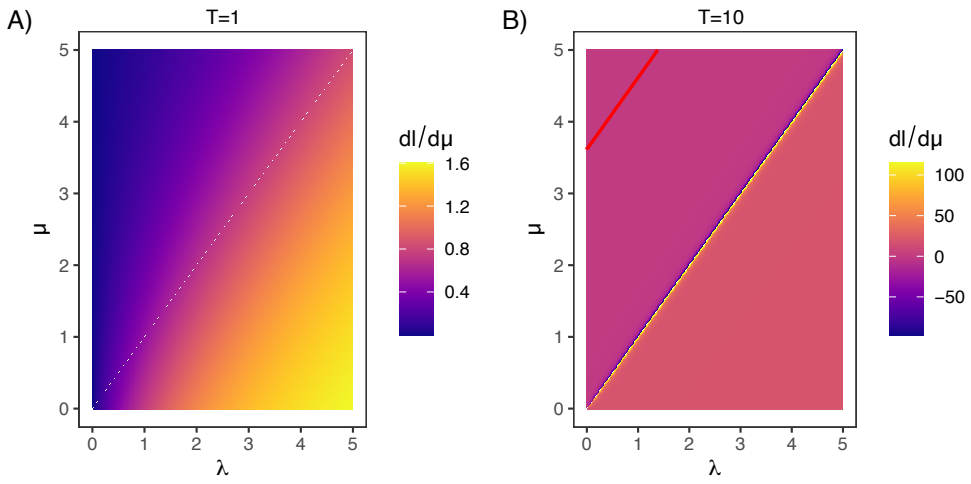


**Fig. 4** Heatmap plots of the partial derivative of the two-taxon log-likelihood function with respect to the extinction rate, $\mu$, across a grid of $\lambda$ and $\mu$ values. Colors indicate the magnitude of the derivative, and the dark red contour line highlights where the derivative equals zero. The left panel corresponds to a time depth of $T$=1, and the right panel to $T$=10. In contrast with the derivative with respect to $\lambda$, the derivative with respect to $\mu$ includes regions where the surface is nearly flat (i.e., derivative near zero), particularly when $\mu \gg \lambda$. These regions become more prominent for older trees, indicating that changes in $\mu$ have diminishing influence on the likelihood surface in those parts of parameter space. This pattern suggests that extinction rates may be weakly identifiable in some scenarios, though no true maximum is present, reinforcing the overall conclusion of non-identifiability from two-taxon trees.

derivative also flattens in the region where $\lambda > \mu$, which is often more biologically plausible.

Together, these results demonstrate that the likelihood function for a two-taxon tree lacks a maximum in either parameter dimension, confirming that speciation and extinction rates cannot be jointly identified from a single bifurcation.

# References

Alfaro ME, Santini F, Brock C, et al (2009) Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proceedings of the National Academy of Sciences, USA 106:13410–13414. https://doi.org/10.1073/pnas.0811087106

Bailey N (1964) The elements of stochastic processes with applications to the natural sciences. Wiley, New York

Beaulieu JM, O'Meara BC (2016) Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Systematic Biology 65:583–601. https://doi.org/10.1093/sysbio/syw022

Bokma F (2008) Bayesian estimation of speciation and extinction probabilities from (in)complete phylogenies. Evolution 62:2441–2445. https://doi.org/10.1111/j.1558-5646.2008.00455.x

Etienne RS, Haegeman B, Stadler T, et al (2012) Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. Proceedings of the Royal Society, B 279:1300–1309. https://doi.org/10.1098/rspb.2011.1439

Kendall DG (1949) Processes and population growth. Journal of the Royal Statistical Society Series B (Methodological) 11:230–282. https://doi.org/10.1111/j.2517-6161.1949.tb00032.x

Maddison WP, Midford PE, Otto SP (2007) Estimating a binary character's effect on speciation and extinction. Systematic Biology 56:701–710. https://doi.org/10.1080/10635150701607033

Maliet O, Hartig F, Morlon H (2014) A model with many small shifts for estimating species-specific diversification rates. Nature Ecology and Evolution 3:1086–1092. https://doi.org/10.1038/s41559-019-0908-0

McKenzie A, Steel M (2000) Distributions of cherries for two models of trees. Mathematical Biosciences 164:81–92. https://doi.org/10.1016/S0025-5564(99)00060-7

Morlon H, Parsons TL, Plotkin JB (2011) Reconciling molecular phylogenies with the fossil record. Proceedings of the National Academy of Sciences, USA 108:16327–16332. https://doi.org/10.1073/pnas.1102543108

Nee S (2001) Inferring speciation rates from phylogenies. Evolution 55:661–668. https://doi.org/10.1111/j.0014-3820.2001.tb00801.x

Nee S, Mooers AO, Harvey PH (1992) The tempo and mode of evolution revealed from molecular phylogenies. Proc Natl Acad Sci 89:8322–8326. https://doi.org/10.1073/pnas.89.17.8322

Noorian F, de Silva A. M., Leong PHW (2016) gramevol: Grammatical evolution in r. Journal of Statistical Software 71:1–26. https://doi.org/10.18637/jss.v071.i01

O'Meara BC, Beaulieu JM (2024) Noise leads to the perceived increase in evolutionary rates over short time scales. PLOS Computational Biology 20:e1012458. https://doi.org/10.1371/journal.pcbi.1012458

Rabosky DL (2006) Likelihood methods for detecting temporal shifts in diversification rates. Evolution 60:1152–1164. https://doi.org/10.1111/J.0014-3820.2006.TB01194.X

Rabosky DL (2009) Heritability of extinction rates links diversification patterns in molecular phylogenies and fossils. Systematic Biology 58:629–640. https://doi.org/10.1093/sysbio/syp069

Rabosky DL (2014) Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. PLOS One 9:e89543. https://doi.org/10.1371/journal.pone.0089543

Rabosky DL, Lovette IJ (2008) Density-dependent diversification in north american wood warblers. Proceedings of the Royal Society, B 275:2363–2371. https://doi.org/10.1098/rspb.2008.0630

Slowinski JB, Guyer C (1993) Testing whether certain traits have caused amplified diversification: an improved method based on a model of random speciation and extinction. American Naturalist 142:1019–1024. https://doi.org/10.1086/285586

Stadler T (2013) How can we improve accuracy of macroevolutionary rate estimates? Systematic Biology 62:321–329. https://doi.org/10.1093/sysbio/sys073

Steel M, Mooers A (2010) The expected length of pendant and interior edges. Applied Mathematical Letters 23:1315–1319. https://doi.org/10.1016/j.aml.2010.06.021

Vasconcelos T, O'Meara BC, Beaulieu JM (2022a) A flexible method for estimating tip diversification rates across a range of speciation and extinction scenarios. Evolution 76:1420–1433. https://doi.org/10.1111/evo.14517

Vasconcelos T, O'Meara BC, Beaulieu JM (2022b) Retiring 'cradles' and 'museums' of biodiversity. The American Naturalist 199:194–205. https://doi.org/10.1086/717412

Vrba ES (1993) Turnover-pulses, the red queen, and related topics. American Journal of Science 293:418–452. https://doi.org/10.2475/ajs.293.A.418

Yule GU (1924) A mathematical theory of evolution, based on the conclusions of dr. j.c. willis. Philos Trans Roy Soc London Ser B 213:21–87. https://doi.org/10.1098/rstb.1925.0002