Title: Noise leads to the perceived increase in evolutionary rates over short time scales

Authors: Brian C. O'Meara[a,1] and Jeremy M. Beaulieu[b,1]

[a]Department of Ecology and Evolutionary Biology, University of Tennessee; Knoxville, Tennessee, 37996-1610 USA.

[b]Department of Biological Sciences, University of Arkansas; Fayetteville, Arkansas, 72701 USA.

[1]To whom correspondence may be addressed: Email: bomeara@utk.edu or jmbeauli@uark.edu

Author contributions: B.C.O. and J.M.B. contributed equally.

Competing Interest Statement: The authors declare that they have no competing interests.

Keywords: evolutionary rates, measurement error, hyperbolic trends, macroevolution, microevolution

This PDF file includes:

Main Text
Figures 1 to 2
Supplementary Figures S1-S10
Supplementary Tables S1
Supplementary Movie S1

Abstract

Interpretations of biodiversity patterns across timescales necessarily assume that fundamental processes of evolution do not change over time. While drift, selection, mutation, speciation may vary in magnitude, there is no need to postulate new mechanisms. A recurring pattern across a variety of biological datasets, from genomes to the fossil record, has shown evolutionary rates increasing toward the present or over shorter time scales, indicating potentially fundamental but unknown processes unifying evolutionary timescales. Here we introduce a set of novel models that assess the relationship between rate and time and demonstrate that these patterns are statistical artifacts of time-independent errors present across ecological and evolutionary datasets, which produce hyperbolic patterns of rates through time. Specifically, no matter how time-independent error and measurements of evolutionary change are distributed, estimated rates exhibit a hyperbolic pattern when plotted against time. Our findings are generally consistent with uniform rates over time but pose new challenges for the rate estimates foundational to other macroevolutionary models.

Introduction

The diverse array of processes across organisms, time, and space serves as a testament to the dynamic nature of evolution, where few universal rules persist. Consequently, biologists take immediate notice when broad patterns manifest across a diverse set of lineages. One such recurring pattern is that evolutionary rates exhibit an exponential increase towards the present or over shorter time scales, highlighting a potentially new fundamental principle in how life evolves. The consistency of this pattern across dimensions of diversity, from genomes to the fossil record (*1-6*), highlight its universality and point to the need for new conceptual bridges connecting disparate timescales of evolutionary change (*7*), despite past work showing potential artifactual causes (*8*). It also challenges the long-held view that the processes playing out in the past generally behave similarly in the present. Reconciling this pattern offers the potential for new insights into the underlying mechanisms that shape biodiversity.

Here we introduce a set of novel models that assess the relationship between rate and time that includes hyperbolic and linear functions of time. We show that one potentially crucial but largely overlooked factor (but see *9, 10*) affecting rate patterns is the impact of empirical errors inherent in rate estimates, and that this apparently has driven the pattern observed across extinction risk, trait evolution, and diversification rate estimates over time, based on model fitting and new randomization tests. We suspect that questions about rates changing over time scales can only be properly examined, across a variety of fields, once the biasing effect of uncertainty is taken into consideration.

Results

Impish Problems. The concern over spurious correlations between ratios and shared factors has a longstanding history in the statistical literature, dating back to Pearson's pioneering work more

than a century ago (*11*). Pearson illustrated this problem by recounting a biologist's study of skeletal measurements, specifically femur and tibia length normalized as fractions of the humerus length. Expecting a high correlation to validate correct groupings, the biologist was unaware that an "imp" had randomly shuffled bones between specimens. Surprisingly, the high correlation would persist even after this randomization, revealing that such spurious relationships arise from inherent properties of the variables rather than indicating meaningful biological connections (*11-12*).

Similarly, when plotting an evolutionary rate against its corresponding denominator, time, the representation becomes a plot of time against its reciprocal (i.e., 1/time), resulting in a relationship that is negatively biased (*5, 6*). In fact, if the numerator is held constant, the slope on a log-log scale must be -1.0 (*1, 14*). However, the numerator in an evolutionary rate, which quantifies the absolute amount of evolutionary change, is unlikely to remain constant across all timescales[6], as seen empirically (*15*). Thus, the observed empirical pattern of rates increasing towards the present may not be the result of a universal artifact, but instead may depend on nuanced conditions or specific contexts within evolutionary processes, something tested here.

Components of Evolutionary Rate Patterns Through Time. In the context of evolutionary biology, a rate is a measure of evolutionary change per unit of time,

$$r(t) = \frac{x(t)}{t}.$$

Evolutionary change, $x(t)$, encompasses a variety of measures, including number of nucleotide substitutions in DNA (*16*), the number of transitions between discrete phenotypes (*17*), speciation and extinction events (*18*), or the absolute change in a continuous trait after some interval of time (*19*). The specifics of the different types of evolutionary rates and how they are estimated are intricate. For instance, one rate may assume changes follow a Poisson distribution

on a nested tree structure (e.g., substitution rates), while others may assume trait changes are drawn from a normal distribution with the rate being the measure of the variance (e.g., Brownian motion). Nevertheless, at its core, an evolutionary rate reduces to some measure of change over time.

While it is common to perceive uncertainty and error as obscuring patterns rather than contributing to them, these uncertainties and errors may be key drivers of the repeated pattern of rates increasing towards the present. For illustrative purposes, we focus on the simplest approach for calculating rates of morphological evolution to demonstrate the impact of biases and errors on the relationship between rate and time. Measurements are assembled as a set of paired comparisons that differ in some trait value (e.g., body size) measured at two separate time points, with ratio of the differences in trait and time being an estimate of the rate,

$$\hat{r}(t) = \frac{|x_2 - x_1|}{t_2 - t_1}, t_2 > t_1.$$

Here x_1 and x_2 are the initial and final values of the trait measurement between two time points, t_1 and t_2, respectively. Expressing a rate in such a way is similar to the "darwin," a widely used unit of evolutionary change first defined by Haldane (*20*).

Various phenomenological patterns can describe how rates change through time. The simplest is that $\hat{r}(t)$ exhibits a constant rate of change, maintaining an expected value regardless of the measurement interval. For example, a mutation rate resulting from DNA copying errors might be anticipated to be consistent when scaling the number of mutations by time within clades of different ages. In other words, even though a clade with a crown age of 5 million years is expected to have far fewer mutations than a 50-million-year-old clade, on a per-time basis the rate will be identical. The biological rationale for such consistency in rate is rooted in the assumption that processes governing rates operate similarly in the past as they do in the present –

125 a foundational premise in much of evolutionary biology. On the other hand, a significant secular
126 change in the mutation rate might result from a general enhancement in DNA copying or repair
127 mechanisms. Thus, we might expect differences between x_2 and x_1 to decrease near the present.

128 Measurements of x_i often reflect the mean for a species and, therefore, represent an
129 estimate of the true value. If each measurement estimate, x_i, carries some level of noise attributed
130 to factors like finite sample size or measurement error, then $\widehat{x_i} = x_i + \varepsilon_i$, with ε_i denoting an
131 error component that can lead to overestimation or underestimation of the overall rate. Thus, we
132 can express the rate estimate as including error:

133 $$\hat{r}(t) = \frac{|(x_2 + \varepsilon_2) - (x_1 + \varepsilon_1)|}{t_2 - t_1}, t_2 > t_1.$$

134 The impact of an explicit error component is readily seen if we assume no differences in the trait
135 measurements, such that x_2 - x_1=0. This reduces a rate estimate as being the difference between
136 the two error terms,

137 $$\hat{r}(t) = \frac{|\varepsilon_2 - \varepsilon_1|}{t_2 - t_1}, t_2 > t_1.$$

138 Since these errors are not inherently time-dependent (e.g., sequencing errors in a DNA analysis),
139 there are no *a priori* reasons for the magnitude of the error in a 5-million-year-old clade would
140 necessarily be any greater or lesser than the magnitude in a 50-million-year-old clade. Thus, a
141 hyperbola will invariably result when the ratio of these differences in error are scaled by time,
142 and then plotted against time. The impact of uncertainty and error is particularly pronounced in
143 models such as those applied to DNA evolution, where differences between taxa are presumed to
144 be changes occurring along branches in a phylogenetic tree. Sequencing errors independent of
145 time can introduce not only an upward bias in the rate but also a bias favoring a hyperbolic
146 relationship between rate and time (*2*). If a constant rate exists, such as from a molecular clock or

net diversification, at large times the magnitude of this rate could be much larger than the magnitude of error divided by time (Fig. 1).

To assess the relative contribution of constant, hyperbolic, and linear functions towards a rate estimate over time, we derived a novel least-squares approach. Our method allowed us to predict changes in observed evolutionary rates sampled through time by minimizing the logarithm of the residual sum of squares between the predicted and observed values. We derived a model that, at its most complex, is given as,

$$\hat{r}(t) = \frac{h}{t} + mt + b.$$

Here h denotes the hyperbolic component [in units of $x(t)$], m is scalar modulating the effect of time up and down linearly [in units of $x(t)t^{-2}$], and b is a constant base rate [in units of $x(t)t^{-1}$]. The full model assumes all three components have impacted the fitted value for $\hat{r}(t)$ in some way. We also fit a range of restrictions to this model where one or more of the parameters are set to zero. For example, restricting h=m=0 is a model in which the rates would be inferred to be constant through time, as one would expect from a process like a molecular clock. The maximum likelihood fit of a model was assessed using the logarithm of the residual sum of squares and first converting this into a measure of the model variance and then into a log-likelihood. To facilitate comparisons across a set of models, we converted the log-likelihood into an Akaike Information Criterion score (AIC); we also assessed confidence regions (e.g., Figs. S1-6) around each of the parameters.

Revisiting Empirical Cases of Age-Dependent Rates. We analyzed five empirical datasets that encompass diverse data types, including substitution rates [Ho *et al.* (*2*)], rates of body size evolution from extant and extinct taxa [measured in darwins; Gingerich (*1*), Uyeda *et al.* (*15*)], speciation rates from phylogenetic trees of extant species [Henao Diaz *et al.* (*3*)], and

contemporary species extinction rates (Barnosky *et al.* (*21*). We also included a dataset of birth rates obtained from 25,000 trees simulated under a Yule process (i.e., pure birth) with the same rate (i.e., 0.10 birth Myr^{-1}) and clade ages drawn from a uniform distribution. Estimates of the birth rates included a correction to account for the biased estimates for smaller tree sizes. This simulated dataset served as a type of control given that the trees were generated without error and thus the best fit model should generally appear as a horizontal line centered on the generating rate.

In all five empirical datasets, we observed a consistent negative trend in rates through time on a log-log scale (Fig. 1). This aligns with published results indicating higher rates closer to the present, declining as time increases. When we fit our least-squares model separately to each empirical dataset, we found, with one exception, that all were best fit by a model that included all three parameters (h, m, and b). The Henao Diaz *et al.* (*3*) data favored a model without a base rate ($b = 0$ speciation events Myr^{-1}), but there were three other models within 2 AIC units from the best fit model, including the full model, and all featured a positive hyperbolic parameter estimate ($\hat{h}$ = 1.71-2.07 speciation events). One consistent finding across all five empirical datasets was that the best fit model included a positive hyperbolic parameter estimate ($\hat{h} > 0$) as well as a non-zero linear component ($\hat{m} \neq 0$).

For the fitted rate values within each dataset, we decomposed the rate to reflect the proportion attributed to each of the three components in our least-squares model. This allowed us to assess how much of the fitted rate comes from each component as a function of time instead of relying solely on interpreting the parameters in the fitted model. This revealed, for example, which portions of the overall model fit were predominantly driven by a linear trend in the rates even in the presence of a non-zero hyperbolic component.

Decomposing the fitted rates in this way revealed that a significant portion of the relationship between rate and time in the empirical datasets is strongly influenced by the hyperbolic component (Fig. 1). This occurred despite model fits indicating that there was signal for both constant and linear components. Interestingly, in the case of body size evolution datasets of Gingrich (*1*) and Uyeda *et al.* (*15*), the hyperbolic component was replaced by the constant component, followed by a small remaining portion showing the linear component as having a substantial contribution to the rate (Fig. 1). Putting these linear trends in the context of the estimate of the *m* parameter, there is a signal for rates of body size evolution to decrease as time increases that contributes strongest to the deepest time points [Gingerich (*1*): $\widehat{m}$ = -0.00007 (CI: -0.00009:-0.00005) Δlog body mass per Myr^{-2}; Uyeda *et al.* (*15*): $\widehat{m}$ = -0.0005 (CI: -0.0006:-0.0005) Δlog body mass per Myr^{-2}]. Similarly, the speciation rates of Henao Diaz *et al.* (*3*) shifted from a predominant signal for the hyperbolic component to the linear component being the strongest contributor towards the rate for the oldest clades. The linear component indicated a trend of speciation rates increasing across these deep time scales [$\widehat{m}$ = 0.00002 (CI: -0.00001:0.00004) speciation events per Myr^{-2}; Fig. 1], though the confidence interval includes zero and at the maximum time only raises the speciation rate by 0.008 events per Myr^{-2}. Moreover, for all the empirical datasets, the simplest *h*00 model, which uses only a hyperbolic component, has only a slightly worse coefficient of determination than the best fitting models, even if there is a substantial difference in AIC (Table S1).

Notable exceptions were the datasets of Ho *et al*. (*2*) and Barnosky *et al.* (*21*), which showed no strong contribution by the linear component. In both cases, the hyperbolic component contributed strongest nearer the present, with the remainder mostly favoring constant rates. There was substantial uncertainty not just in the parameter estimates for these datasets (Figs S5-S6 and Table S1), but also which component contributed to the fitted rate generally, which we suspect

was due to the small sample sizes of these datasets (n=17 and n=11, respectively). In the pure birth simulation, the overall trend was a horizontal trend line driven almost exclusively by the constant component, as expected, which was centered near the generating rate in our simulation ($\hat{b}$ = 0.106 (CI: 0.103:0.109) birth events Myr^{-1} vs b_{true} = 0.10 birth events Myr^{-1}).

Deriving an appropriate null expectation. The strong contribution of the hyperbolic component does not rule out the possibility that a biological process could underlie the increasing rates towards the present. For example, for rates of body size evolution there might exist a finite range of viable body sizes for terrestrial vertebrates, with the potential for doubling over short time scales but reaching a fixed maximum numerator over longer periods. The traditional null expectation assumes no relationship between rate and time and is represented by a horizontal line. However, this null expectation makes a strong assumption that the amount of reconstructed evolutionary change, $x(t)$, is perfectly correlated with reconstructed time. We expect some degree of imperfect correlation, i.e., at its most extreme, the evolutionary change could be entirely random in relation to time. In such instances, we are plotting random rates against time, which manifests exactly as a hyperbola.

To generate a new empirical null, we developed a procedure similar to that of Sheets and Mitchell (*8*), we first separated each empirical rate into its numerator and denominator components. The numerator was obtained either directly from the original data, or by multiplying the empirical rate by their respective denominator of measured time. We then shuffled the numerators at random across a given empirical dataset and divided them by the unshuffled times to obtain a new set of rates and times. This not only follows the long tradition of using reshuffling to get null distributions when examining ratios against common factors (*11*), but also has the advantage of keeping the same distributions of times and amount of change as

242 the original data, which can matter for some of the biases here (*12*). Our expectation was that the
243 observed relationships between rate and time will differ from results where evolutionary change
244 and time are completely random with respect to one another.

245 As expected, when we randomized the numerator relative to the denominator of the
246 empirical rates, the relationship between rate and time was solely driven by the hyperbolic
247 component across all examined datasets (Fig. 2, OR comparison type). However, we also
248 surprisingly observed substantial similarity between the randomized and empirical results. They
249 were so similar, in fact, that when model parameters were optimized on the randomized data and
250 applied to the corresponding empirical data, the resulting coefficient of determination (R^2),
251 which indicates explanatory power, was nearly identical or sometimes greater than that of the
252 actual empirical fit (Fig. 2). For example, in our largest empirical dataset, Uyeda *et al*. (*15*) with
253 5,886 observations, we predicted 86.9% of the variation using an *hmb* model fit to the actual
254 data, but fitting to randomized data explains on average 85.6% of the variation.

255 We hasten to acknowledge that some differences in the R^2 among empirical and
256 randomized fits is expected given the inferred shifts in the contribution of three components
257 through time. For example, while the empirical dataset of Uyeda *et al*. (*15*) is largely consistent
258 with a hyperbolic distribution (Fig. 1), the rates nearer the maximum age are largely driven by
259 the linear component. Nevertheless, as a practical matter, the fact that we can take measurements
260 like amount of change within *Daphnia* over a summer (*22*), the amount of change between late
261 Cretaceous *Tyrannosaurus* and Triassic *Coleophysis* over 98 million years (*23*), and many
262 datasets in between, shuffle these values, and have the model fit nearly as well on the original
263 data—including explaining over 80% of the variation—suggests that the apparent patterns arise
264 from fitting noise. By contrast, the Yule dataset, where there is variance due to the model (a
265 given simulation condition can create trees of very different sizes) but no error in the number of

taxa used, the tree, or the total height. In this case, the empirical pattern was very different from a hyperbola. Even an error creating a bias, such as censoring the data with zero rates without using a correction, creates a pattern very different from the hyperbola as a consequence of noise (Fig. S7).

Discussion

Our study demonstrates that the pattern of increased evolutionary rates towards the present, observed across various axes of biodiversity, such as contemporary extinction rates (*21*) and macroevolutionary rates in morphology and molecular studies (*1, 2, 15*), closely resembles what we would expect from purely random data. Even within large datasets, the observed hyperbolic pattern closely mirrors noise, except for certain periods where a signal-to-noise ratio would suggest constancy in rates. Furthermore, it is worth noting that while our focus was on ecological and evolutionary datasets, similar conflicts between rates measured over different time scales exist in other fields, including the Hubble tension regarding the cosmic expansion rate (*24*). For instance, incorporating Milky Way Cepheids, stars whose brightness fluctuates radially over time, instead of distant stars, increases the expansion estimate (*25*), aligning with our expectations.

Our findings do not imply that assessing patterns of rates over time is impossible. For instance, in dendrochronology, annual tree growth is often estimated using cores with a relatively consistent number of years, rather than increasing the number of years for older samples. A recent study on contemporary bird extinction dynamics (*26*) employed 100-year rolling windows to estimate average extinction rates over time, and contrary to expectation, did not find the highest rate at present. Approaches like these, which maintain consistent time intervals,

289 minimize the likelihood of secular trends in time as well as numerator errors and are likely robust
290 to the statistical artifacts uncovered here.

291 An important implication of our results is that any errors present in measurements of
292 evolutionary change not only greatly affect rate estimation overall, but also any comparative test
293 that examines rate shifts within a single phylogenetic tree. Consider a scenario where we want to
294 compare evolutionary rates between two clades, with one nested within the other and thus
295 differing in their overall crown age. Any inaccuracies in trait measurements or character state
296 assignments within the tree could lead to the erroneous inference that the younger clade has a
297 higher rate, simply due to less overall time in terms of branch lengths to mitigate the effects of
298 these errors (Fig. S10). This discrepancy also creates a concerning disconnect between the
299 expected error rates for specific phylogenetic comparative models, derived from simulations, and
300 their actual behavior in empirical settings when assessing rate differences within and among
301 clades. That is, null models for many comparative tests may fit far worse in an empirical setting
302 than might be indicated by careful simulation study conducted in the absence of measurement
303 error (*27*), providing alternative explanations with artificially inflated support. Even with
304 Bayesian approaches that attempt to integrate over uncertainty in results, data are still typically
305 assumed to be free of errors, and this being incorrect will tend to lead to increased rates,
306 especially where the magnitude of the effect of the error is high relative to the magnitude of the
307 effect of the signal.

308 We anticipate that our findings will encourage further development of approaches that
309 acknowledge measurement error when estimating rates. In models of continuous trait evolution,
310 measurement error can be addressed by directly inputting the standard error of the species' mean
311 for a trait to the model (*28*). However, this practice is not yet standard, but at the very least,
312 measurement error could be incorporated as a parameter of the model. In other comparative

313 models, especially those related to the evolution of discrete characters (e.g., nucleotides or
314 higher-level phenotypic characters exhibiting a finite state space), rates are determined without
315 methods to mitigate potential errors, like inaccuracies in character state assignments. Such errors
316 have a clear impact of inflating rates over shorter time scales (e.g., Fig. 1 and Fig. S9).

317 One unexplored issue is that while our study solely focused on errors in measurements of
318 evolutionary change, errors in the temporal component underlying a rate can also significantly
319 influence rate estimates. In certain scenarios, such as diversification models, we suspect errors in
320 branch lengths might counteract the observed patterns, at least partially. Therefore, given the
321 potential biases in molecular age estimates (*29-31*), it is prudent to prioritize the development of
322 integrative approaches that address errors not only in measurements of evolutionary change, but
323 also in the underlying branch lengths.

324 Finally, our findings demonstrate the importance of using informed nulls as expectations
325 of a process (*32*), especially as errors have a biased effect on rate estimates. Understanding how
326 rates vary both within and among phylogenies will always remain a key part of biology, as they
327 can point to important principles governing biodiversity across disparate timescales. However,
328 statistical artifacts can still impishly affect our results, particularly when we dismiss uncertainties
329 in our measurements as inconsequential.

330

331

332 **Acknowledgments**

333 This work presented here was substantially influenced by James Boyko's advocacy in
334 challenging the artifactual nature of age rate scaling. Conversations with Dan Simberloff, Nina
335 Fefferman, Paul Armsworth, Xingli Giam, Naomi O'Meara, Adam Siepielski, and Andy
336 Alverson were also of substantial help in generating ideas and providing constructive feedback.

This study was funded by grants from National Science Foundation (grant DEB−1916539 awarded to B.C.O. and grant DEB−1916558 awarded to J.M.B.).

Materials and Methods

Overall workflow. Other than the Yule simulations themselves (see below), all analyses were done in R 4.2.3 (*33*) using *targets* (*34*). This creates a reproducible workflow that allows for caching intermediate results. All scripts and substantial outputs are available at https://github.com/bomeara/hyperbolic_rates . Other packages used included *ggplot* (*35*), *ape* (*36*), *dplyr* (*37*), *Rmpfr* (*38*), *tidyr* (*39*), *nloptr* (*40*), *dentist* (*41*), and *scales* (*42*).

Empirical data and Yule simulation. We fit our models to five empirical data sets and rates generated from a simulation of Yule trees. We used the body size dataset compiled by Uyeda *et al.* (*15*, their "Dryad7.csv"), which integrates contemporary field studies, historical field data, and the fossil record for mammals, squamates, and birds. This was filtered for points that were BodySizeCorrelated==1, but no further filtering was done. Due to its historical importance, we also used the morphological rate data set of Gingerich (*1*). We used the mitochondrial substitution rate (measured in sites Myr^{-1}) data of Ho *et al.* (*2*). Finally, we used the rates of contemporary species extinction (E/Msy, or extinctions per million species-years) from Barnosky *et al.* (*21*).

The original study of Henao Diaz *et al.* (*3*) estimated diversification rates (i.e., speciation, extinction, and net diversification rates) from these trees. However, these were not available (Matt Pennell, personal communication). We therefore re-estimated diversification rates by fitting a constant birth-death model to each tree using modified functions extracted from the R package *TreePar* (*43*). Specifically, we re-parameterized the likelihood function to search for the

361 maximum likelihood estimates of turnover, τ, which is speciation + extinction, and extinction
362 fraction, ε, which reflects the ratio of extinction rate to speciation rate. We conditioned the
363 likelihood on survival (i.e., probability that the observed tree could have gone extinct by time, *t*).
364 For each tree, we transformed estimates of τ and ε back to the original speciation and extinction
365 variables.

366 In the case of Gingerich (*1*), Ho *et al.* (*2*), and Barnosky *et al.* (*21*) the original data were
367 not readily available. We therefore used plotdigitizer.com to manually digitize the points from
368 the log-log plots, as they allowed easiest visibility of individual points. For Ho *et al.* (*2*) we
369 recreated Fig. 1*A* from (*7*). For Gingerich (*1*), in cases where there were multiple data points per
370 symbol, and so we recorded one point for each replicate. For example, a symbol of "5" became
371 five points at that location, whereas an "X" means 10 or more. In the latter case, we simply
372 encoded these as ten points.

373 For the Yule simulated rates, we generated a vector of 25,000 ages from a uniform
374 distribution, starting at log(1 Myr) and ending at log(50 Myr). We then transformed the ages
375 back to their original units. This ensured that the sampling was approximately even on a log-log
376 plot. We then simulated a tree at each of these ages in the R package *TreeSim* (*44*) with a known
377 birth rate of $\lambda = 0.10$ births Myr^{-1}. Each simulation began with two lineages and terminated once
378 the tree reached the specified age. Estimates of the birth rate were obtained analytically using our
379 newly derived unbiased estimator, $\hat{\lambda}^*$, for Yule birth rates (see section *Deriving an unbiased*
380 *estimator for the Yule birth rate* section below).

381

382 ***Least-squares model.*** We implemented a set of least-squares models into an R package, *hyperr8*,
383 whose namesake is based on our expectation of hyperbolas and as an homage to Mike
384 Sanderson's *r8s* software (*45*). The basic idea is to minimize the logarithm of the residual sum of

squares between the log predicted and log observed rates coming from a model which, in its most general form, predicts a rate estimate as:

$$\hat{r}(t) = \frac{h}{t} + mt + b.$$

This general model is referred to as the *hmb* model after its free parameters *h*, *m*, and *b* that correspond to the hyperbolic, linear, and constant components. We can fix any of the parameters to 0 to create simpler models nested within this general one. For example, *h*0*b* fixes *m* at 0 and estimates *h* and *b*. The simplest model in the set of possible models restricts *h*=*m*=0 and the rates are assumed constant through time. To avoid some extreme rates driving the fit, we used the log of the empirical rates and the log of the predicted rates when calculating the residual sum of squares. If any empirical data set rates contain rates of zero, as was true in our pure birth simulation for trees that started and ended with two taxa, we used a log1p transform on all rates. This is done automatically in the *hyperr8* software.

For each empirical data set we fit and compare a set of models that range in complexity, as well as assess uncertainty in their inferred model parameters. We accomplished this by taking advantage of the fact that one can convert the residual sum of squares to a log-likelihood. We first calculated the maximum likelihood estimate for the variance of the residuals, $\widehat{\sigma^2}$, by dividing the residual sum of squares by the number of observations. The log-likelihood is then calculated as,

$$log(L) = -.5n * log(\widehat{\sigma^2}),$$

and we maximize this to obtain the MLE of model parameters for a given model. The log-likelihood and the number of free parameters were also used to compute the Akaike Information Criterion (*46*, AIC). This allowed us to use *dentist* (*41*) to compute uncertainty around the parameter estimates, and ΔAIC *Adding uncertainty in our plots* section below).

Our optimization procedure starts with a diverse set of starting points and algorithms from the *nloptr* package (*40*) in R, to fine-tune a given model. Initially, we optimize model parameters using fixed starting points with the NLOPT_LN_SBPLX (*47*) algorithm. Subsequently, a different set of fixed starting points is applied to the same algorithm, followed by random starting points centered on the best values found in the first round of optimization with NLOPT_LN_SBPLX. The optimization process cycles through the NLOPT_LN_BOBYQA (*48*), NLOPT_LN_SBPLX, and NLOPT_LN_NEWUOA_BOUND (*49*) algorithms to mitigate the risk of suboptimal fits. If a better log-likelihood is discovered during the parameter search, those parameters and log-likelihood are used for comparison with other models. This search effort is applied to both the original data and each of the randomized datasets. The resulting fits are provided in Table S1.

Contribution of each parameter to the fitted rate. Using the best-fitting model, we calculated the values of h/t, $m*t$, and b for any given value of t. Since these are usually summed to obtain the fitted rate, $\hat{r}(t)$, it is exceedingly rare that a fitted rate will be estimated as being less than zero. For example, the rate could decrease with time when m is less than one, but a high enough value for b would ensure that $m*t + b$ remains above zero over the range of times examined. Nevertheless, we took the absolute value of each component anyway when calculating its overall contribution to the rate. If $h = 2$, $m = -1$, and $b = 0$, then at time $t = 1$ the overall rate is 2/1 + -1*1 + 0 = 1, but the contribution of the linear component, m, is |-1| / (|2|+|-1|) = $\frac{1}{3}$, while at time t=0.5 the overall rate is 2/0.5 + -1*0.5 + 0 = 3.5 and the contribution m is |-0.5| / (|4| + |-0.5|) = $\frac{1}{9}$.

The rightmost column in Figure 1 of the main text shows the predicted rate as a band, but we also wanted the color of that band to communicate the relative contribution of each

component. This was accomplished by using *geom_ribbon* within *ggplot2* (*35*) with ribbon widths proportional to the contribution of each component. However, this visually makes the ribbons look narrower when the rate has a slope with high magnitude, so we estimated what the slope is locally, which was then used to inflate the ribbon width on steeper parts of the predicted rate line.

Adding uncertainty in our plots. Overall uncertainty in the rate estimates as we all as individual estimates of h, m, and b were inferred using *dentist* (*41*). Functions within *dentist* work by "denting" the likelihood surface, trying sample points around the rim to approximate the confidence interval one would calculate from a chi-square distribution with the same number of degrees of freedom (e.g., a difference in likelihood of 1.92 log likelihood units for a single parameter). In some ways, this is similar to sampling a credibility interval in a Bayesian analysis, but without the possibility of rescuing a ridge in likelihood space through the use of priors. This gives us a set of points inside a region that has "good enough" likelihood. For well-behaved surfaces with a unimodal peak, this will look approximately like an ellipse when any pair of parameters are plotted. To be conservative, we take the minimum and maximum of each parameter inside the confidence region (i.e., the highest and lowest values for h, m, and b) and use these to calculate the range of possible values for the component contributions. For example, we compute the contributions of hyperbolic, linear, and constant components for h_{max}, m_{max}, b_{max}, then do it again for h_{max}, m_{max}, b_{min}, then for h_{max}, m_{min}, b_{min}, and so forth until we have tried all eight combinations at a given time point. The right-hand column of Figure 1 in the main text was created in this way, where the vertical thickness of the bands is due to their range of uncertainty in their proportions. We used *dentist*'s defaults for estimating the likelihood width to use as well as number of steps. For figures on the surfaces, we increased the number of steps to 10,000 and

455 reran the analyses to create denser plots that are easier to see. The results from the dentist for
456 each data set are shown in Figures S1-6.
457
458 ***Randomization procedure.*** As described in the main text, we used a randomization procedure to
459 generate a new empirical null to compare against our empirical fits. This involved first
460 separating each empirical rate into its numerator and denominator components. Some of the
461 datasets included information on the numerator (e.g., difference in log body size for datasets
462 whose rates are measured in darwins) and the denominator (e.g., time separating the pair of
463 values being compared), while others included only rate and time. For the latter, we imputed the
464 numerator by multiplying the rate by time. We then shuffled the numerators at random across a
465 given empirical data set, and then divided them by the unshuffled times to obtain a new set of
466 rates and times. This was repeated 100 times for each dataset.
467 Rates derived solely from phylogenetic trees, such as those in the Henao Diaz *et al.* (*3*)
468 dataset and our pure birth simulations, present a unique challenge for our randomization
469 procedure. Typically, rates are assessed across the entire tree, which reflects the total time span
470 of all branch lengths. However, the time indicated on the *x*-axis typically represents the age of
471 the tree, which, although related (*50*), are far from perfectly correlated. Utilizing the age of the
472 tree to calculate the numerator would significantly underestimate the rate. In such cases, we
473 opted to use the total branch length of the tree to calculate both the numerator and the
474 denominator when recalculating the rate, while still employing the tree's age for the *x*-axis and
475 model fitting. We do note that for the Yule simulation dataset, we did re-run all analyses using
476 total tree height as opposed to clade age and did not find any substantive and qualitative
477 differences from the results presented in the main text (Fig. S7).
478

Comparing empirical and randomized data sets. We commonly use statistical approaches like bootstrapping, jackknifing, or randomizations to estimate parameters from resampled data as a means of testing for robustness. Here, we took an additional step by evaluating the predictive accuracy of these parameters derived from resampled data on the original empirical data. The first step is assessing how well the best model fitted from the empirical data predicts the empirical rates. This is done by calculating the coefficient of determination (R^2) between the empirical rates and predictions from the fitted model (using log rates for the calculation, with the exception of the pure birth simulation where we used log1p given the presence of rates of zero). In Figure 2, this is OO, the comparison of the empirical original (O) data fit to the original (O) parameters. We also compared the fit of the empirical original data to predictions from the randomized data (OR). This allowed us to see how much the random data "null" predicts the original data. It is possible that, even if the original data were pure noise, the fit would still be worse when using a different randomized dataset since it was not tuned to the focal dataset. To check this, we also computed the R^2 between a randomized dataset and the fitted values from a different randomized dataset (the RR column). Our expectation was that if the reshuffled data substantially differed from the empirical data, the fitted values from the reshuffled data would explain little, if any, of the variation in the empirical rate. By contrast, if the empirical pattern is largely noise, as we suspected, there will be little difference between the datasets.

We also computed *p*-values for detecting any significant differences between the original data to original fit (OO) to original data to randomized fit (OR), and the OR fit to the randomized data with different randomized data fit (RR). One issue with this is differences in sample sizes. With OO, for a given dataset and model, there is only one calculation of R^2. For OR, if we have 100 randomizations there are 100 values for R^2 (original with the first randomized dataset, original with the second randomized dataset, and so forth). For RR, since we do not fit the same

503 randomized dataset to itself, there are 99*100 = 9900 comparisons. A comparison of one value
504 versus 100 using something like a *t*-test or Wilcoxon rank-sum test will have far less power than
505 comparing 100 values to 9900 values, making the *p*-values between the OO-OR and OR-RR
506 comparisons incomparable. Also, the low power of a one to 100 comparison means that even if
507 the R^2 for OO is ten times bigger than any value in OR, the *p*-value from a Wilcoxon rank-sum
508 test will not be significant, biasing the result to look as if there are no significant differences
509 between fits from original or random data. Ironically, this lack of power would support our
510 hypothesis that there is no difference between empirical and random data. To avoid this risk, we
511 used percentiles to compute the *p*-value for the OO to OR comparison, with the *p*-value equalling
512 2*max(percentile, 1-percentile), where the percentile is where the OO value is relative to the
513 distribution of OR values. In order to have a comparable value for OR to RR, we took a single R^2
514 value from the set of RR, computed its percentile given the values from OR, converted this to a
515 *p*-value, and then repeated this for 1,000 random draws from the RR distribution. The average of
516 these was taken as the *p*-value for the OR-RR comparison.

517 We also stress the importance of distinguishing biological significance from statistical
518 significance, something long stressed in the literature (*51*). In the context of our work, a *p*-value
519 is the probability of two distributions being exactly equal. For example, in the Uyeda *et al.* (*15*)
520 dataset, it is clear from Fig. 1 in the main text that while the empirical dataset is largely
521 consistent with a hyperbolic distribution, the few points of maximum age are more consistent
522 with a linear distribution. This is reflected in Fig. 2 in the main text, where the fit of the
523 empirical data to predictions from randomized data are slightly worse than fit to empirical data
524 predictions. Due to the large size of the dataset, there is little variance in the former, so the latter
525 is outside the observed distribution and thus is quite significantly different. Biologically, though,
526 an R^2 of 86.9% using the empirical data is not substantially better than an R^2 of 85.6% using

randomized data. It is heartening that on a huge dataset the pattern is not exactly identical to random, but this reflects the rates nearer the maximum age being largely driven by the linear component (see Fig. 1 main text).

As a visual aid, we have also made an animated gif comparing the original and randomized datasets. Given issues with drawing rates of zero for the pure birth simulation, we arbitrarily assigned those rates to be 10% of the minimum value of the nonzero rates for ease in plotting. This is available in our supplemental documents.

Deriving an unbiased estimator for the Yule birth rate. Prior to the development of our least-squares approach described above, we first simulated trees under a pure birth (Yule) process to explore potential artifacts or confounding behaviors that might lead to rates exponentially increasing towards the present when the generating rate is set to be constant through time. We generated a vector comprising 1000 clade ages, uniformly sampled in log space along a line ranging from 1 Myr to 50 Myr. We then simulated a tree at each of these ages with a known birth rate of λ = 0.10 births Myr^{-1}. The simulation began with two lineages and terminated once the tree reached the specified age. Estimates of the birth rate were obtained analytically using the MLE estimator of the birth rate, λ, as derived in (*52*),

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

where denominator represents the sum of the edge lengths, x_i, in a tree. We then plotted all 1000 rates against time and fit a loess curve to assess the trend through time.

One immediate observation was the impact of zero rates – i.e., trees starting at n=2, but which fail to bifurcate after time, t. In an empirical context, such estimates are almost always either discarded (i.e., censored) or not included due to the fact such clades are inherently

uninteresting [evolutionary "minivans" (*53*)]. However, not including these zero rates produces an ascertainment bias where the higher rates nearer the present, which are a result of trees going on a run of speciation events early, are not properly counterbalanced by the higher instances of zero rate trees over shorter time scales. As a result, the mean shifts up where the no change results are more frequent, which we suspect explains the exponentially increasing birth rate estimates towards the present in our Yule simulations (Fig. S8).

Indeed, when zero rate trees are included, the exponential increase disappears entirely, but instead with the overall rate estimates showing a *downward* bias towards the present. In other words, when examining birth rates across time, as we do here, after correcting for ascertainment bias (e.g., only including trees ≥3 taxa) by including the zero rate trees the trend shows an opposite trend of rates slightly *decreasing* towards the present (Fig. S8). A regression line on the log1p transformed data showed a significant positive slope (slope = 0.013), with clade age still accounting for over 6% of the variation. Again, this is due to the higher frequency of trees that failed to speciate over shorter time intervals. When these trees are included, they effectively drive the regression line down.

The remaining downward bias suggests that the Yule birth rate, $\hat{\lambda}$, is a biased estimator of λ, which is not unusual for likelihood-based estimators. Here we quantify the remaining bias by relying heavily on the formulas from Steel and Mooers (*50*). A key insight from (*50*) was that the expected average edge length in a tree grown under a Yule process is actually $\frac{1}{2\lambda}$, as opposed to $\frac{1}{\lambda}$ that might follow from the assumption of exponentially distributed wait times before any given lineage splits on a Yule tree. This is because each speciation event contributes to the average edge length, and each event creates two lineages, leading to the factor of 1/2 in the expression.

For comprehensive mathematical proofs under various conditions (i.e., conditioning) we refer the reader to (*50*).

We use the expected average edge length, $\frac{1}{2\lambda}$, to express the expected value of the MLE estimator of λ as,

$$E(\hat{\lambda}) = E\left[\frac{n-2}{\frac{(2n-2)}{2\lambda}}\right],$$

where the total length is defined as the expected average length $\frac{1}{2\lambda}$ multiplied by the $2n$-2 edges. This expression simplifies to,

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

which we then use to derive the bias of the MLE estimate of $\hat{\lambda}$ by subtracting λ from $E(\hat{\lambda})$:

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

The expected value is indeed less than the true value and therefore provides an underestimate of λ. To correct for this, we can multiply $\hat{\lambda}$ by a "correction" multiplier of $\frac{n}{n-1}$ that converges to 1 as n increases.

Note that when n is small, and zero rates are not included, this correction will amplify higher rates, potentially exacerbating the trend of exponentially increasing rates toward the present. However, when zero rates are included, the rates show a remarkably constant trend through time (Fig. S8). A regression line on the log1p transformed data showed a slope that is essentially zero (slope=0.005), with time explaining less than <1% of the variation in the birth rates (R^2 = 0.004). Thus, throughout the main text, we exclusively focus on estimates of $\hat{\lambda}^*$ through time.

593 ***Yule sensitivity to excluding zero rates.*** As mentioned above, most diversification studies and
594 simulations typically exclude zero rates from their analyses, which can introduce bias. Excluding
595 zero rates fails to properly balance higher rates nearer the present with the higher instances of
596 zero-rate trees over shorter time scales (see Fig. S7). For our least-squares model, removing zero
597 rates from our Yule simulation dataset necessarily shifted the signal towards the hyperbolic
598 component (Fig. S7). We were concerned that if our randomization procedure incorrectly
599 attributed exponentially increasing rates towards the present to errors when there were none, it
600 would also raise questions about the empirical results, which also omitted zero rates. To test this,
601 we reran all analyses on the Yule simulation dataset with the zero rates excluded. As expected,
602 the strongest signal nearest the present was for the hyperbolic component (Fig. S7). In this case,
603 the randomization creates a pattern very different from the hyperbola as a consequence of noise.
604
605 ***Within and among tree rate bias.*** One implication of our study is that errors in measuring
606 evolutionary changes not only greatly affect rate estimation overall, but also any comparative test
607 that examines rate shifts within a single tree. For instance, consider a scenario where we want to
608 compare evolutionary rates between two sister clades that differ in age. Any inaccuracies in
609 branch lengths or character state assignments within the tree could lead to the erroneous
610 inference that the younger clade has a higher rate, simply due to less overall time to mitigate the
611 effects of errors.
612 To demonstrate this, we conducted simulations of character state transitions and
613 estimated transition rates with and without error. We ran 1000 simulations each on two types of
614 64-taxon trees that represented extremes in tree balance: pectinate (i.e., "caterpillar" trees) and
615 perfectly balanced trees, where diversity is evenly distributed between sister clades. The age of
616 each tree was varied by scaling clade age using values from a vector of 1,000 ages drawn from a

uniform distribution, ranging from log(5 Myr) to log(20 Myr). We used a simple transition model that assumed equal transition rates between states 0 and 1, and we chose generating rates of 0.01 transitions Myr^{-1} for pectinate and 0.05 transitions Myr^{-1} for balanced to ensure that a large number of invariant or saturated data sets were not produced across our age range. Finally, we estimated the transition rates in the R package *corHMM* (*54*) on the resulting data sets with and without error, with error being introduced by choosing six taxa at random and changing their true state to the opposite and incorrect one.

The results of these simulations are shown in Figures S9 and S10. First, in the absence of error the estimated transitions rates are, on average, constant across time and are centered on the generating value. However, when we introduce error, the rates are significant and negatively correlated with time, which is consistent with findings presented in the main text. Importantly, if we were to compare the rates between a clade that is say 5-million-years-old that is sister to a clade of say million-years-old, rates be higher in the younger clade simply as a byproduct of errors. Particularly intriguing is the observation that pectinate trees tend to mitigate error impacts more effectively than balanced trees, irrespective of the generating rate (not depicted). Plotting results based on total time (i.e., the sum of all branch lengths) reveals a notable yet intuitive trend – namely, that a young pectinate tree contains significantly more time overall than a balanced tree of the same size. Thus, this underscores the challenges in accurately estimating rates within and among trees that contain errors.

Other sources of bias. Note that the biases arising from error are unlikely to be the only biases playing a role in rate estimation. As we have written elsewhere (*54*), there is a likely substantial effect of human selection biases. Scientists study *Daphnia* over a summer because the rate is high enough that they can expect to find a pattern, ignoring all the other groups with multiple

generations over that period but less evidence for size change. Scientists study dinosaur size differences over 98 million years because their overall similarity is high enough that a comparison seems reasonable; other groups that have changed substantially enough over such a long span of time that they are no longer considered comparable "things" are ignored. Within groups, we focus on the ones where traits are variable: no one studies gain or loss of having skulls in mammals, nor gain or loss of woodiness in oaks, or gain or loss of powered flight in cnidarians, instead focusing on the groups that have elevated enough rates to be practical to study. There are also biases about sizes of groups to study and the way we identify groups, taking the tens of millions of clades on the tree of life and only choosing to assign names to some of them, often seemingly based on key characters arising on the branch leading to a group. We choose to study "X-idae" because that clade received the family name, rather than one node rootward or tipward. The remarkable thing about our study is that while those biases no doubt play a role, the fact that just randomization creates a nearly identical pattern to the actual data suggests that error alone has a dramatic effect all on its own.

Caveats. This study's outcomes represent a conflict of interest for us, as scientists whose careers are based on extensive work developing new ways to estimate rates of evolution in a variety of contexts (*53-57*). Rates are crucial for understanding biology, but their usefulness is diminished, at least in the short term, by inherent time-dependence primarily caused by noise. Although we aimed for objectivity in interpreting our results, we may have inadvertently downplayed the impact of the statistical artifacts we uncovered.

There were also analysis choices that could affect communication of results. For example, we found that the parameter estimates from some of the empirical data sets were slightly outside the distributions of parameter estimates from the randomizations. This could be

spun to demonstrate how significantly different the results are, but in practice the differences have little impact (as shown by the similarity in the R^2 results).

While we did not set a minimum or maximum dataset size, we also did not sample every published comparison of rates versus time. Instead, we limited our focus to the datasets that have been given attention recently and span a variety of rates ranging from molecules to extant species and trees to fossils (*7*). Finally, we developed a robust optimization procedure into the *hyperr8s* package, starting with multiple starting points and using algorithms in *nloptr* and *dentist* to explore parameter space. However, no algorithm guarantees success, and incorrect estimates could result in poorer fits but unlikely to produce the hyperbolic patterns we discovered here.

Finally, one issue with the original datasets is non-independence of the data points. This includes the standard phylogenetic non-independence as a consequence of shared ancestry (*19*), but also stems from the way some of the comparisons were calculated. For example, Uyeda *et al.* (*15*) compiled information from dozens of studies, leading to over 6,031 comparisons, but there are only 1,684 distinct samples used. To ensure comparability with earlier results, we did not correct for either of these issues.

Data availability. All scripts and substantial outputs are available at https://github.com/bomeara/hyperbolic_rates.

References

1. P. D. Gingerich, Rates of evolution: effects of time and temporal scaling. *Science* 222, 159-161 (1983). doi: https://www.science.org/doi/10.1126/science.222.4620.159.
2. S. Y. W. Ho, M. J. Phillips, A. Cooper, A. J. Drummond, Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution* 22, 1561-1568 (2005). doi: https://doi.org/10.1093/molbev/msi145.
3. L. Henao Diaz, L. J. Harmon, M. T. C. Sugawara, M. W. Pennell. 2019. Macroevolutionary diversification rates show time dependency. *Proceedings of the National Academy of Sciences, U.S.A.* 116, 7403-7408 (2019). doi: https://doi.org/10.1073/pnas.1818058116.
4. L. J. Harmon, M. W. Pennell, L. F. Henao Diaz, J. Rolland, B. N. Sipley, J. C. Uyeda, Causes and consequences of apparent time scaling across all estimated evolutionary rates. *Annual Review of Ecology, Evolution, and Systematics* 52, 587-609 (2021). doi: https://doi.org/10.1146/annurev-ecolsys-011921-023644.
5. A. Hallam, Evolutionary size increase and longevity in Jurassic bivalves and ammonites. Nature 258, 493-496 (1975). doi: https://doi.org/10.1038/258493a0.
6. S. J. Gould, Smooth curve of evolutionary rate: a psychological and mathematical artifact. *Science* 226, 994-996 (1983). doi: https://www.science.org/doi/10.1126/science.6505682.
7. J. Rolland, L. F. Henao-Diaz, M. Doebeli, R. Germain, L. J. Harmon, L. L. Knowles, L. H. Low, J. E. Mank, A. Machac, S. P. Otto, M. Pennell, N. Salamin, D. Silvestro, M. Sugawara, J. Uyeda, C. E. Wagner, D. Schluter, Conceptual and empirical bridges between micro- and macroevolution. *Nature Ecology and Evolution* 7, 1181-1193 (2023). doi: https://doi.org/10.1038/s41559-023-02116-7.

8. H.D. Sheets, C.E. Mitchell, Uncorrelated change produces the apparent dependence of evolutionary rate on interval. *Paleobiology* 27, 429-445 (2001). doi: https://doi.org/10.1666/0094-8373(2001)027<0429:UCPTAD>2.0.CO;2
9. G. Hunt, Fitting and comparing models of phyletic evolution: random walks and beyond. *Paleobiology* 32, 578-601 (2006). doi: https://doi.org/10.1666/05070.1
10. G. Hunt, "Evolutionary patterns within fossil lineages: model-based assessment of modes, rates, punctuations and process" in *Evolution to Geobiology: Research Questions Driving Paleontology at the Start of a New Century.* Paleontological Society Papers, P.H. Kelley, R.K. Bambach, Eds. (Paleontological Society, 2008), pp. 117-131. doi: https://doi.org/10.1017/S1089332600001649.
11. K. Pearson, Mathematical contributions to the theory of evolution.—On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the Royal Society, London* 60, 359-367 (1897). doi: https://doi.org/10.1098/rspl.1896.0076.
12. F. Chayes, On ratio correlation in petrography. *Journal of Geology* 57, 239-254 (1949). doi: https://doi.org/10.1086/625606.
13. W. R. Atchley, C. T. Gaskins, D. Anderson, Statistical properties of ratios. I. Empirical results. *Systematic Biology* 25, 137-148 (1976). doi: https://doi.org/10.2307/2412740.
14. S. P. De Lisle, E. I. Svensson, Evolutionary change is remarkably constant across timescales. bioRxiv [Preprint] (2023). doi: https://doi.org/10.1101/2023.12.02.569704.
15. J. C. Uyeda, T. F. Hansen, S. J. Arnold, J. Pienaar, The million-year wait for macroevolution bursts. *Proceedings of the National Academy of Sciences, U.S.A.* 108, 15908-15913 (2011). doi: https://doi.org/10.1073/pnas.1014503108.

16. T. H. Jukes, C. R. Cantor, "Evolution of protein molecules" in *Mammalian Protein Metabolism* (Academic Press, New York), pp. 21-132. doi: http://dx.doi.org/10.1016/B978-1-4832-3211-9.50009-7.

17. M. Pagel, Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society, London* 255, 37-45 (1994). doi: https://doi.org/10.1098/rspb.1994.0006.

18. S. Nee, R. M. May, P. H. Harvey, The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society, London* 344, 305-311 (1994). doi: https://doi.org/10.1098/rstb.1994.0068.

19. J. Felsenstein, Phylogenies and the comparative method. American Naturalist 125, 1-15 (1985). doi: https://www.jstor.org/stable/2461605.

20. J. B. S. Haldane, Suggestions as to the quantitative measurement of rates of evolution. *Evolution* **3**, 51-56 (1949). doi: https://doi.org/10.2307/2405451.

21. A. D. Barnosky, N. Matzke, S. Tomiya, G. O. U. Wogen, B. Swartz, T. B. Quental, C. Marshall, J. L. McGuire, E. L. Lindsey, K. C. Maguire, B. Mersey, E. A. Ferrer, Has the Earth's sixth mass extinction already arrived? *Nature* 471, 51-57 (2011). doi: https://doi.org/10.1038/nature09678.

22. A. J. Tessier, A. Young, M. Leibold, Population dynamics and body-size selection in *Daphnia*. *Limnology and Oceanography* 37, 1-13 (1992). doi: https://doi.org/10.4319/lo.1992.37.1.0001.

23. E. H. Colbert, Evolution of horned dinosaurs. *Evolution* 2, 145-163 (1948). doi: https://doi.org/10.1111/j.1558-5646.1948.tb02737.x.

24. M. Kamionkowski, A. G. Riess, The Hubble tension and early dark energy. *Annual Review of Nuclear and Particle Science* 73, 153-180 (2023). doi: https://doi.org/10.1146/annurev-nucl-111422-024107.

25. W. Cardona, M. Kunz, V. Pettorino, Determining Ho with Bayesian hyper-parameters. *Journal of Cosmology and Astroparticle Physics* 3, 56 (2017). doi: 10.1088/1475-7516/2017/03/056.

26. R. Cooke, F. Says, T. Andermann, T. M. Blackburn, M. J. Steinbauer, A. Antonelli, S. Faurby, Undiscovered bird extinctions obscure the true magnitude of human-driven extinction waves. *Nature Communications* 14, 8116 (2023). doi: https://doi.org/10.1038/s41467-023-43445-2.

27. D. Silvestro, A. Kostikova, G. Litsios, P. B. Pearman, N. Salamin, Measurements errors should always be incorporated in phylogenetic comparative analysis. *Methods in Ecology and Evolution* 6, 340-346 (2015). doi: https://doi.org/10.1111/2041-210X.12337.

28. A. R. Ives, P. E. Midford, T. Garland, Jr, Within-species variation and measurement error in phylogenetic comparative methods. *Systematic Biology* 56, 252-270 (2007). doi: https://doi.org/10.1080/10635150701313830.

29. J. M. Beaulieu, B. C. O'Meara, P. Crane, M. J. Donoghue, Heterogeneous rates of molecular evolution and diversification could explain the Triassic age estimate for angiosperms. *Systematic Biology* 64, 869-878 (2015). doi: https://doi.org/10.1093/sysbio/syv027

30. J. W. Brown, S. A. Smith, The past sure is tense: on interpreting phylogenetic divergence time estimates. *Systematic Biology* 67, 340-353 (2018). doi: https://doi.org/10.1093/sysbio/syx074.

31. G. E. Budd, R. P. Mann, Two notorious nodes: a critical examination of relaxed molecular clock age estimates of the bilaterian animals and placental mammals. *Systematic Biology* In press (2023). doi: https://doi.org/10.1093/sysbio/syad057.

32. D. S. Simberloff, Taxonomic diversity of island biotas. Evolution 24, 23-47 (1970). doi: https://doi.org/10.1111/j.1558-5646.1970.tb01738.x.

30. R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2023). https://www.R-project.org/.

31. W. M. Landau, The *targets* R package: a dynamic Make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. *Journal of Open Source Software* 6, 2959, (2021). doi: https://doi.org/10.21105/joss.02959.

32. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).

33. E. Paradis, K. Schliep, *ape* 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526-528 (2019). doi: https://doi.org/10.1093/bioinformatics/bty633.

34. H. Wickham H, R. François, L. Henry, K. Müller, D. Vaughan, *dplyr*: A Grammar of Data Manipulation. R package version 1.1.4 (2023). doi: https://CRAN.R-project.org/package=dplyr.

35. M. Maechler, *Rmpfr*: R MPFR - Multiple Precision Floating-Point Reliable. R package version 0.9-2 (2023). doi: https://CRAN.R-project.org/package=Rmpfr.

36. H. Wickham, D. Vaughan, M. Girlich, *tidyr*: Tidy Messy Data. R package version 1.3.1, (2024). doi: https://CRAN.R-project.org/package=tidyr.

37. S. G. Johnson, The NLopt nonlinear-optimization package (2023). https://nlopt.readthedocs.io/en/latest

38. J. D. Boyko, B. C. O'Meara, dentist: computing uncertainty by sampling points around maximum likelihood. *bioRxiv* 2023.01.10.523430 [preprint] (2023). doi: https://doi.org/10.1101/2023.01.10.523430.

39. H. Wickham, T. Pedersen, D. Seidel, *scales*: Scale Functions for Visualization. R package version 1.3.0, (2023). doi: https://CRAN.R-project.org/package=scales.

40. T. Stadler, On incomplete sampling under birth-death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology* 261, 58-66 (2009). doi: https://doi.org/10.1016/j.jtbi.2009.07.018.

41. T. Stadler, Simulating trees on a fixed number of extant species. *Systematic Biology* 60, 676-684 (2011). doi: https://doi.org/10.1093/sysbio/syr029.

42. M. J. Sanderson, r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301-302 (2003). doi: https://doi.org/10.1093/bioinformatics/19.2.301

43. K. P. Burnham, D. R. Anderson, *Model Selection and Multimodal Inference*. (Springer Press, 2002).

44. T. Rowan, "Functional stability analysis of numerical algorithms", Thesis, University of Texas, Austin (1990).

45. M. J. D. Powell, The BOBYQA algorithm for bound constrained optimization without derivatives. (Department of Applied Mathematics and Theoretical Physics, Cambridge England, technical report NA2009/06 2009).

46. M. J. D. Powell, The NEWUOA software for unconstrained optimization without derivatives. Proc. 40th Workshop on Large Scale Nonlinear Optimization (Erice, Italy, 2004).

47. M. Steel, A. Mooers, The expected length of pendant and interior edges of a Yule tree. *Applied Mathematics Letters* 23, 1315-1319 (2010). doi: https://doi.org/10.1016/j.aml.2010.06.021.

48. S. Nakagawa, I. C. Cuthill, Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews* 82, 591-605 (2007). doi: https://doi.org/10.1111/j.1469-185X.2007.00027.x.

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

50. J. M. Beaulieu, B. C. O'Meara, Can we build it? Yes we can, but should we use it? Assessing the quality and value of a very large phylogeny of campanulid angiosperms. *American Journal of Botany* 105, 417-432 (2018). doi: https://doi.org/10.1002/ajb2.1020

51. J. M. Beaulieu, B. C. O'Meara, M. J. Donoghue, Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Systematic Biology* 62, 725-737 (2013). doi: https://doi.org/10.1093/sysbio/syt034.

52. B. C. O'Meara, C. Ane, M. J. Sanderson, P. C. Wainwright, Testing for continuous trait evolution using likelihood. *Evolution* 60, 922-933 (2006). doi: https://doi.org/10.1111/j.0014-3820.2006.tb01171.x.

53. J. M. Beaulieu, D.-C. Jhwueng, C. Boettiger, B. C. O'Meara, Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution* 66, 2369-2382 (2012). doi: https://doi.org/10.1111/j.1558-5646.2012.01619.x.

54. J. M. Beaulieu, B. C. O'Meara, Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Systematic Biology* **65**, 583-601 (2016). doi: https://doi.org/10.1093/sysbio/syw022.

Figures

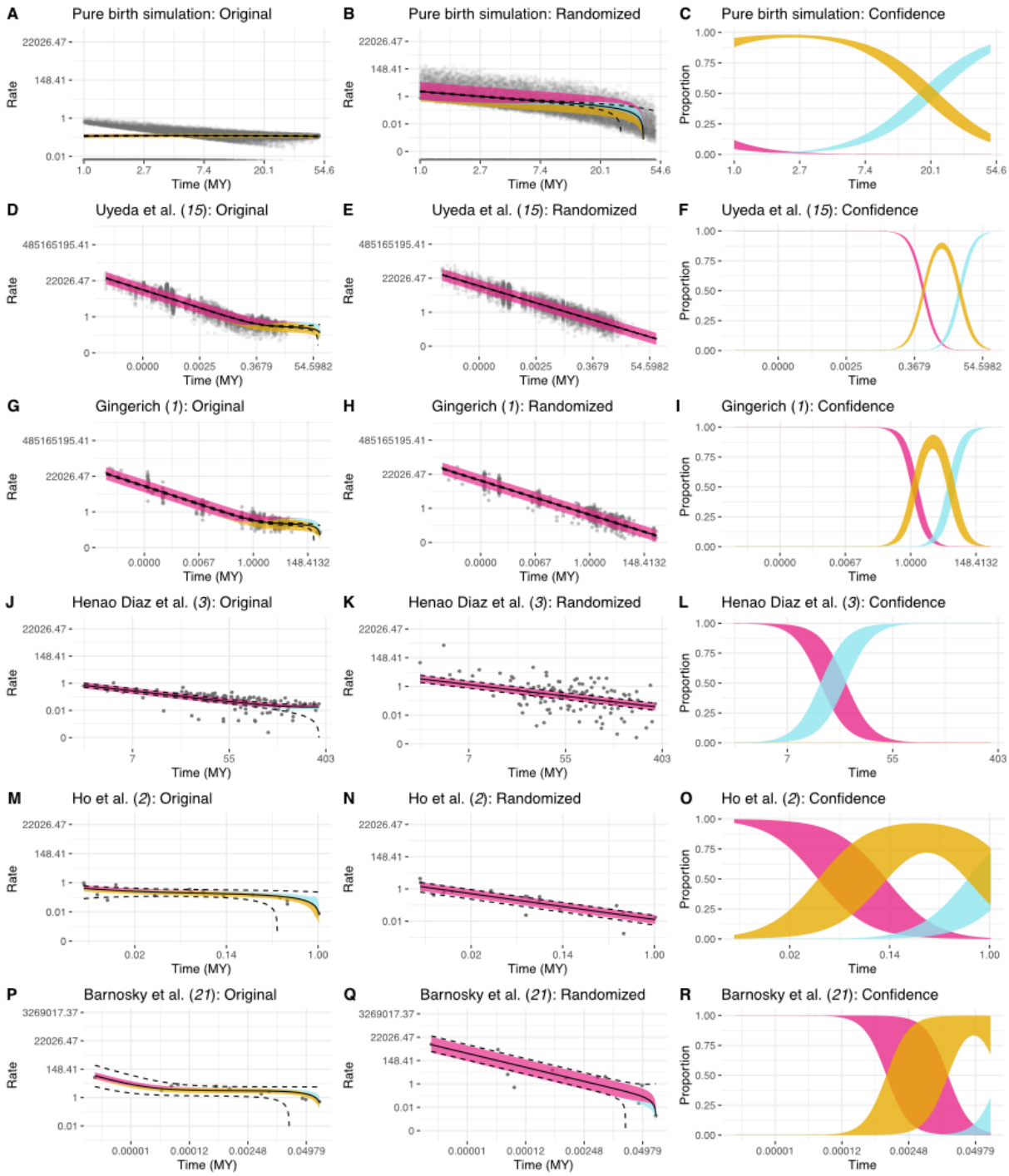
A
Pure birth simulation: Original
B
Pure birth simulation: Randomized
C
Pure birth simulation: Confidence
D
Uyeda et al. (15): Original
E
Uyeda et al. (15): Randomized
F
Uyeda et al. (15): Confidence
G
Gingerich (1): Original
H
Gingerich (1): Randomized
I
Gingerich (1): Confidence
J
Henao Diaz et al. (3): Original
K
Henao Diaz et al. (3): Randomized
L
Henao Diaz et al. (3): Confidence
M
Ho et al. (2): Original
N
Ho et al. (2): Randomized
O
Ho et al. (2): Confidence
P
Barnosky et al. (21): Original
Q
Barnosky et al. (21): Randomized
R
Barnosky et al. (21): Confidence
Rate
Proportion
Time (MY)
Time

Figure 1. Comparisons of log-linear trends in evolutionary rates: simulated data vs. empirical studies. While simulated data exhibit constant birth rates through time, empirical datasets consistently reveal a significant negative correlation between rate and time. The first column is for the original data; the second column is using the randomization procedure. Time is plotted along the horizontal axis, and rate (log-scaled) along the vertical axis. Individual rate estimates are depicted as dots, with a black line representing regression from the best fitting model, and dashed lines indicating the 95% confidence interval around the regression. Additionally, the thick line illustrates the relative impact of the various components from our least-squares model on the overall rate in the best fitting model: dark pink represents the hyperbolic component (which appears linear on this log-log plot), goldenrod denotes a constant rate component, and blue signifies a linear component. Note that for the pure birth simulation there are some rates that are zero, which are placed at negative infinity on a log scale. Finally, the third column illustrates the proportion of the rate from each model component over time, with band thickness indicating uncertainty in that proportion.

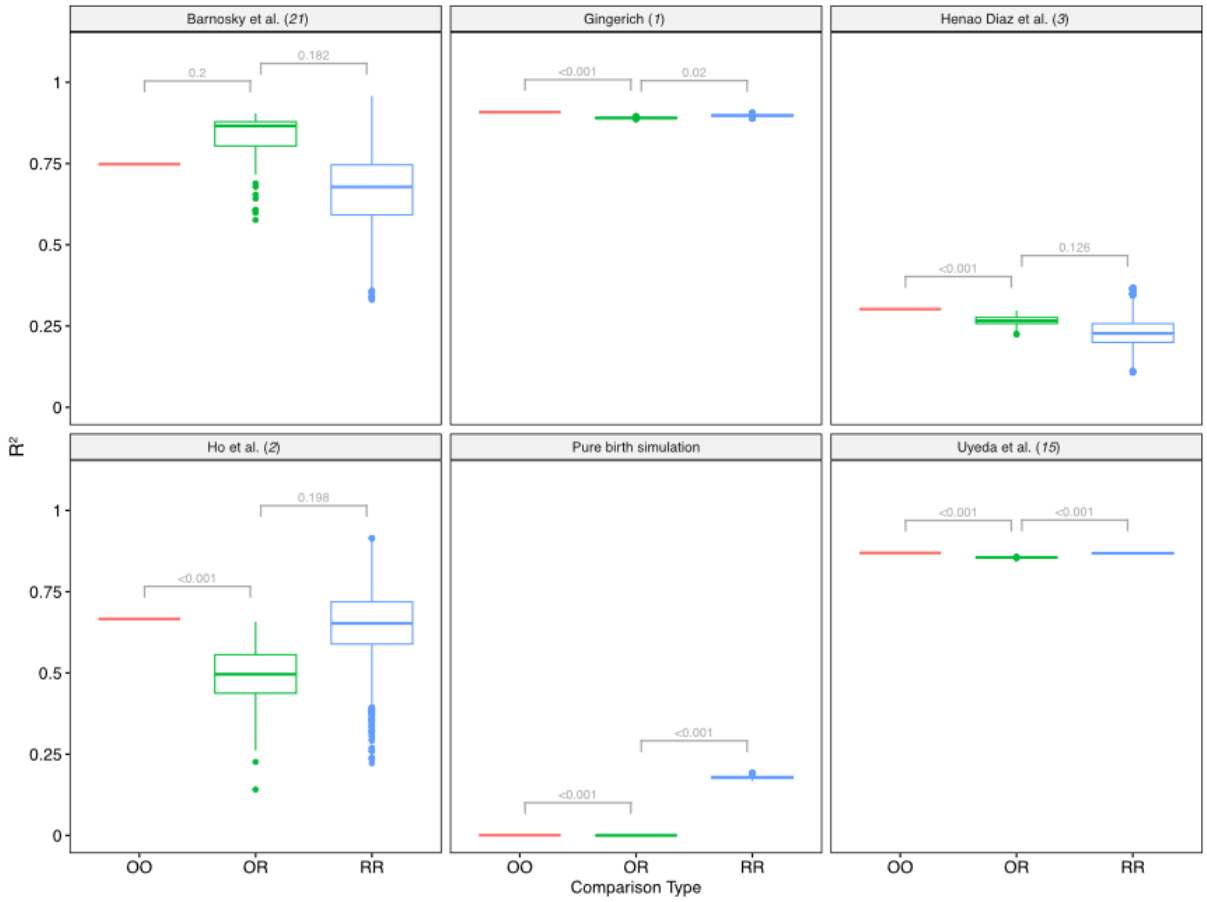

Figure 2. Comparisons of coefficient of determination (R^2) between the observed fit and fits from randomizing the numerator relative to the denominator of the empirical rates. First, obtained the R^2 of the empirical original (O) data fit to the original (O) parameters (comparison type OO). Second, we evaluated the predictive accuracy of the parameters estimated from resampled data applied back to the original empirical data. This involved first separating each empirical rate into its numerator and denominator components, shuffling the numerators at random across a given empirical dataset, and then divided them by the unshuffled times to obtain a new set of rates and times, repeated 100 times. We then compared the fit of the empirical original data to predictions from the randomized data (comparison type OR). This allowed us to see how much the random data predicts the original data. It is possible that, even if the original data were pure noise, the fit would still be worse when using a different randomized dataset since it was not tuned to the focal dataset. To check this, we also computed the R^2 between a randomized dataset and the fitted values from a different randomized dataset (RR comparison type).

Supplementary Figures and Tables

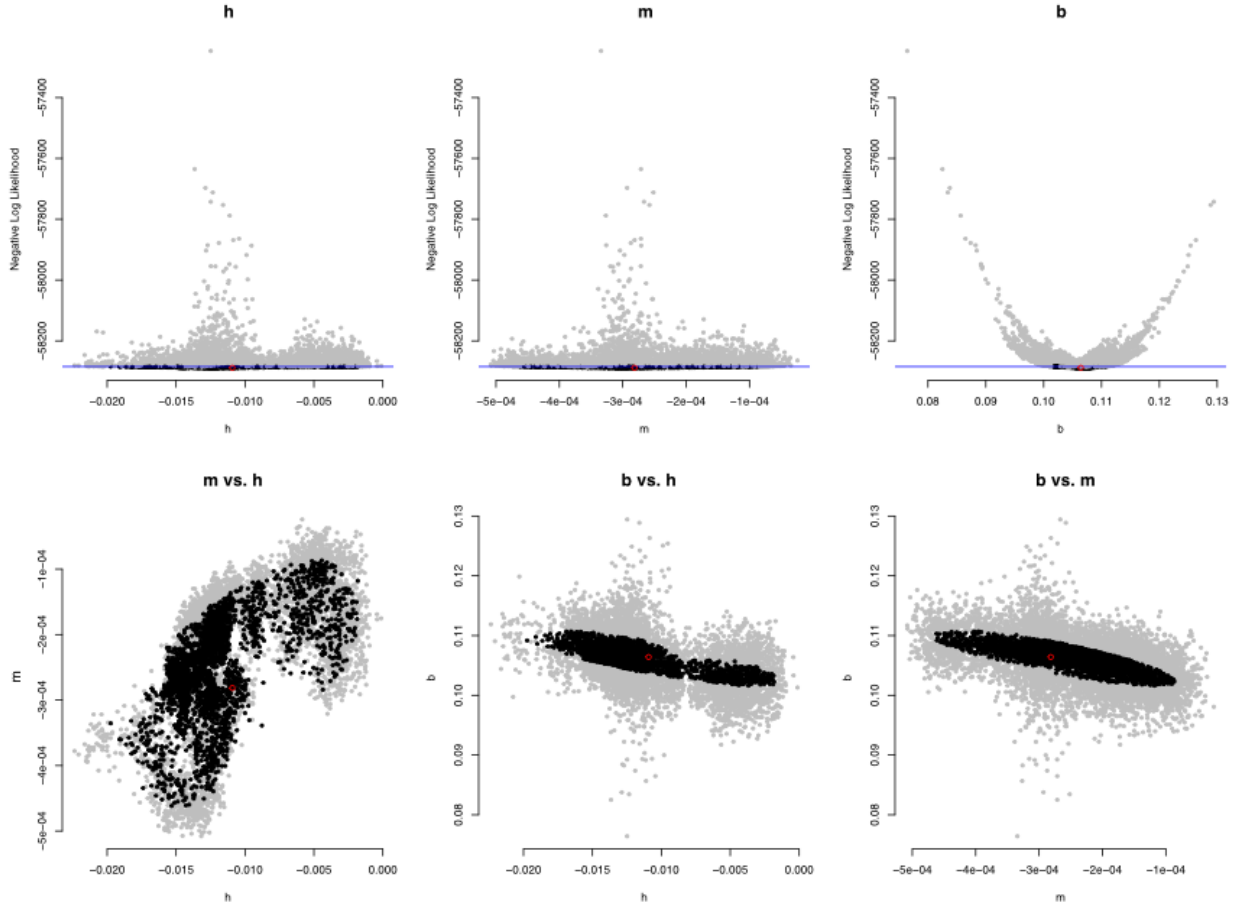

Figure S1: Overall uncertainty in the individual estimates of *h*, *m*, and *b* for the Yule simulation dataset. The top row displays univariate confidence regions for each parameter fitted to the Yule simulation dataset. The best model, denoted as *hmb*, allowed all parameters to vary freely. Sampled parameter values are represented by dots, with gray dots indicating values outside the confidence region and black dots inside. The blue line indicates the boundary between these points. This representation offers a flattened view of the multidimensional analysis; within the horizontal range of black points, some gray points may exist where values for other parameters lead to poor likelihoods. The red circle marks the maximum likelihood estimate. The second row presents bivariate plots, with colors indicating confidence regions as described above. In an effective analysis, the black region in these plots should form an ellipsoid shape, surrounded by gray. To obtain a conservative estimate from the points, such as those in the rightmost column of Fig. 1 in the main text, a rectangular prism was placed around the clusters of black points. The range of values, such as the proportion of hyperbolic weight, was derived from calculations at the vertices of this prism. The overall approach of *dentist* resembles Markov Chain Monte Carlo in Bayesian analysis but does not rely on prior distributions. Instead, it focuses on establishing bounds rather than defining a region that comprises a certain proportion of overall probability.

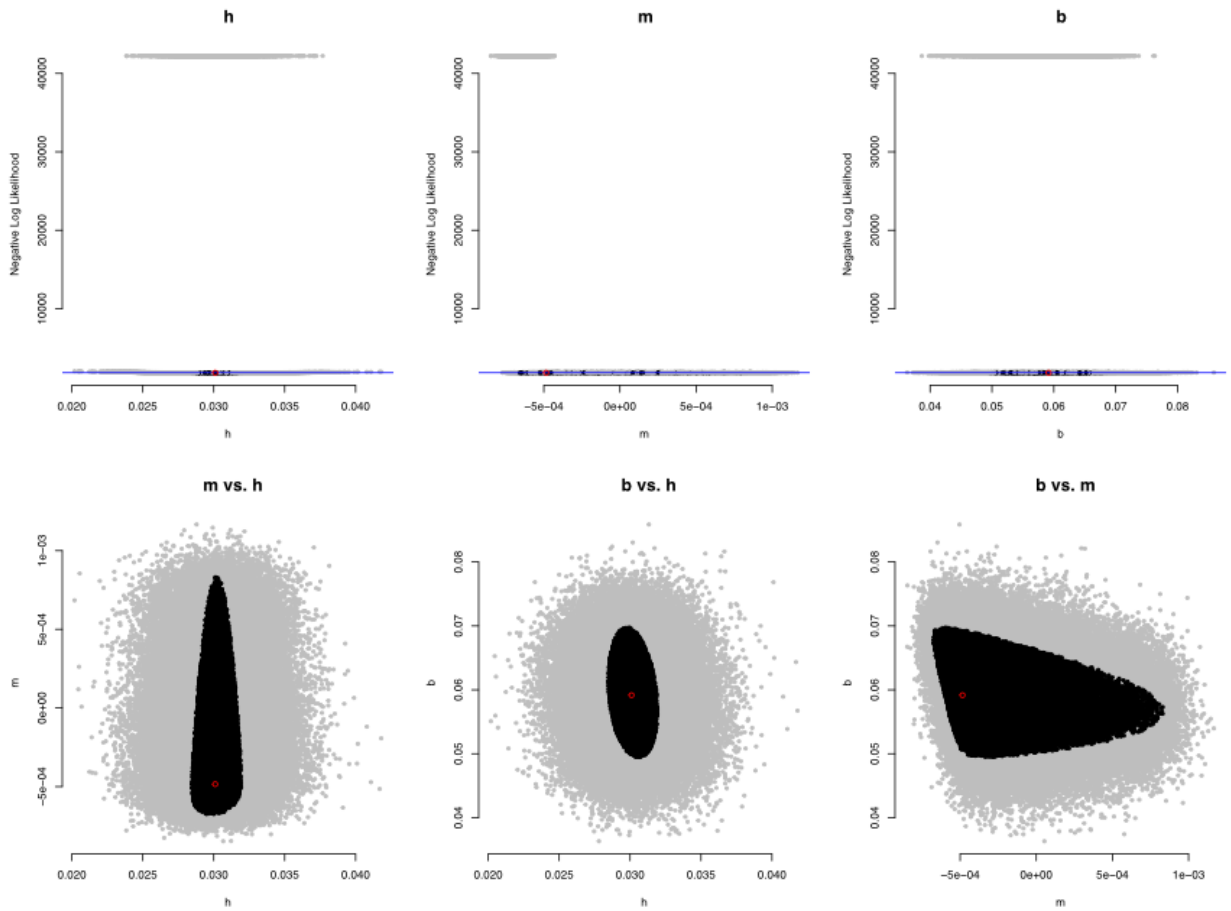

Figure S2. Same *dentist* plot as in Fig. S1 but showing parameter uncertainty for the Uyeda *et al.* (*15*) dataset; the best model was *hmb* so all parameters were free to vary.

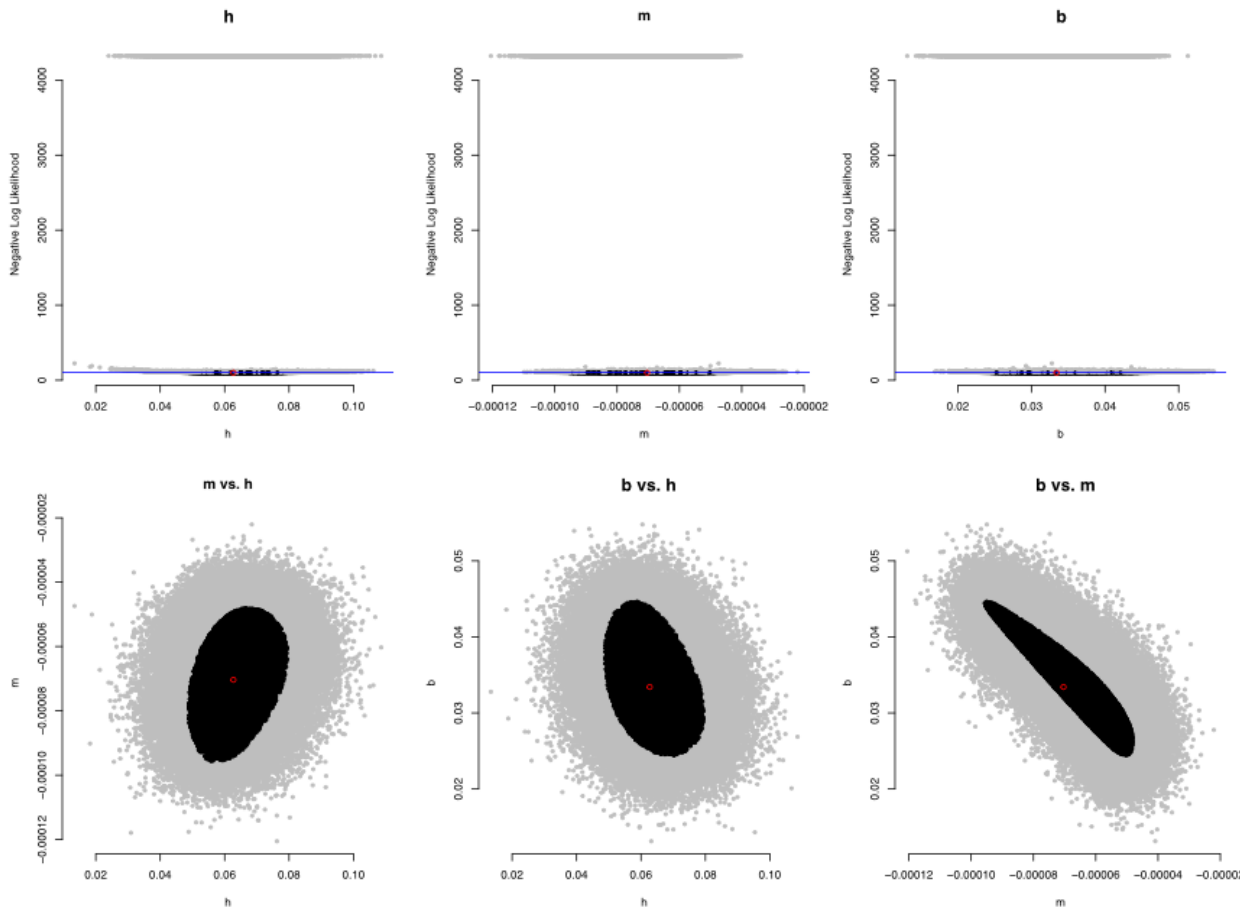

Figure S3. Same *dentist* plot as in Fig. S1 but showing parameter uncertainty for the Gingrich (*1*) dataset; the best model was *hmb* so all parameters were free to vary.

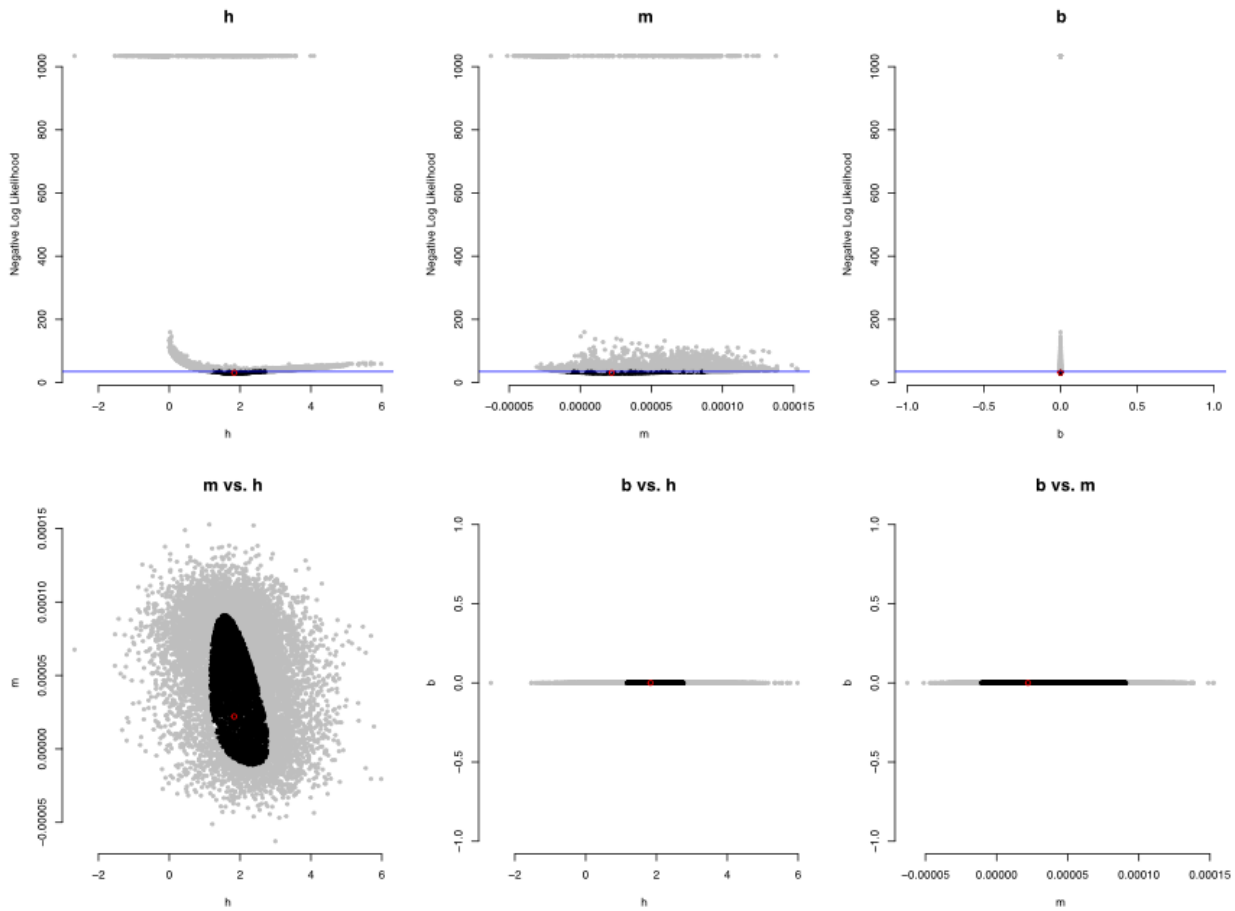

Figure S4. Same *dentist* plot as in Fig. S1 but showing parameter uncertainty for the Henao Diaz *et al.* (*3*) dataset; the best model was *hm*0 so the *b* parameter is fixed at zero.

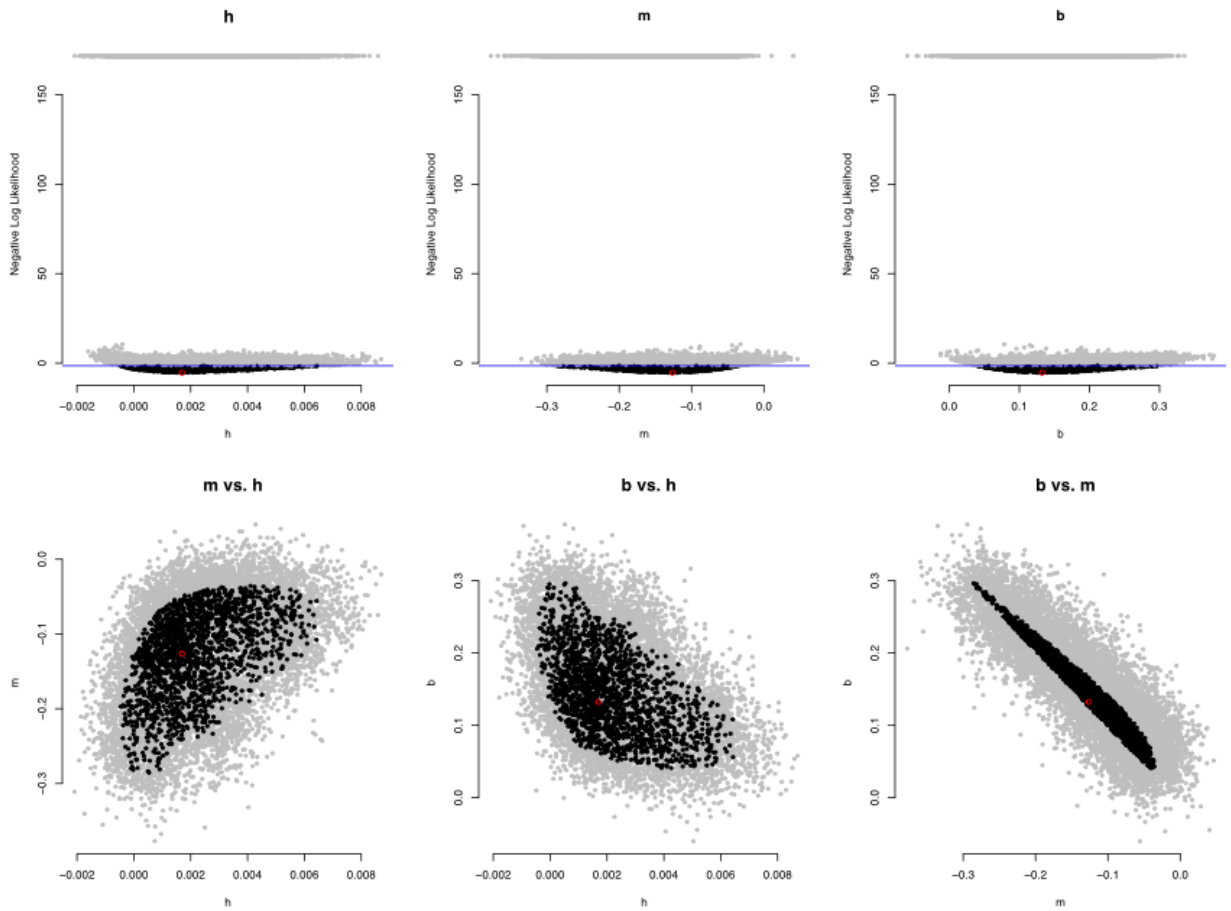

Figure S5. Same *dentist* plot as in Fig. S1 but showing parameter uncertainty for the Ho *et al.* (*2*); the best model was *hmb* so all parameters were free to vary.

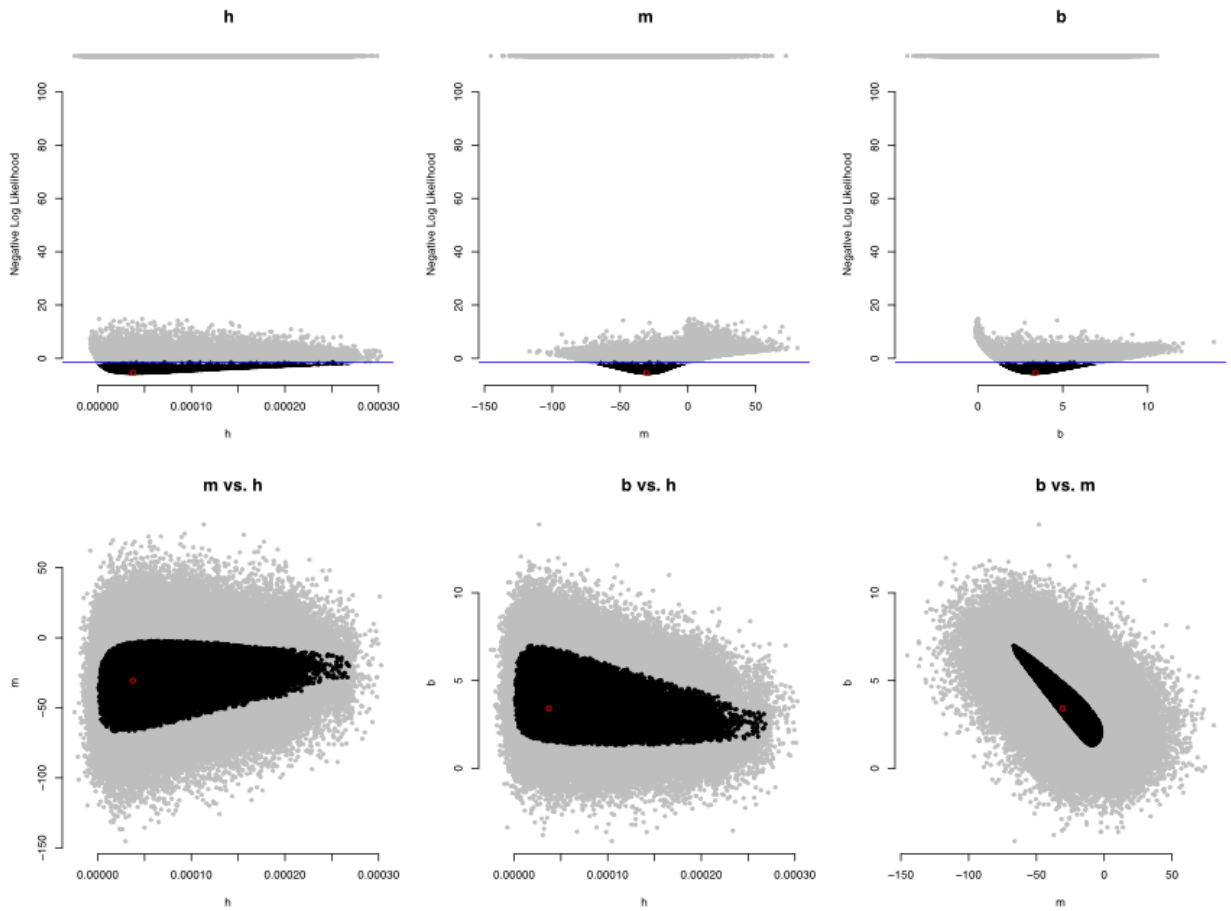

Figure S6. Same *dentist* plot as in Fig. S1 but showing parameter uncertainty for the Barnosky *et al.* (*21*); the best model was *hmb* so all parameters were free to vary.

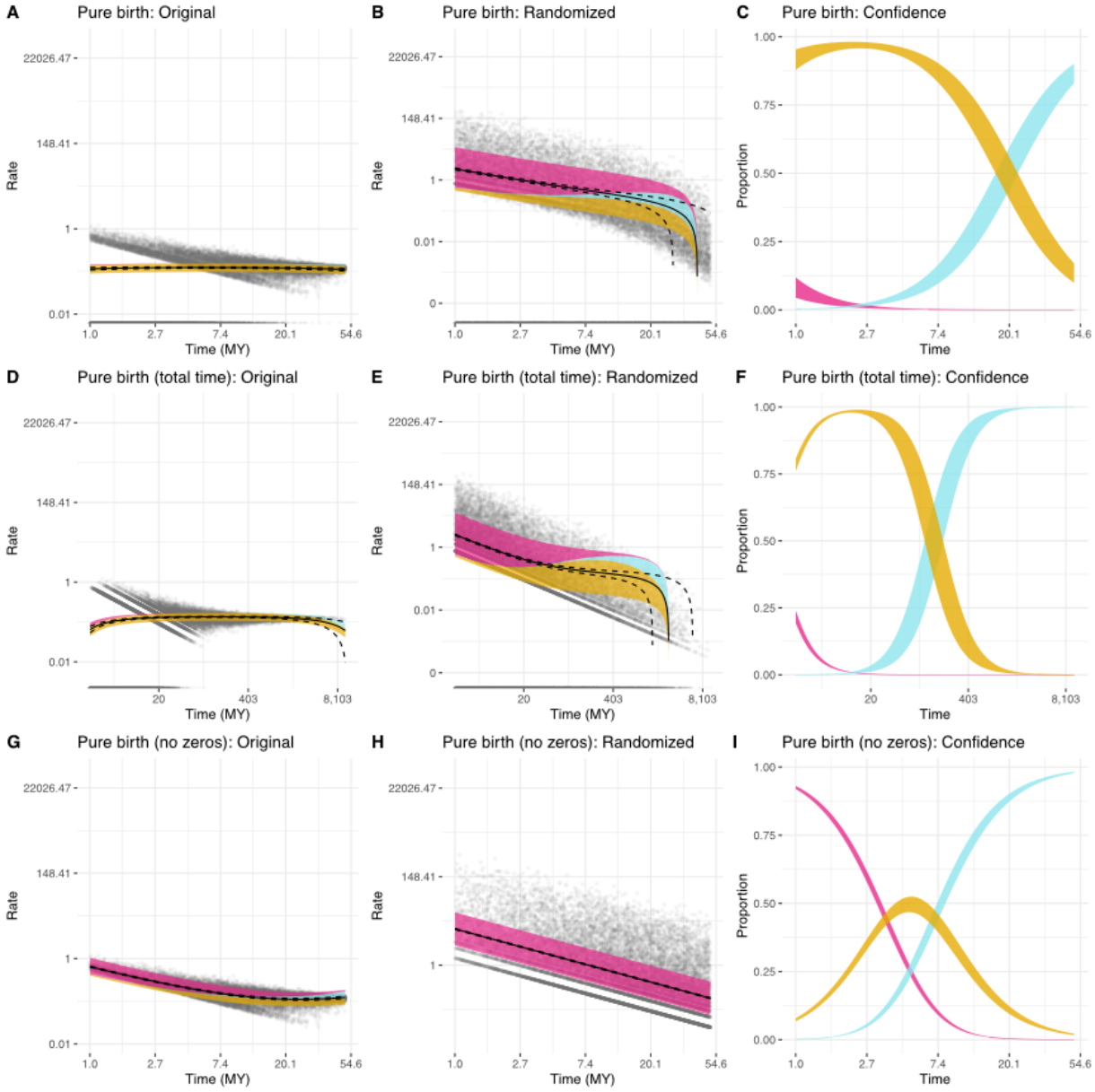
A
Pure birth: Original
B
Pure birth: Randomized
C
Pure birth: Confidence
D
Pure birth (total time): Original
E
Pure birth (total time): Randomized
F
Pure birth (total time): Confidence
G
Pure birth (no zeros): Original
H
Pure birth (no zeros): Randomized
I
Pure birth (no zeros): Confidence
Rate
Proportion
Time (MY)
Time
22026.47
148.41
1
0.01
0
1.00
0.75
0.50
0.25
0.00
1.0
2.7
7.4
20.1
54.6
20
403
8,103

Figure S7. Rates over time from different alterations of the Yule simulation dataset presented in the main text. The first column is for the original data; the second column is using the randomization procedure. The horizontal axis is time; the vertical axis is rate (both on a log scale). Dots show individual rate estimates; the black line shows the regression from the best fitting model and the dashed lines show the 95% confidence interval around that regression. The thick line shows the relative impact of the magnitude of each component on the overall rate in the best fitting model: dark red is from the hyperbolic component (which would be linear on this log-log plot), goldenrod is from a constant rate component, and blue is from a linear component. Note that for the pure birth simulation there are some rates that are zero, which are placed at negative infinity on a log scale; *ggplot2* handles these by plotting them along the *x*-axis. The third column shows the proportion of the rate from each component over time, with the thickness of the bands representing the uncertainty in that proportion.

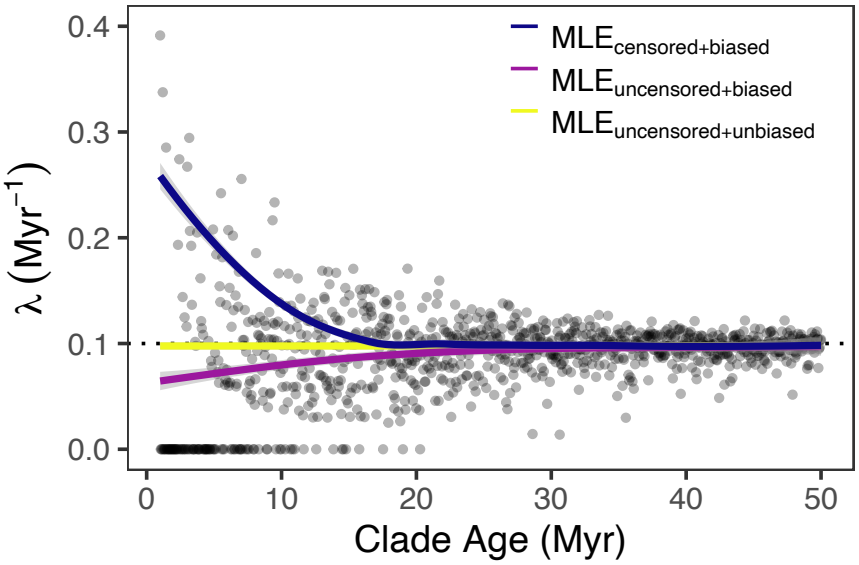

Figure S8. Trends of the Yule birth rate through time under various conditions of the data and the estimator. We simulated 1000 Yule trees that assumed a constant rate of 0.1 birth My^{-1}, (dotted black line) regardless of the time. When zero rate tree are excluded (i.e., censored), and we rely on the MLE of the birth rate, $\hat{\boldsymbol{\lambda}}$, we find that rates exponentially increase towards the present (blue line, MLE$_{censored+biased}$). However, when we include the zero rate trees the line dramatically shifts downwards, with rates nearer the present being lower than in deeper time frames (purple line, MLE$_{uncensored+biased}$). This suggests that the MLE estimator of the birth is biased. When we apply a "correction" to account for the bias, we see that the rates are indeed constant through time (yellow line, MLE$_{uncensored+unbiased}$), which is consistent with the simulation scenario.

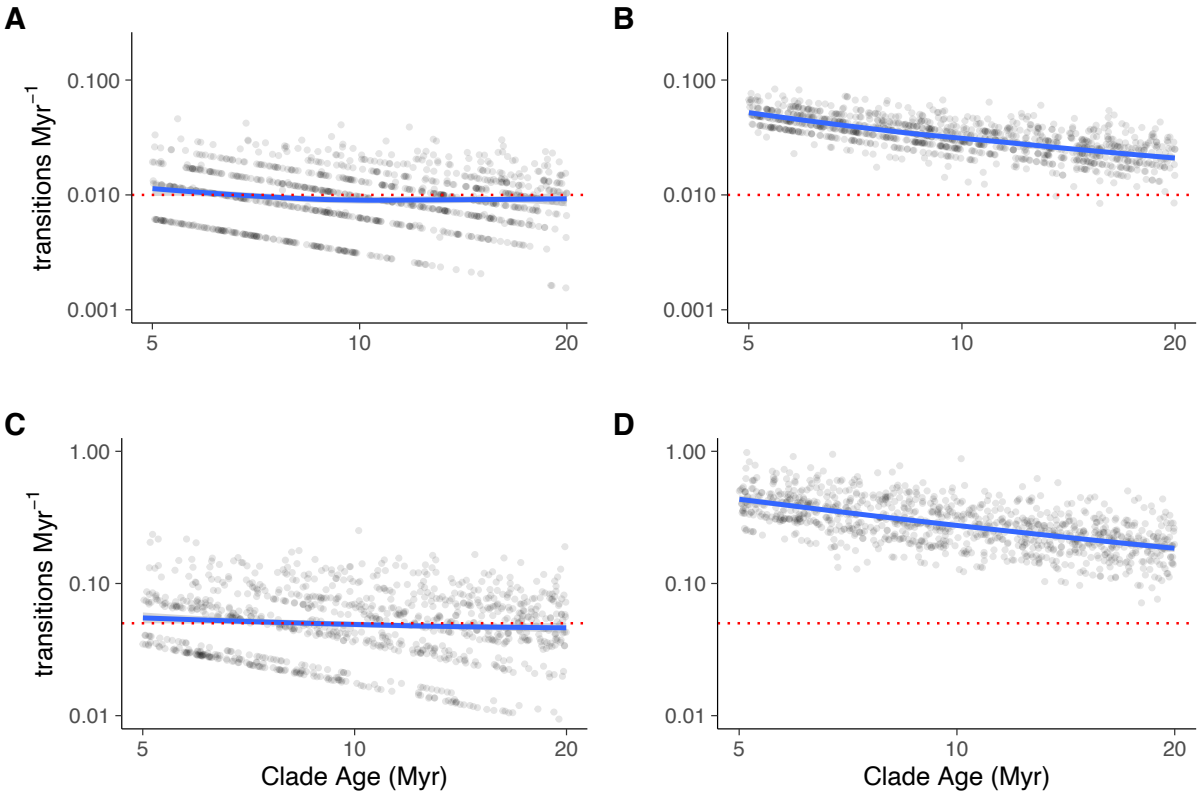

Figure S9. Log-linear plot of estimated transition rates with and without error as a function of clade age obtained from simulated character state transitions. The figure displays the results of state transition rates estimated at various time points, using data simulated on two types of 64-taxon trees representing extremes in tree balance: pectinate (resembling "caterpillar" trees) and perfectly balanced trees, where diversity is evenly distributed between sister clades. The true rates were assumed to be constant over time, indicated by the red dashed line. The first column (a and c) illustrates rates with no errors in the datasets and constant rates over time, depicted by the blue trend line. The second column (b and d) shows rate estimates from the same data but with errors introduced by randomly selecting several taxa and changing their true state to the incorrect one before re-estimating the rate. These errors result in an artificial negative relationship between the transition rate and clade age.

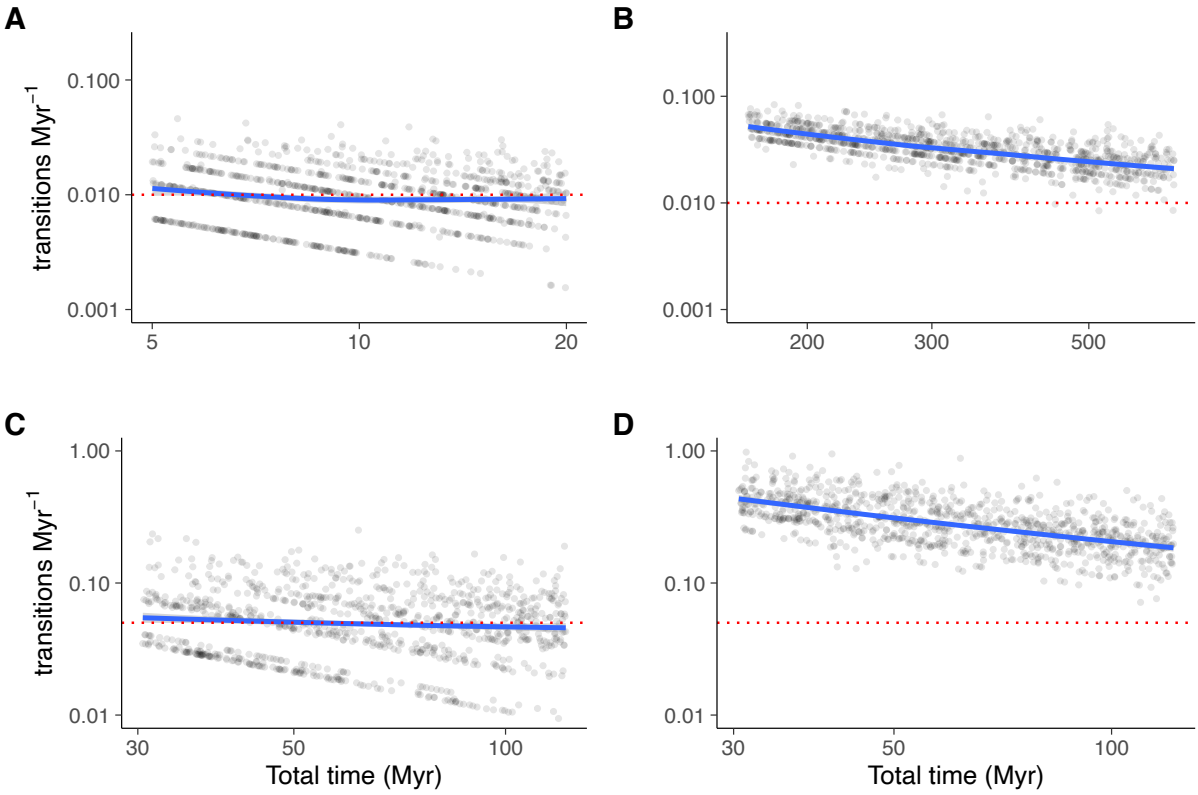

Figure S10. Log-linear plot of estimated transition rates with and without error as a function of sum of all branch lengths obtained from simulated character state transitions. Note that this figure is similar to Fig. S9 but replaces clade age on the vertical axis with the total time represented by a given tree (i.e., the sum of all branch lengths). Note the difference in scale for both axes. Pectinate trees are less affected by errors compared to balanced trees due to the greater overall time represented in the former.

Table S1. Summary of the results obtained from fitting a least-squares model to each empirical dataset and various iterations of our Yule simulated datasets. Rows highlighted in bold indicate the models with the lowest AIC overall from the seven models fitted to each dataset. The values in parentheses beneath each of the model parameters denote the two ends of the 95% confidence interval obtained from our uncertainty analyses.

Dataset	Model	n	# Pars	AIC	ΔAIC	$\hat{h}$	$\hat{m}$	$\hat{b}$	R^2 (OO)
Ho et al. (2005)	h00	17	1	6.29	10.60	0.009 (0.004, 0.021)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.514
Ho et al. (2005)	0m0	17	1	35.47	39.78	0.000 (0.000, 0.000)	2.112 (0.274, 4.077)	0.000 (0.000, 0.000)	0.514
Ho et al. (2005)	hm0	17	2	8.24	12.55	0.008 (0.003-0.022)	0.004 (-0.02, 0.045)	0.000 (0.000, 0.000)	0.485
Ho et al. (2005)	00b	17	1	10.02	14.33	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.136 (0.052, 0.356)	0.000
Ho et al. (2005)	h0b	17	2	4.95	9.26	0.004 (0.000, 0.015)	0.000 (0.000, 0.000)	0.043 (-0.005, 0.197)	0.360
Ho et al. (2005)	0mb	17	2	-2.70	1.61	0.000 (0.000, 0.000)	-0.190 (-0.356, -0.097)	0.199 (0.112, 0.370)	0.585
Ho et al. (2005)	**hmb**	**17**	**3**	**-4.31**	**0.00**	**0.001** **(0.000, 0.002)**	**-0.156** **(-0.282, -0.088)**	**0.163** **(0.117, 0.289)**	**0.666**
Gingerich (1983)	h00	515	1	440.11	234.39	0.154 (0.128, 0.186)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.890
Gingerich (1983)	0m0	515	1	2196.52	1990.79	0.000 (0.000, 0.000)	2.440 (0.885, 6.906)	0.000 (0.000, 0.000)	0.890
Gingerich (1983)	hm0	515	2	402.63	196.91	0.116 (0.089, 0.152)	<0.001 (0.00, 0.001)	0.000 (0.000, 0.000)	0.880
Gingerich (1983)	00b	515	1	1430.65	1224.93	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.614 (0.375, 0.993)	0.000
Gingerich (1983)	h0b	515	2	238.97	33.25	0.067 (0.052, 0.085)	0.000 (0.000, 0.000)	0.027 (0.019, 0.037)	0.902

Gingerich (1983)	0mb	515	2	1417.41	1211.68	0.000 (0.000, 0.000)	-0.001 (-0.002, -0.001)	0.657 (0.409, 1.027)	0.037
Gingerich (1983)	**hmb**	**515**	**3**	**205.72**	**0.00**	**0.063 (0.049, 0.079)**	**0.000 (0.000, 0.000)**	**0.035 (0.024, 0.045)**	**0.908**
Uyeda et al. (2011)	h00	5886	1	4385.94	713.25	0.039 (0.037, 0.041)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.856
Uyeda et al. (2011)	0m0	5886	1	23480.11	19807.42	0.000 (0.000, 0.000)	1539 (1183, 2012)	0.000 (0.000, 0.000)	0.856
Uyeda et al. (2011)	hm0	5886	2	3837.35	164.66	0.032 (0.030, 0.034)	0.039 (0.037, 0.043)	0.000 (0.000, 0.000)	0.865
Uyeda et al. (2011)	00b	5886	1	15608.25	11935.56	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	7.792 (6.728, 8.832)	0.000
Uyeda et al. (2011)	h0b	5886	2	3673.33	0.64	0.031 (0.029, 0.032)	0.000 (0.000, 0.000)	0.058 (0.049, 0.068)	0.869
Uyeda et al. (2011)	0mb	5886	2	15595.30	11922.61	0.000 (0.000, 0.000)	-0.079 (-0.086, -0.043)	7.736 (6.779, 8.707)	0.005
Uyeda et al. (2011)	**hmb**	**5886**	**3**	**3672.69**	**0.00**	**0.030 (0.029, 0.032)**	**-0.001 (-0.001, 0.001)**	**0.059 (0.051, 0.069)**	**0.869**
Henao Diaz et al. (2019)	h00	113	1	66.61	0.59	2.075 (1.455, 2.961)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.290
Henao Diaz et al. (2019)	0m0	113	1	184.28	118.26	0.000 (0.000, 0.000)	0.001 (0.000, 0.001)	0.000 (0.000, 0.000)	0.290
Henao Diaz et al. (2019)	**hm0**	**113**	**2**	**66.01**	**0.00**	**1.832 (1.312, 2.747)**	**<0.001 (0.000, 0.000)**	**0.000 (0.000, 0.000)**	**0.302**
Henao Diaz et al. (2019)	00b	113	1	104.58	38.57	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.036 (0.024, 0.048)	0.000
Henao Diaz et al. (2019)	h0b	113	2	67.00	0.99	1.719 (1.175, 2.901)	0.000 (0.000, 0.000)	0.004 (-0.004, 0.005)	0.296
Henao Diaz et al. (2019)	0mb	113	2	95.12	29.11	0.000 (0.000, 0.000)	<0.001 (0.000, 0.000)	0.048 (0.031, 0.071)	0.107

Henao Diaz et al. (2019)	hmb	113	3	67.57	1.56	1.914 (1.102, 2.922)	<0.001 (0.000, 0.000)	-0.005 (-0.008, 0.010)	0.305
Barnosky et al. (2011)	h00	11	1	21.14	25.88	0.003 (0.000, 0.006)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.869
Barnosky et al. (2011)	0m0	11	1	35.76	40.50	0.000 (0.000, 0.000)	3369 (33.95, 3669)	0.000 (0.000, 0.000)	0.869
Barnosky et al. (2011)	hm0	11	2	15.58	20.32	0.001 (0.000, 0.007)	28.43 (2.892, 522.5)	0.000 (0.000, 0.000)	0.393
Barnosky et al. (2011)	00b	11	1	6.36	11.10	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	3.316 (0.963, 5.939)	0.000
Barnosky et al. (2011)	h0b	11	2	1.62	6.36	<0.001 (0.000, 0.007)	0.000 (0.000, 0.000)	2.041 (0.581, 5.823)	0.476
Barnosky et al. (2011)	0mb	11	2	1.82	6.56	0.000 (0.000, 0.000)	-43.19 (-114.4, 9.925)	4.626 (2.078, 11.62)	0.459
Barnosky et al. (2011)	**hmb**	**11**	**3**	**-4.74**	**0.00**	**<0.001 (0.000, 0.003)**	**-30.28 (-64.21, -2.043)**	**3.394 (1.329, 6.643)**	**0.748**
Pure birth simulation	h00	25000	1	-106768.27	9802.38	0.205 (0.198, 0.211)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	**<0.001**
Pure birth simulation	0m0	25000	1	-106544.04	10026.61	0.000 (0.000, 0.000)	0.004 (0.004, 0.004)	0.000 (0.000, 0.000)	0.000
Pure birth simulation	hm0	25000	2	-113065.13	3505.52	0.173 (0.168, 0.179)	0.003 (0.003, 0.004)	0.000 (0.000, 0.000)	0.001
Pure birth simulation	00b	25000	1	-116558.35	12.31	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.100 (0.098, 0.102)	0.000
Pure birth simulation	h0b	25000	2	-116556.73	13.92	-0.000 (-0.002, -0.001)	0.000 (0.000, 0.000)	0.100 (0.099, 0.102)	**<0.001**
Pure birth simulation	0mb	25000	2	-116563.08	7.57	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.102 (0.099, 0.104)	0.000
Pure birth simulation	**hmb**	**25000**	**3**	**-116570.65**	**0.00**	**-0.011 (-0.013, -0.005)**	**0.000 (0.000, 0.000)**	**0.106 (0.103, 0.109)**	**0.001**

Pure birth simulation (total time)	h00	25000	1	-103689.73	13412.62	0.335 (0.320, 0.350)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.020
Pure birth simulation (total time)	0m0	25000	1	-101203.47	15898.89	0.000 (0.000, 0.000)	<0.001 (0.000, 0.000)	0.000 (0.000, 0.000)	**<0.001**
Pure birth simulation (total time)	hm0	25000	2	-105418.33	11684.03	0.332 (0.318, 0.346)	<0.001 (0.000, 0.000)	0.000 (0.000, 0.000)	0.016
Pure birth simulation (total time)	00b	25000	1	-116558.35	544.01	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.100 (0.098, 0.102)	0.000
Pure birth simulation (total time)	h0b	25000	2	-117074.25	28.10	-0.119 (-0.133, -0.107)	0.000 (0.000, 0.000)	0.113 (0.111, 0.116)	0.021
Pure birth simulation (total time)	0mb	25000	2	-116557.50	544.85	0.000 (0.000, 0.000)	<0.001 (0.000, 0.000)	0.100 (0.098, 0.102)	**<0.001**
Pure birth simulation (total time)	**hmb**	**25000**	**3**	**-117102.35**	**0.00**	**-0.128 (-0.143, -0.113)**	**<0.001 (0.000, 0.000)**	**0.116 (0.112, 0.118)**	**0.022**
Pure birth simulation (no zeros)	h00	17108	1	-13484.72	27137.75	1.457 (1.436, 1.478)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.574
Pure birth simulation (no zeros)	0m0	17108	1	12626.46	53248.93	0.000 (0.000, 0.000)	0.012 (0.011, 0.012)	0.000 (0.000, 0.000)	0.574
Pure birth simulation (no zeros)	hm0	17108	2	-38555.09	2067.37	0.784 (0.775, 0.792)	0.002 (0.002, 0.002)	0.000 (0.000, 0.000)	0.670
Pure birth simulation (no zeros)	00b	17108	1	-20597.06	20025.40	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.130 (0.128, 0.131)	0.000
Pure birth simulation (no zeros)	h0b	17108	2	-39580.23	1042.23	0.473 (0.465, 0.482)	0.000 (0.000, 0.000)	0.074 (0.073, 0.075)	0.671
Pure birth simulation (no zeros)	0mb	17108	2	-25320.95	15301.51	0.000 (0.000, 0.000)	-0.002 (-0.002, -0.002)	0.169 (0.166, 0.170)	0.245
Pure birth simulation (no zeros)	**hmb**	**17108**	**3**	**-40622.46**	**0.00**	**0.577 (0.563, 0.590)**	**0.001 (0.001, 0.001)**	**0.046 (0.043, 0.048)**	**0.690**

Movie S1. An animated gif comparing the original empirical data and five different randomizations of the numerator with respect to the denominator.