Full 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.

Abstract
Across a variety of biological datasets, from genomes to conservation to the fossil record, evolutionary rates appear to increase toward the present or over short time scales. This has long been seen as an indication of processes operating differently at different time scales, even potentially as an indicator of a need for new theory connecting macroevolution and microevolution. Here we introduce a set of 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. We show that plotting a noisy numerator divided by time versus time leads to the observed hyperbolic pattern; in fact, randomizing the amount of change over time generates patterns functionally identical to observed patterns. Ignoring errors can not only obscure true patterns but create novel patterns that have long misled scientists.

Introduction

Biology is characterized by diversity: how a modern moss survives and evolves is very different from how Cambrian trilobites did the same. Biologists thus 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 arguably point to the need for new conceptual bridges connecting disparate timescales of evolutionary change [7,8], despite past work showing potential artifactual causes [9]. It also challenges the long-held view that the processes playing out in the past generally behave similarly in the present—that is, life seems to evolve "faster" now. Understanding 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 [10,11]) 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 directly into consideration.

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 [12]. 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 [12,13].

Similarly, when plotting an evolutionary rate against its corresponding denominator, time, the representation becomes a plot of time against its reciprocal (i.e., k/time vs 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 between two time points, is unlikely to remain constant across all timescales, as seen empirically [15]. Our hypothesis is that the observed hyperbolic pattern comes about due to this reciprocal relationship, especially given uncertainty in rates. We show mathematically why this might be the case, develop an approach for estimating the magnitude of the effect, and do randomizations and simulations to suggest that this is driving nearly all the observed empirical patterns.

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 varied. For instance, one model 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*]: with the darwin, the x_i are the traits measured in log space.

Various phenomenological patterns can describe how rates change through time. The simplest is that $\hat{r}(t)$ is a constant rate of change, maintaining a constant value regardless of the measurement interval. For example, under a molecular clock, a mutation rate resulting from DNA copying errors is expected to be constant across clades of different ages. In other words, even though a pair of taxa sharing a common ancestor 5 million years ago are expected to have far fewer mutations than a pair sharing an ancestor 50 million years ago, 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 – a foundational premise in much of evolutionary biology. On the other hand, if mutation rates were increasing towards the present (perhaps due to a decrease in the protective ozone layer, or loss of function of mutation repair enzymes), then we would expect the rates to increase near the present.

Measurements of x_i often reflect the mean for a species or even a measurement of a single individual and, therefore, represent an estimate of the true value. If each measurement estimate, x_i, carries some level of noise attributed to factors like finite sample size (how representative is this particular flower of the species as a whole) or measurement error (how long is this rather stretchy squid), then $\widehat{x_i} = x_i + \varepsilon_i$, with ε_i denoting an error component that can lead to overestimation or underestimation of the measurement. Thus, we can express the rate estimate as including error:

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

We can reorder the terms to get:

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

The first term on the right hand-side, reflects the underlying model. For example, with a constant nonzero rate the first term should be constant: larger trait differences (numerator) are balanced out by larger time intervals (denominator). However, the effect of the second fraction, the error, is quite

different. Since these errors are not inherently time-dependent (e.g., sequencing errors in a DNA analysis or measurement of body lengths in dinosaurs), there are no *a priori* reasons for the magnitude of the error in a 5-million-year-old clade would necessarily be any greater or lesser than the magnitude in a 50-million-year-old clade – the numerator will come from a consistent, time independent, distribution. But this is divided by different amounts of time in the denominator. Thus, a hyperbola will invariably result from the second term when the ratio of these differences in error are scaled by time, and then plotted against time (even if the time estimate itself also has error, as it inevitably does). Thus, the overall pattern of empirical rates versus time is whatever the true pattern is plus a hyperbola coming from measurement error. At short times, under most models we expect the true difference in trait values to be small, while the uncertainty in measurements may remain high, leading to the hyperbola term dominating.

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 $x(t)t^{-1}$]. Essentially, the $\frac{h}{t}$ represents the $\left(\frac{\varepsilon_2-\varepsilon_1}{t_2-t_1}\right)$ term in the equation above, while the $mt + b$ terms represent the $\left(\frac{x_2-x_1}{t_2-t_1}\right)$ term, if we are willing to assume a simple model where the underlying rate varies linearly with time (including the possibility of it being constant). 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.

Results

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

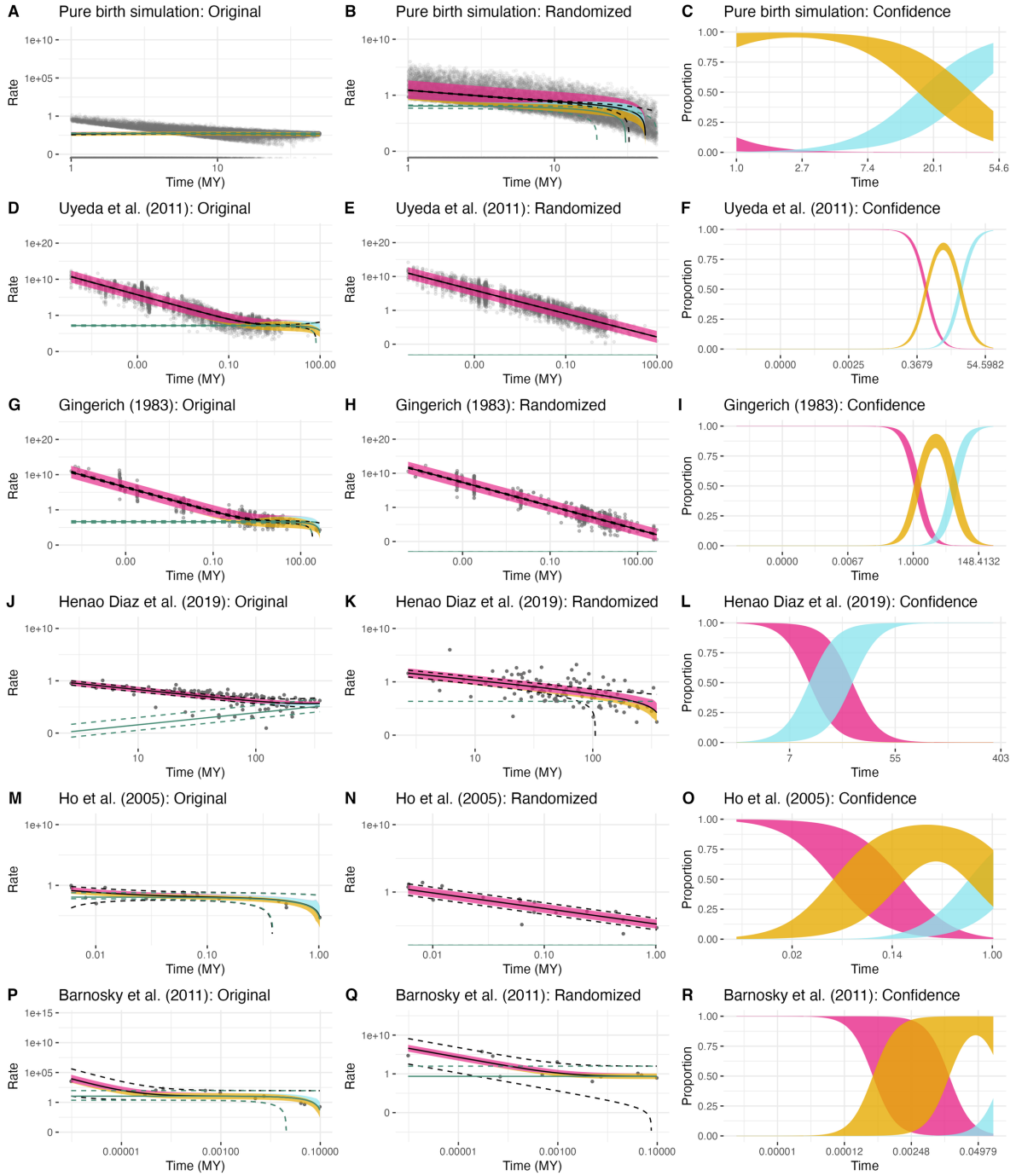

Figure 1 The first column of figures is for the original data; the second column is using the randomization procedure. Time is plotted along the horizontal axis, and rate along the vertical axis, both on log scales. The first row of figures comes from a pure birth simulation, while the other rows come from empirical datasets. Points represent observed rate and time pairs. For the pure birth simulation there are some rates that are zero, which are correctly placed at negative infinity on a log scale but appear at the bottom of the plot here. The black solid and dashed lines represent the prediction (and 95% confidence interval) from the best-fitting hmb model. Note the close match between the predictions from the hmb model and the empirical data, even where the point distribution flattens out in D and G. The aquamarine lines show the prediction from the best hmb model of what the underlying rates would be in the absence of measurement error. The thick line follows the black prediction line; colors indicate 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 component that changes linearly with time. Finally, the third column illustrates the proportion of the rate from each model component over time, with band thickness indicating uncertainty in that proportion.

(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 ($\hat{h}$ = 1.31-2.75 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 ($\hat{m} \neq 0$). This is similar to results from De Lisle and Svennson [14], who also found that a substantial proportion of the empirical pattern could be explained by plotting a ratio versus its denominator, without specifying a particular cause. While it can be hard to estimate raw uncertainty from the empirical datasets (the data are often the observed difference in measurements without uncertainty), in two datasets there are actually multiple observations of rate for a single time point, Gingerich [1] and Uyeda *et al.* [15]. Using just the latter (for the former, we had to estimate the point positions, creating additional uncertainty), there were 653 time intervals where there were two or more observations of rates. The median coefficient of variation across these intervals was 0.59, meaning the standard error was 59% the value of the mean; in 12% of the time periods, the standard error was greater than the mean.

For the fitted rate values within each dataset, we decomposed the rate to reflect the proportion of the rate attributable 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. Decomposing the fitted rates in this way revealed that a significant portion of the relationship between rate and time in the empirical datasets is caused by the hyperbolic component (Fig. 1). This occurred despite model fits indicating that there was signal for both constant and linear components. Any influence of other components only came with long time intervals: exactly what would be expected if there is largely time-independent noise in measurements that only becomes outweighed with signal when the true difference between the measurements is high.

One question is how this problem can be addressed. We performed simulation approaches to assess how well our approach could work at estimating the true underlying

rates when the presumed error was removed. Our approach outperformed DeLisle and Svennson's [14] approach of comparing linear regression slopes to -1, as it more accurately identified trend directions. However, our method sometimes incorrectly detected trends when the generating rate was perfectly constant, especially under high noise conditions. Given the apparent large uncertainty in empirical datasets, even in cases where, say, the trend parameter was nonzero [for example, the value of -0.00007 (CI -0.00009: -0.00005) for *m* for Gingerich [1]], we would avoid drawing strong biological conclusions from this. We are somewhat more confident in estimates for the *y* intercept of the rate. For example, having a rate of 0.059 (CI 0.051-0.069) for Uyeda et al. [15] is reasonable, especially as we expect there to be a positive rate of overall evolution.

Deriving an appropriate null expectation. A potential explanation for the interest in the severe empirical decline in rates over time is that it conflicts with the *de facto* null scientists often envision, which is a constant rate across time. One way to construct a null is to reshuffle the data and plot rates randomly with respect to times and observe whether they create a horizontal line. However, the actual data are the differences in trait measurements and the differences in time. To generate an appropriate 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 [12], but also has the advantage of keeping the same distributions of times and amount of change as the original data, which can matter for some of the biases here [13]. Our expectation was that the observed relationships between rate and time will differ from results where evolutionary change and time are completely random with respect to one another.

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

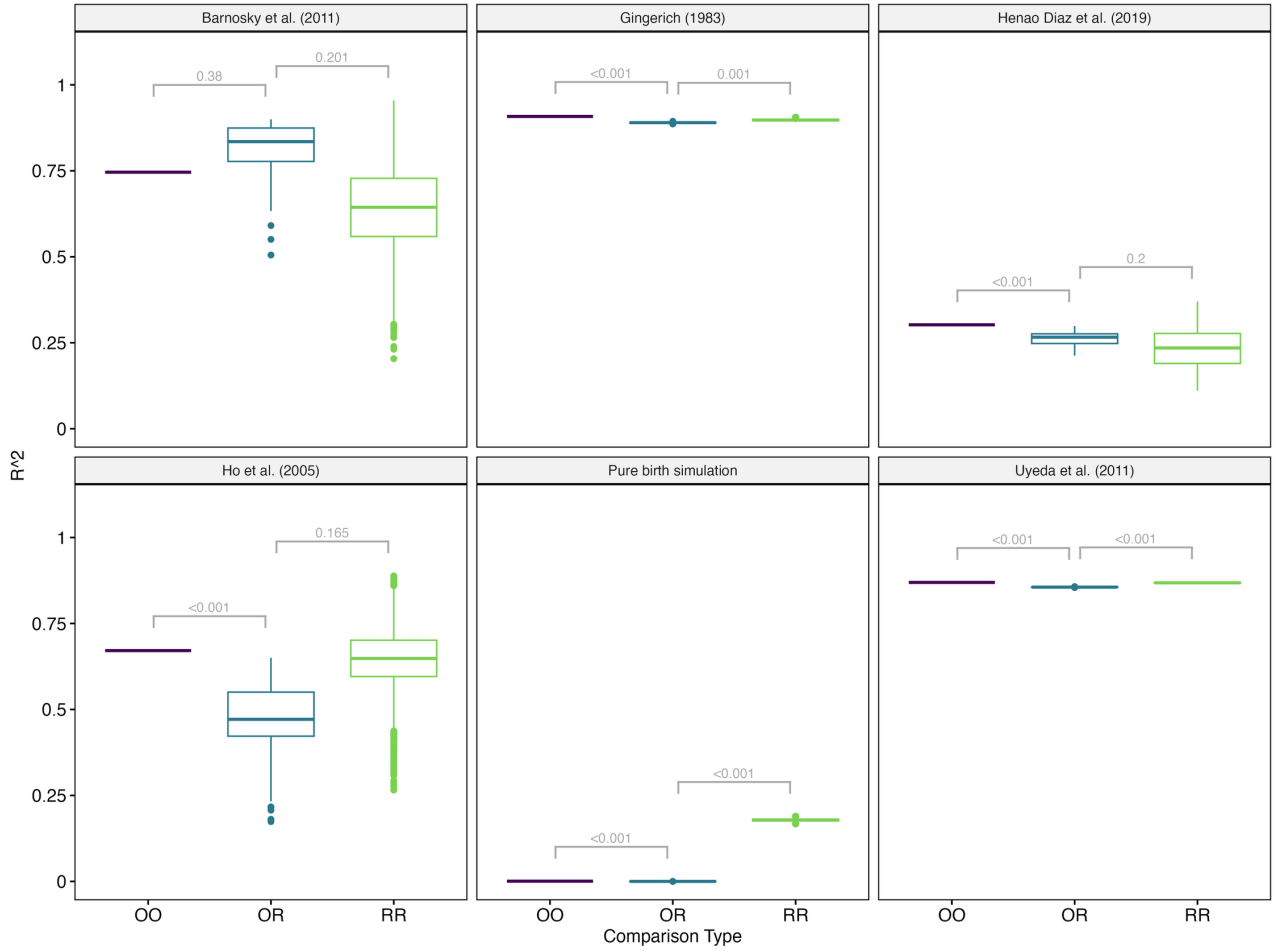

Figure 2 Comparisons of coefficient of determination (R^2) between the observed fit and fits from randomizing the numerator relative to the denominator of the empirical rates. First, we obtained the R^2 of the empirical original (O) data fit to the original (O) parameters (comparison type OO). Second, 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).

85.6% of the variation (Supplemental Table S1).

We hasten to acknowledge that some differences in the R^2 among empirical and randomized fits is expected given the inferred shifts in the contribution of three components through time. For example, while the empirical dataset of Uyeda *et al.* [15] is largely consistent with a hyperbolic distribution (Fig. 1), the rates nearer the maximum age are largely driven by the linear component—that is, the effect of $\left(\frac{\varepsilon_2-\varepsilon_1}{t_2-t_1}\right)$ is less than that of $\left(\frac{x_2-x_1}{t_2-t_1}\right)$ at deep times. Nevertheless, as a practical matter, the fact that we can take measurements like amount of change within *Daphnia* over a summer [22], the amount of change between late Cretaceous *Tyrannosaurus* and Triassic *Coleophysis* over 98 million years [23], and many datasets in between, shuffle these

values, and have the model fit nearly as well on the original data—including explaining over 80% of the variation—suggests that the apparent patterns arise from fitting noise. By contrast, the Yule dataset, where there is variance due to the model (a 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 at deeper times where there may be enough signal to overcome the hyperbolic pattern from noise. 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 [24,25].

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, should likely be robust to the statistical artifacts uncovered here. However, the traditional approach of simply taking rates and plotting versus times is fundamentally flawed and practically uninformative, even with corrections. One could easily add more complexity to our linear *hmb* model, for example (such as different rates in different discrete times, rates correlating with some external variable, and more), but at least with the amount of variation we see in the empirical datasets here we would not believe any results are biologically informative. Moreover, though our envisioned noise mechanism alone generates data nearly indistinguishable from empirical data, it is far from the only bias or source of error in these data. Estimates of time are biased and uncertain. There is also substantial ascertainment bias in which datasets scientists examine. Over short time periods, we only look at cases where we expect a detectable amount of change; over long time periods, we only compare things similar enough to make comparison sensible (e.g., body length between two species of dinosaurs, not body length between a dinosaur and a tree). For diversification analyses, clades below a certain size are often excluded, creating a bias in recovered

rates, especially for clades originating more recently.

An implication of our results is that any errors present in measurements of evolutionary change not only greatly affect rate estimation overall, but also any comparative test that examines rate shifts within a single phylogenetic tree. Consider a scenario where we want to compare evolutionary rates between a paraphyletic group of taxa and clade within it (such as seed size evolution in nonflowering plants versus angiosperms). Any inaccuracies in trait measurements or character state assignments within the tree could lead to the erroneous inference that the subclade has a higher rate, simply due to less overall time in terms of branch lengths to mitigate the effects of these errors (Fig. S10). This discrepancy also creates a concerning disconnect between the expected error rates for specific phylogenetic comparative models, derived from simulations, and their actual behavior in empirical settings when assessing rate differences within and among groups. That is, null models for many comparative tests may fit far worse in an empirical setting than might be indicated by careful simulation study conducted in the absence of measurement error [27], providing alternative explanations with artificially inflated support. Even with Bayesian approaches that attempt to integrate over uncertainty in results, data are still typically assumed to be free of errors, and this being incorrect will tend to lead to increased rates, especially where the magnitude of the effect of the error is high relative to the magnitude of the effect of the signal. This may also pose substantial issues to molecular and fossil dating analyses.

We anticipate that our findings will encourage further development of approaches that acknowledge measurement error when estimating rates. In models of continuous trait evolution, measurement error can be addressed by directly inputting the standard error of the species means for a trait to the model [28]. However, this practice is not yet standard, but at the very least, measurement error could be incorporated as a parameter to estimate in the model. The current standard approach forces it to be zero, which is a very strong, and likely very inaccurate, assumption. In other comparative models, especially those related to the evolution of discrete characters (e.g., nucleotides or higher-level phenotypic characters exhibiting a finite state space), rates are determined without methods to mitigate potential errors, like inaccuracies in character state assignments. Such errors have a clear impact of inflating rates over shorter time scales (e.g., Fig. 1 and Fig. S9). More generally, this points to the need to incorporate the potential effects of measurement error and other uncertainty in causing a pattern, not just noise. This has been dismissed as a potential cause of the pattern here [8] but both randomization and simulation presented here show how the empirical pattern can come directly from error in the numerator.

One unexplored issue is that while our study solely focused on errors in measurements of evolutionary change, errors in the temporal component underlying a rate can also significantly influence rate estimates. In certain scenarios, such as diversification models, we suspect errors in branch lengths might counteract the observed patterns, at least partially.

Therefore, given the potential biases in molecular age estimates [29-31], it is prudent to prioritize the development of integrative approaches that address errors not only in measurements of evolutionary change, but also in the underlying branch lengths.

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

Materials and Methods

Overall workflow. Other than the Yule simulations themselves (see below), all analyses were done in R 4.4.0 [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 ; DOI 10.5281/zenodo.12538611. The R package developed for this project, *hyperr8*, is at https://github.com/bomeara/hyperr8. 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 maximum likelihood estimates of turnover, τ, which is speciation + extinction, and extinction fraction, ε, which reflects the ratio of extinction rate to speciation rate. We conditioned the likelihood on survival (i.e., probability that the observed tree could have gone extinct by time, t). For each tree, we transformed estimates of τ and ε back to the original speciation and extinction variables.

In the case of Gingerich [1], Ho *et al.* [2], and Barnosky *et al.* [21] the original data were not readily available. We therefore used plotdigitizer.com to manually

digitize the points from the log-log plots, as they allowed easiest visibility of individual points. For Ho *et al.* [2] we recreated Fig. 1*A* from [7]. For Gingerich [1], in cases where there were multiple data points per symbol, and so we recorded one point for each replicate. For example, a symbol of "5" became five points at that location, whereas an "X" means 10 or more. In the latter case, we encoded these as ten points.

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

Least-squares model. We implemented a set of least-squares models into an R package, *hyperr8*, whose namesake is based on our expectation of hyperbolas and as an homage to Mike 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 reran the analyses to create denser plots that are easier to see. The results from the dentist for each data set are shown in Figures S1-6.

Randomization procedure. As described in the main text, we used a randomization procedure to generate a new empirical null to compare against our empirical fits. This involved first separating each empirical rate into its numerator and denominator components. Some of the datasets included information on the numerator (e.g., difference in log body size for datasets whose rates are measured in darwins) and the denominator (e.g., time separating the pair of values being compared), while others included only rate and time. For the latter, we imputed the numerator by multiplying the rate by time. We then shuffled the numerators at random across a given empirical data set, and then divided them by the unshuffled times to obtain a new set of rates and times. This was repeated 100 times for each dataset.

Rates derived solely from phylogenetic trees, such as those in the Henao Diaz *et al.* [3] dataset and our pure birth simulations, present a unique challenge for our randomization procedure. Typically, rates are assessed across the entire tree, which reflects the total time span of all branch lengths. However, the time indicated on the x-axis typically represents the age of the tree, which, although related [50], are far from perfectly correlated. Utilizing the age of the tree to calculate the numerator would significantly underestimate the rate. In such cases, we opted to use the total branch length of the tree to calculate both the numerator and the denominator when recalculating the rate, while still employing the tree's age for the x-axis and model fitting. We do note that for the Yule simulation dataset, we did re-run all analyses using total tree height as opposed to clade age and did not find any substantive and qualitative differences from the results presented in the main text (Fig. S7).

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

We also stress the importance of distinguishing biological significance from statistical significance, something long emphasized in the literature [51]. In the context of our work, a *p*-value is the probability of two distributions being exactly equal. For example, in the Uyeda *et al.* [15] dataset, it is clear from Fig. 1 in the main text that while the empirical dataset is

largely consistent with a hyperbolic distribution, the few points of maximum age are more consistent with a linear distribution. This is reflected in Fig. 2 in the main text, where the fit of the empirical data to predictions from randomized data are slightly worse than fit to empirical data predictions. Due to the large size of the dataset, there is little variance in the former, so the latter is outside the observed distribution and thus is quite significantly different. Biologically, though, 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 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 $\lambda = 0.10$ 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.

Yule sensitivity to excluding zero rates. As mentioned above, most diversification studies and simulations typically exclude zero rates from their analyses, which can introduce bias. Excluding zero rates fails to properly balance higher rates nearer the present with the higher instances of zero-rate trees over shorter time scales (see Fig. S7). For our least-squares model, removing zero rates from our Yule simulation dataset necessarily shifted the signal towards the

hyperbolic component (Fig. S7). We were concerned that if our randomization procedure incorrectly attributed exponentially increasing rates towards the present to errors when there were none, it would also raise questions about the empirical results, which also omitted zero rates. To test this, we reran all analyses on the Yule simulation dataset with the zero rates excluded. As expected, the strongest signal nearest the present was for the hyperbolic component (Fig. S7). In this case, the randomization creates a pattern very different from the hyperbola as a consequence of noise.

Within and among tree rate bias. One implication of our study is that errors in measuring evolutionary changes not only greatly affect rate estimation overall, but also any comparative test that examines rate shifts within a single tree. For instance, consider a scenario where we want to compare evolutionary rates between two sister clades that differ in age. Any inaccuracies in branch lengths or character state assignments within the tree could lead to the erroneous inference that the younger clade has a higher rate, simply due to less overall time to mitigate the effects of errors.

To demonstrate this, we conducted simulations of character state transitions and estimated transition rates with and without error. We ran 1000 simulations each on two types of 64-taxon trees that represented extremes in tree balance: pectinate (i.e., "caterpillar" trees) and perfectly balanced trees, where diversity is evenly distributed between sister clades. The age of 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 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.

Evaluating fit to simulated distributions. To assess how well the new method in the hyperr8 package performs, we generated data under a variety of scenarios. The first was a constant rate of 0.4 changes per time unit (which went from 0.01 to 50-time units). The second was a rate that started at 0.6 changes per time unit and increased with a slope of 0.01 per time unit. The third was a rate that started at 1 change per time unit and decreased with a slope of -0.01 per time unit. The fourth was a sine wave with a period of 26-time units. This is something that our method should perform poorly at, as such periodicity is not a component of the model (this model is inspired by [53]). For added realism, we added error through using times with noise (standard deviation of 10% of the mean). Ten replicate datasets of 500 and 5000 data points were simulated by taking the expected rates, multiplied them by time to get the numerator, and then sampled from a normal distribution with sd of 0.000001, 0.01, 0.1, and 1 and mean at the predicted numerator, then divided by time. Some of this variation seems quite large, but it spans the empirical data. For example, the Uyeda et al. [15] dataset had 79 time points where the inferred rate at that time point had a coefficient of variation (ratio of standard deviation to mean) greater than 1. Results are shown in Figure S11, Figure S12, and Table S2.

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 [54-58]. 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 because 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.

Acknowledgments

This work presented here was substantially influenced by James Boyko's advocacy in challenging the artifactual nature of age rate scaling. Conversations with Dan Simberloff, Nina Fefferman, Paul Armsworth, Xingli Giam, Naomi O'Meara, Adam Siepielski, and Andy Alverson were also of substantial help in generating ideas and providing constructive feedback. This study was funded by grants from the National Science Foundation (grant DEB−1916539 awarded to B.C.O. and grant DEB−1916558 awarded to J.M.B.).

References

1. Gingerich PD. Rates of evolution: effects of time and temporal scaling. Science 1983; 222: 159-161.
2. Ho SYW, Phillips MJ, Cooper A, Drummond AJ. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Molecular Biology and Evolution 2005; 22: 1561-1568.
3. Henao Diaz L, Harmon LJ, Sugawara MTC, Pennell MW. 2019. Macroevolutionary diversification rates show time dependency. Proceedings of the National Academy of Sciences, U.S.A. 2019; 116: 7403-7408.

4. Harmon LJ, Pennell MW, Henao Diaz LF, Rolland J, Sipley BN, Uyeda JC. Causes and consequences of apparent time scaling across all estimated evolutionary rates. Annual Review of Ecology, Evolution, and Systematics 2021; 52: 587-609.
5. Hallam A. Evolutionary size increase and longevity in Jurassic bivalves and ammonites. Nature 1975; 258: 493-496.
6. Gould SJ. Smooth curve of evolutionary rate: a psychological and mathematical artifact. Science 1983; 226: 994-996.
7. Rolland J, Henao Diaz LF, Doebeli M, Germain R, Harmon LJ, Knowles LL, et al. Conceptual and empirical bridges between micro- and macroevolution. Nature Ecology and Evolution 2023; 7: 1181-1193.
8. Hansen TF. Three modes of evolution? Remarks on rates of evolution and time scaling. Journal of Evolutionary Biology 2024; Forthcoming. doi: https://doi.org/10.1093/jeb/voae071.
9. Sheets HD, Mitchell CE. Uncorrelated change produces the apparent dependence of evolutionary rate on interval. Paleobiology 2001; 27: 429-445.
10. Hunt G. Fitting and comparing models of phyletic evolution: random walks and beyond. Paleobiology 2006; 32: 578-601.
11. Hunt G. Evolutionary patterns within fossil lineages: model-based assessment of modes, rates, punctuations and process. In Kelley PH, Bamback RK, editors. Evolution to Geobiology: Research Questions Driving Paleontology at the Start of a New Century. Paleontological Society; 2008. pp. 117-131.
12. Pearson K. 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 1897; 60: 359-367.
13. Chayes F. On ratio correlation in petrography. Journal of Geology 1949; 57: 239-254.
14. De Lisle SP, Svensson EI. Evolutionary change is remarkably constant across timescales. bioRxiv [Preprint] 2024. Available from: https://doi.org/10.1101/2023.12.02.569704.
15. Uyeda JC, Hansen TF, Arnold SJ, Pienaar J. The million-year wait for macroevolution bursts. Proceedings of the National Academy of Sciences, U.S.A. 2011; 108: 15908-15913.
16. Jukes TH, Cantor CR. Evolution of protein molecules. In Munro HN. Mammalian Protein Metabolism. New York: Academic Press; 1969. pp. 21-132.
17. Pagel M. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proceedings of the Royal Society, London 1994; 255: 37-45
18. Nee S, May RM, Harvey PH, The reconstructed evolutionary process. Philosophical Transactions of the Royal Society, London 1994; 344: 305-311.
19. Felsenstein J. Phylogenies and the comparative method. American Naturalist 1985; 125: 1-15
20. Haldane JBS. Suggestions as to the quantitative measurement of rates of evolution. Evolution 1949; 3: 51-56.
21. Barnosky AD, Matzke N, Tomiya S, Wogen GOU, Swartz B, Quental TB, et al. Has the Earth's sixth mass extinction already arrived? Nature 2011; 471: 51-57.

22. Tessier AJ, Young A, Leibold M. Population dynamics and body-size selection in *Daphnia*. Limnology and Oceanography 1992; 37: 1-13.
23. Colbert EH. Evolution of horned dinosaurs. Evolution 1948; 2: 145-163.
24. Kamionkowski M, Riess AG. The Hubble tension and early dark energy. Annual Review of Nuclear and Particle Science 2023; 73: 153-180.
25. Cardona W, Kunz M, Pettorino V. Determining H_o with Bayesian hyper-parameters. *Journal of Cosmology and Astroparticle Physics* 2017; 3: 56.
26. Cooke R, Says F, Andermann T, Blackburn TM, Steinbauer MJ, Antonelli A, Faurby S. Undiscovered bird extinctions obscure the true magnitude of human-driven extinction waves. Nature Communications 2023; 14: 8116.
27. Silvestro D, Kostikova A, Litsios G, Pearman PB, Salamin N. Measurements errors should always be incorporated in phylogenetic comparative analysis. Methods in Ecology and Evolution 2015; 6: 340-346.
28. Ives AR, Midford PE, Garland, Jr. T. Within-species variation and measurement error in phylogenetic comparative methods. Systematic Biology 2007; 56: 252-270.
29. Beaulieu JM, O'Meara BC, Crane P, Donoghue MJ. Heterogeneous rates of molecular evolution and diversification could explain the Triassic age estimate for angiosperms. Systematic Biology 2015; 64: 869-878.
30. Brown JW, Smith SA. The past sure is tense: on interpreting phylogenetic divergence time estimates. Systematic Biology 2018; 67: 340-353.
31. Budd GE, Mann RP. Two notorious nodes: a critical examination of relaxed molecular clock age estimates of the bilaterian animals and placental mammals. Systematic Biology 2023; Forthcoming. doi: https://doi.org/10.1093/sysbio/syad057.
32. Simberloff DS. Taxonomic diversity of island biotas. Evolution 1970; 24: 23-47.
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. Landau WM. The *targets* R package: a dynamic Make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. Journal of Open Source Software 2021; 6: 2959.
32. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.
33. Paradis E, Schliep K. *ape* 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 2019; 35: 526-528.
34. Wickham H, François R, Henry L, Müller K, Vaughan D. *dplyr*: A Grammar of Data Manipulation. R package version 1.1.4. 2023. doi: https://CRAN.R-project.org/package=dplyr.
35. Maechler M. *Rmpfr*: R MPFR - Multiple Precision Floating-Point Reliable. R package version 0.9-2. 2023. doi: https://CRAN.R-project.org/package=Rmpfr.
36. Wickham H, Vaughan D, Girlich M. *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. Available from: https://nlopt.readthedocs.io/en/latest
38. Boyko JD, O'Meara BC. dentist: quantifying uncertainty by sampling points around maximum likelihood estimates. Methods in Ecology and Evolution 2024. Forthcoming. doi: https://doi.org/10.1111/2041-210X.14297.
39. Wickham H, Pedersen T, Seidel D. *scales*: Scale Functions for Visualization. R package version 1.3.0. 2023. doi: https://CRAN.R-project.org/package=scales.
40. Stadler T. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. Journal of Theoretical Biology 2009; 261: 58-66.
41. Stadler T. Simulating trees on a fixed number of extant species. Systematic Biology 2011; 60: 676-684.
42. Sanderson MJ. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 2003; 19: 301-302.
43. Burnham KP, Anderson DR. Model Selection and Multimodal Inference. New York: Springer-Verlag Press; 2002.
44. Rowan T. Functional stability analysis of numerical algorithms. Ph.D Thesis, University of Texas, Austin 1990.
45. Powell MJD. The BOBYQA algorithm for bound constrained optimization without derivatives. Department of Applied Mathematics and Theoretical Physics, Cambridge England, technical report NA2009/06. 2009.
46. Powell MJD. The NEWUOA software for unconstrained optimization without derivatives. Proc. 40th Workshop on Large Scale Nonlinear Optimization. Erice, Italy. 2004.
47. Steel M, Mooers A. The expected length of pendant and interior edges of a Yule tree. Applied Mathematics Letters 2010; 23: 1315-1319.
48. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biological Reviews 2007; 82: 591-605.
49. Nee S. Inferring speciation rates from phylogenies. Evolution 2001; 55: 661-668.
50. Beaulieu JM, O'Meara BC. 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 2018; 105: 417-432.
51. Beaulieu JM, O'Meara BC, Donoghue MJ. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. Systematic Biology 2013; 62: 725-737.
52. O'Meara BC, Ane C, Sanderson MJ, Wainwright PC. Testing for continuous trait evolution using likelihood. Evolution 2006; 60: 922-933.
53. Raup DM, Sepkoski Jr, JJ. Periodicity of extinctions in the geologic past. Proceedings of the National Academy of Sciences 1984; 81: 801-805.
54. Beaulieu JM, Jhwueng D-C, Boettiger C, O'Meara BC. Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. Evolution 2012; 66: 2369-2382.

55. Beaulieu JM, O'Meara BC. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Systematic Biology 2016; 65: 583-601

Supporting Information

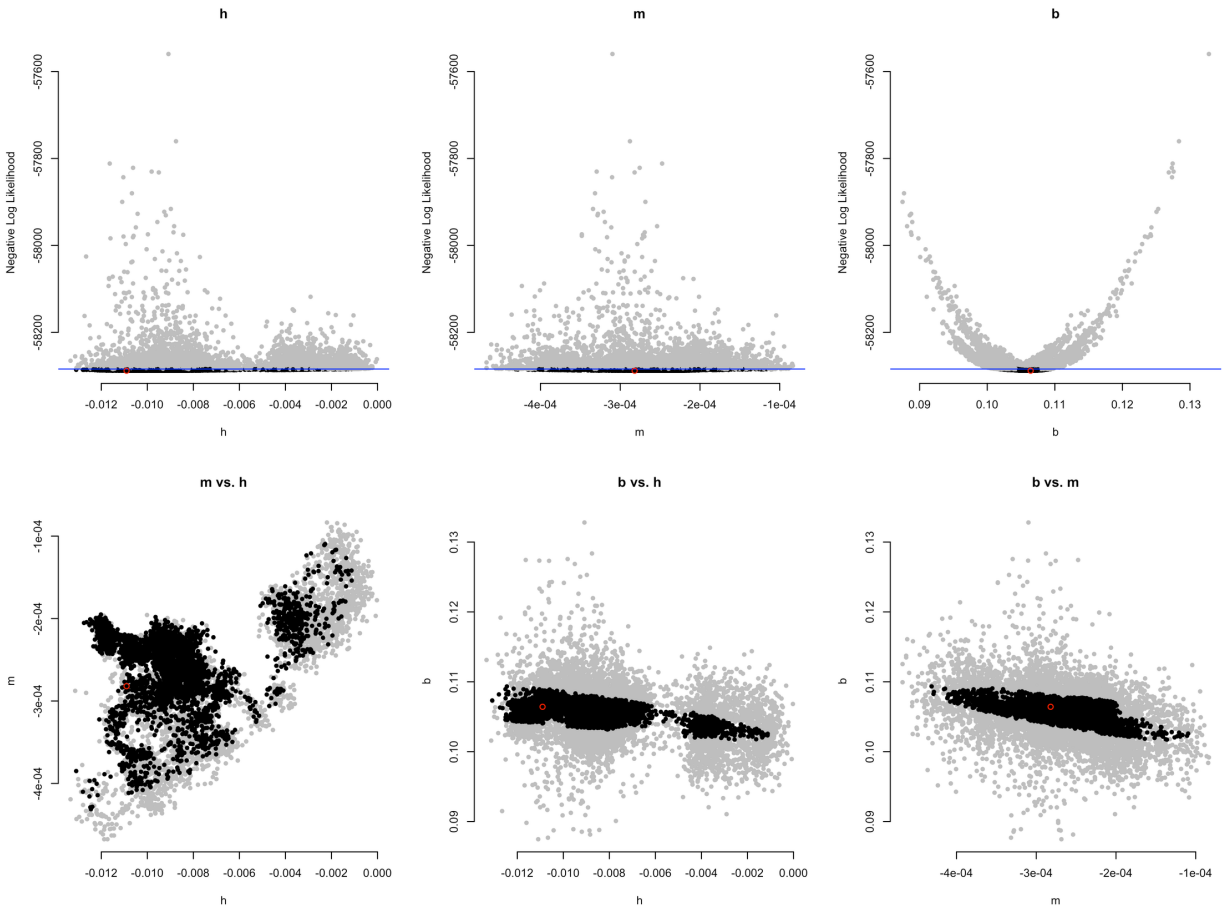

Figure S1. Overall uncertainty in the individual estimates of h, m, and b for the Yule simulation dataset.

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.

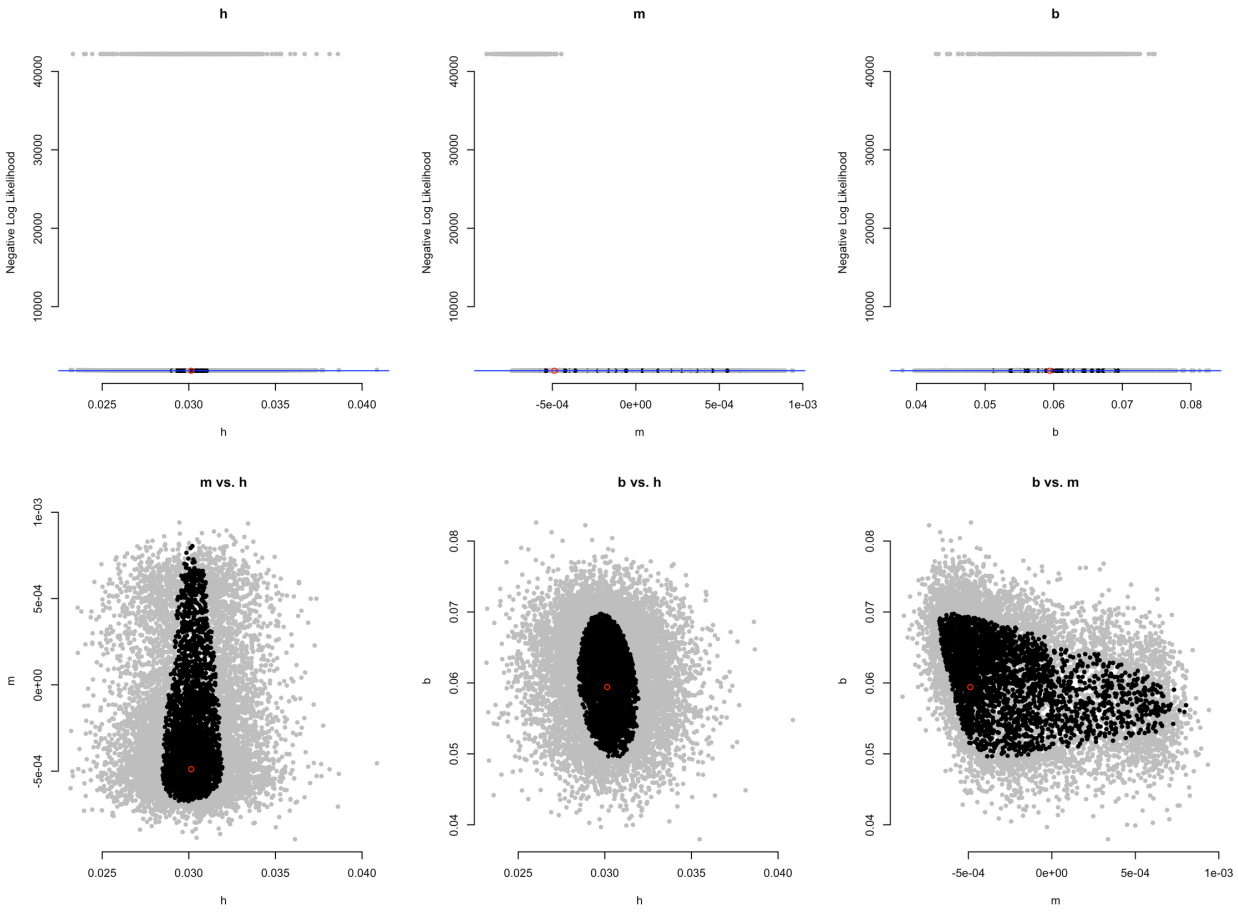

Fig. S2. Same dentist plot as in Fig. S1 but showing parameter uncertainty for the Uyeda et al. [15] dataset.

The best model was *hmb* so all parameters were free to vary.

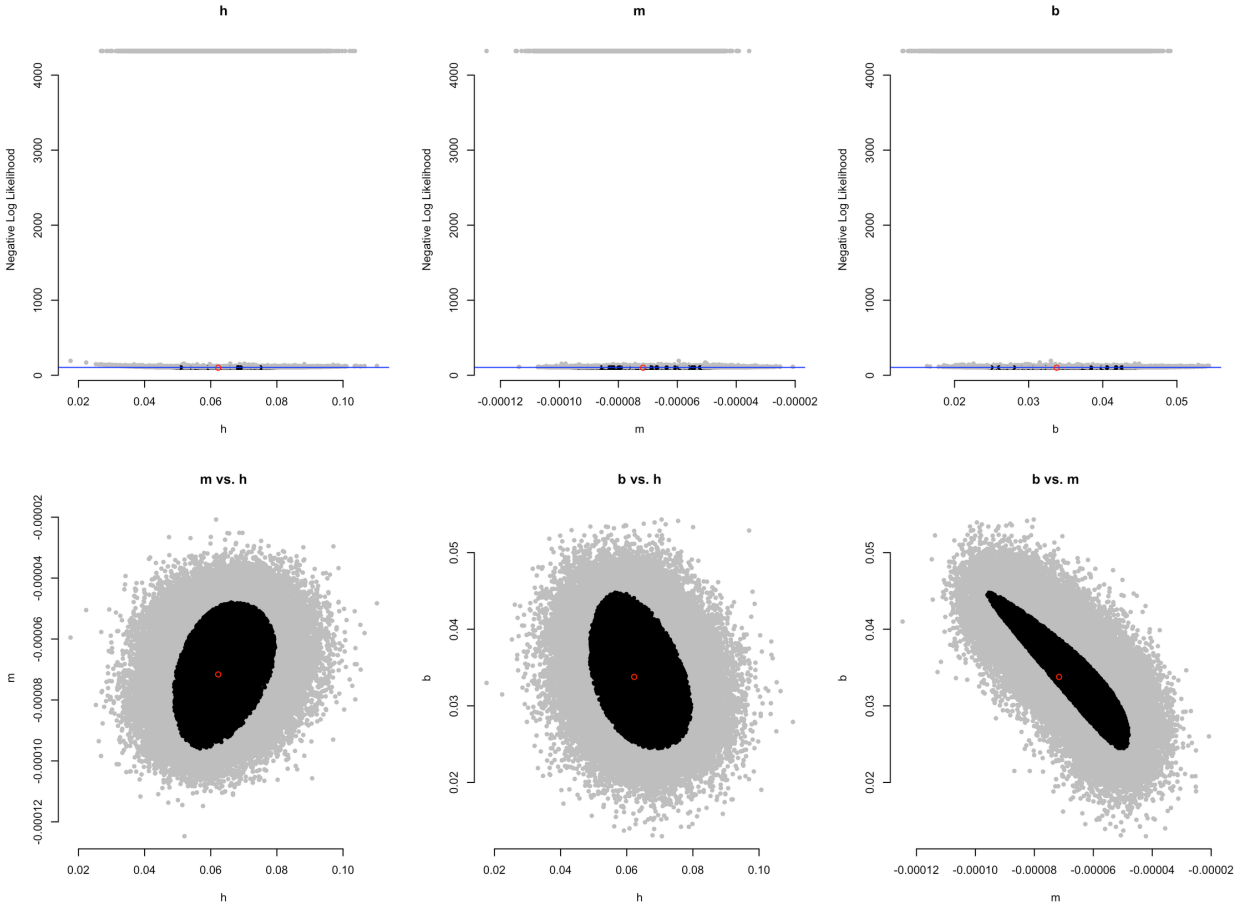

Fig. S3. Same dentist plot as in Fig. S1 but showing parameter uncertainty for the Gingrich [1] dataset.

The best model was *hmb* so all parameters were free to vary.

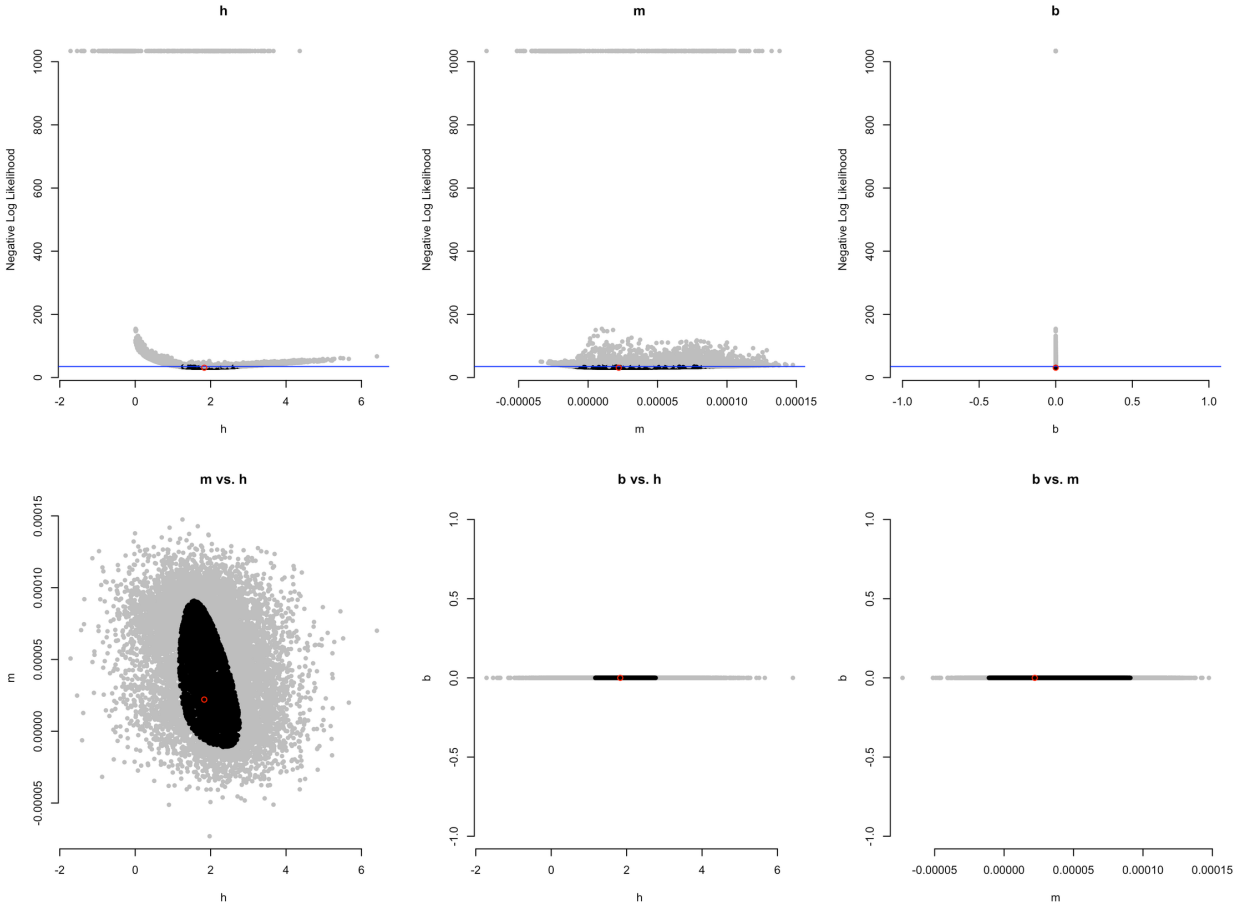

Fig. S4. Same dentist plot as in Fig. S1 but showing parameter uncertainty for the Henao Diaz et al. [3] dataset.

The best model was *hm0* so the *b* parameter is fixed at zero.

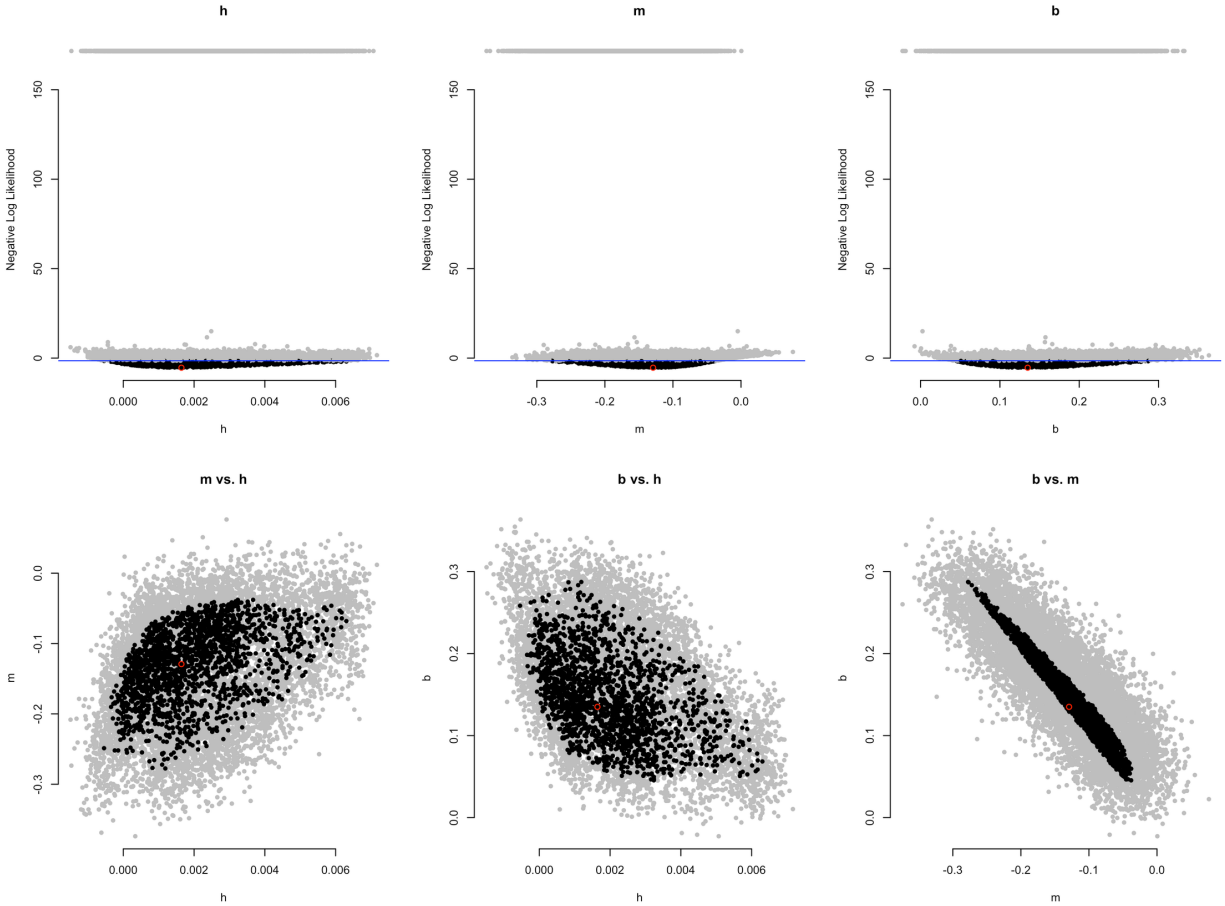

Fig. S5. Same dentist plot as in Fig. S1 but showing parameter uncertainty for the Ho et al. [2].

The best model was *hmb* so all parameters were free to vary.

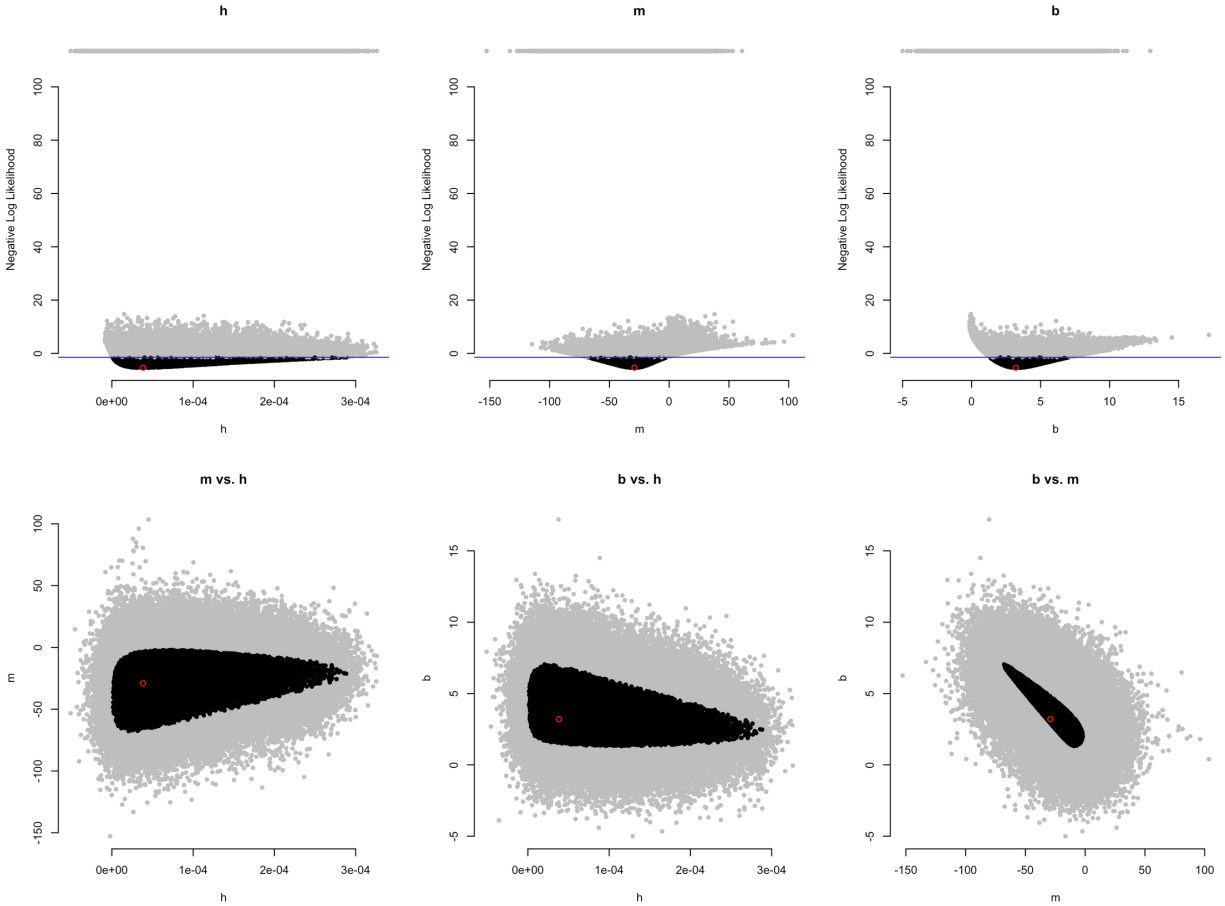

Fig. S6. Same dentist plot as in Fig. S1 but showing parameter uncertainty for the Barnosky et al. [21].

The best model was *hmb* so all parameters were free to vary.

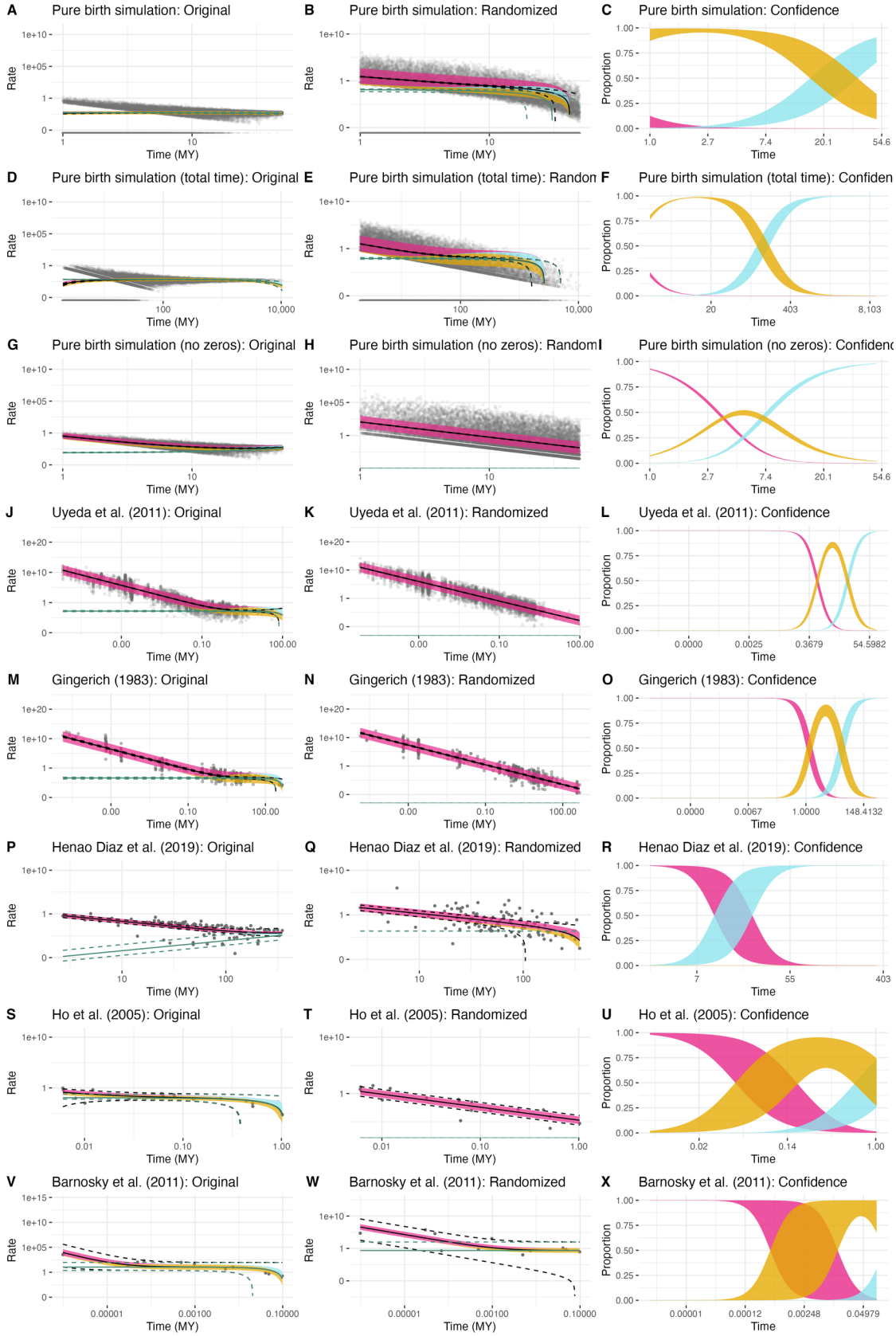

Fig. S7. Figure 1, but with additional ways to handle the Yule simulation dataset presented in the main text.

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.

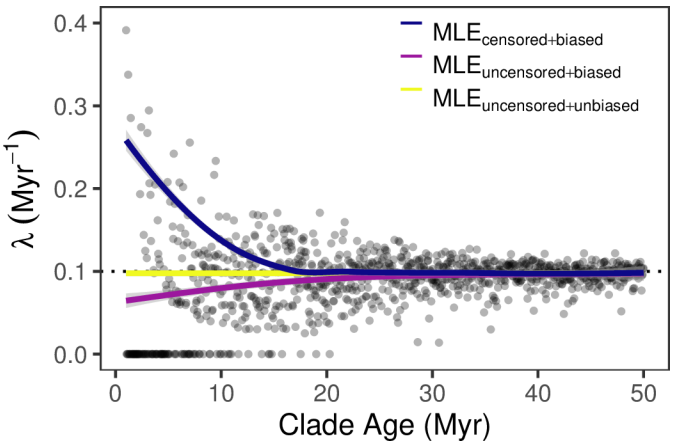

Fig. S8. Trends of the Yule birth rate through time under various conditions of the data and the estimator.

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, $\hat{\lambda}$, we find that rates exponentially increase towards the present (blue line, MLEcensored+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, MLEuncensored+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, MLEuncensored+unbiased), which is consistent with the simulation scenario.

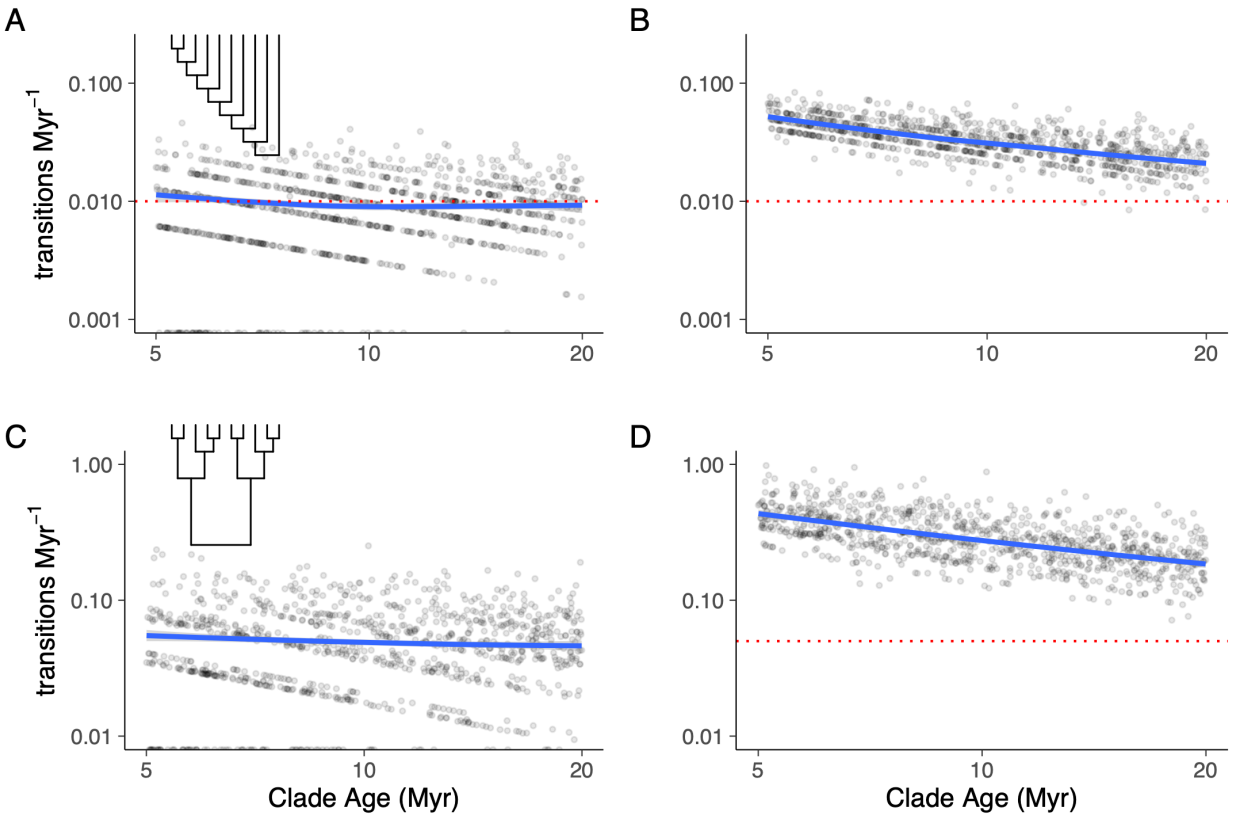

Fig. S9. Log-linear plot of estimated transition rates with and without error as a function of clade age obtained from simulated character state transitions.

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.

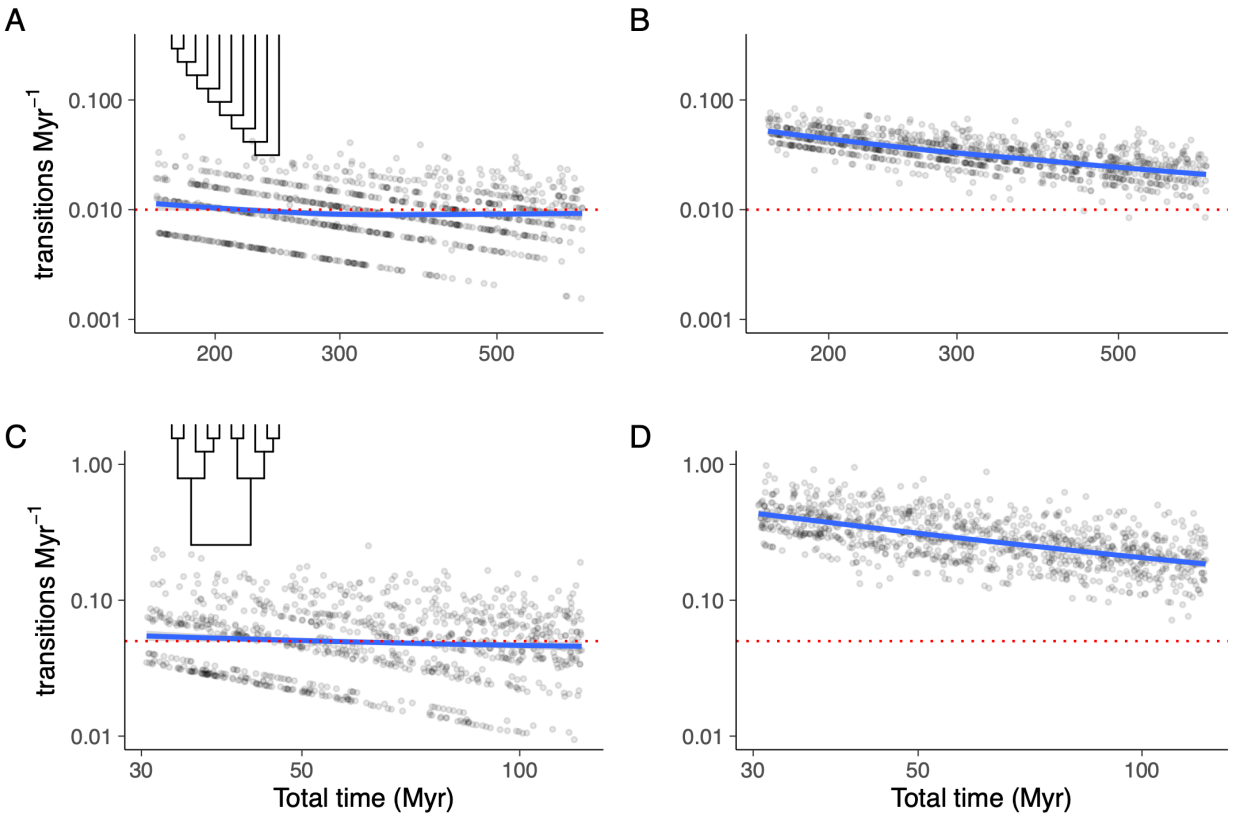

Fig. S10. Log-linear plot of estimated transition rates with and without error as a function of sum of all branch lengths obtained from simulated character state transitions.
Note that this figure is similar to Fig. S9 but replaces clade age on the vertical axis with the total time represented by a given 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.

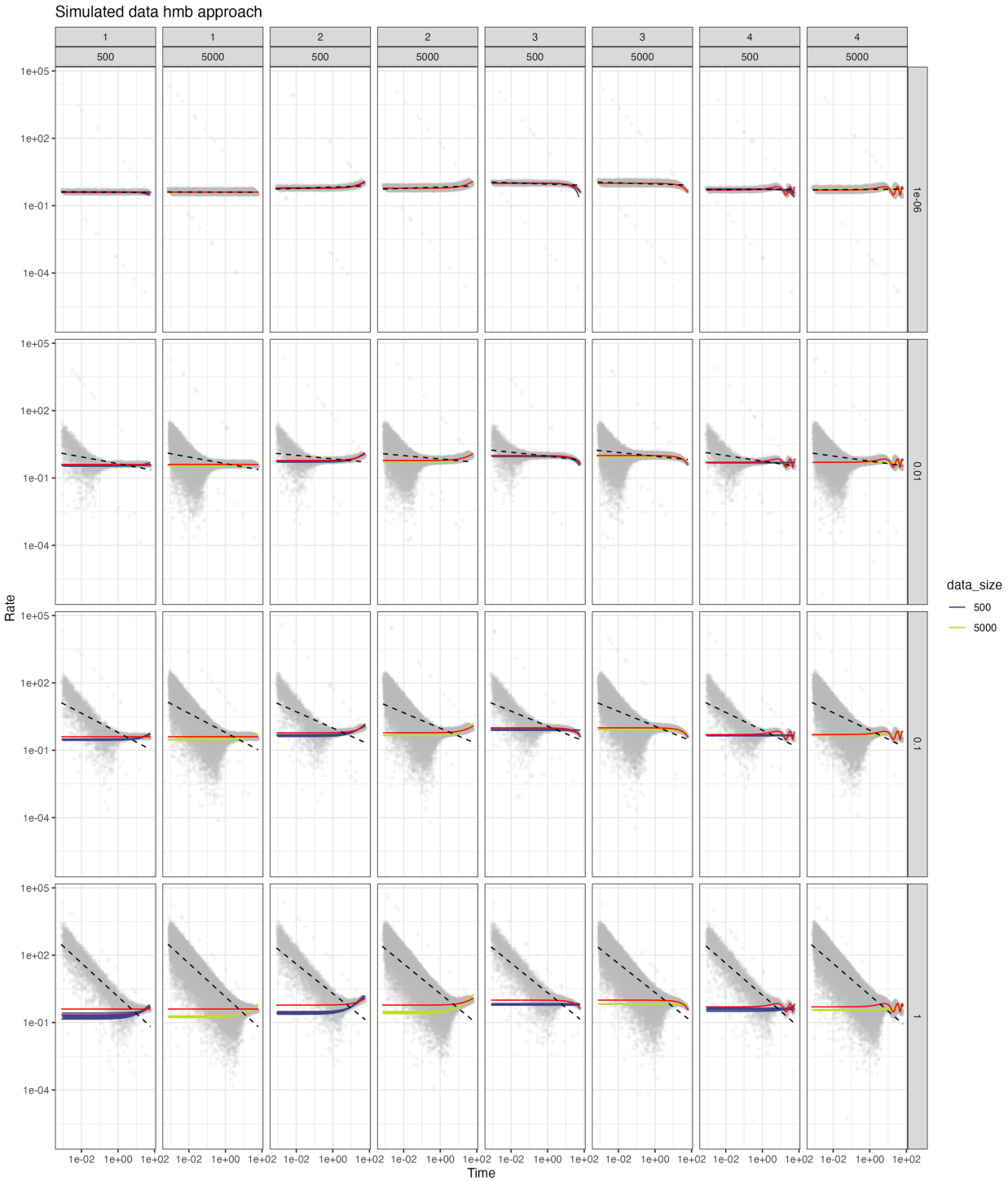

Fig. S11. Simulation results for different scenarios of the *hmb* model described in the main text.

Each dot represents a simulated comparison. Within each subplot, the x axis represents time and the y axis empirical rate, both on a log scale. The red line represents the true rate at that time under the model, the other solid lines the reconstructed rates under the best-fitting hmb model for each replicate dataset. The dashed line shows a linear fit to the points. The subgraphs are arranged in columns by generating models 1, 2, 3, and 4 and by dataset sizes of 500 or 5000

points and in rows by the standard deviation used in the simulation. Model 1 is a constant rate; models 2 and 3 are increasing and decreasing rates with increased time, respectively; model 4 is a sine wave with a periodicity of 26 million years. Note that with low measurement error, the clouds of points resemble the generating model (red line) but with increasing amounts of noise the clouds begin to resemble the points found in the empirical datasets.

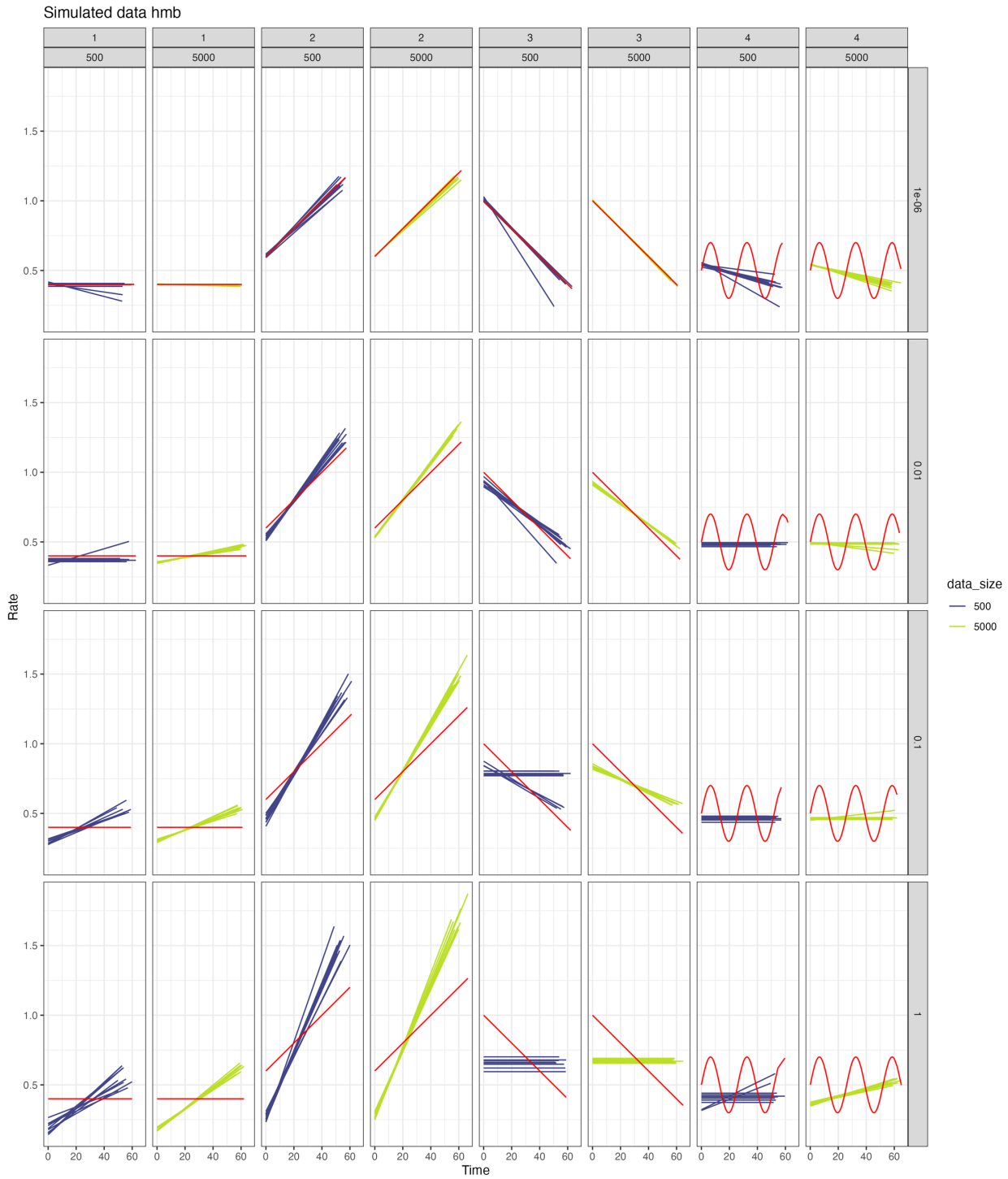

Fig. S12. This shows multiple simulations and is arranged in the same way as Figure S11. Each purple or green line represents the hmb model fit for the generating rate from a different replicate; the true generating rate is in red. Note that the axes in this plot are not log transformed.

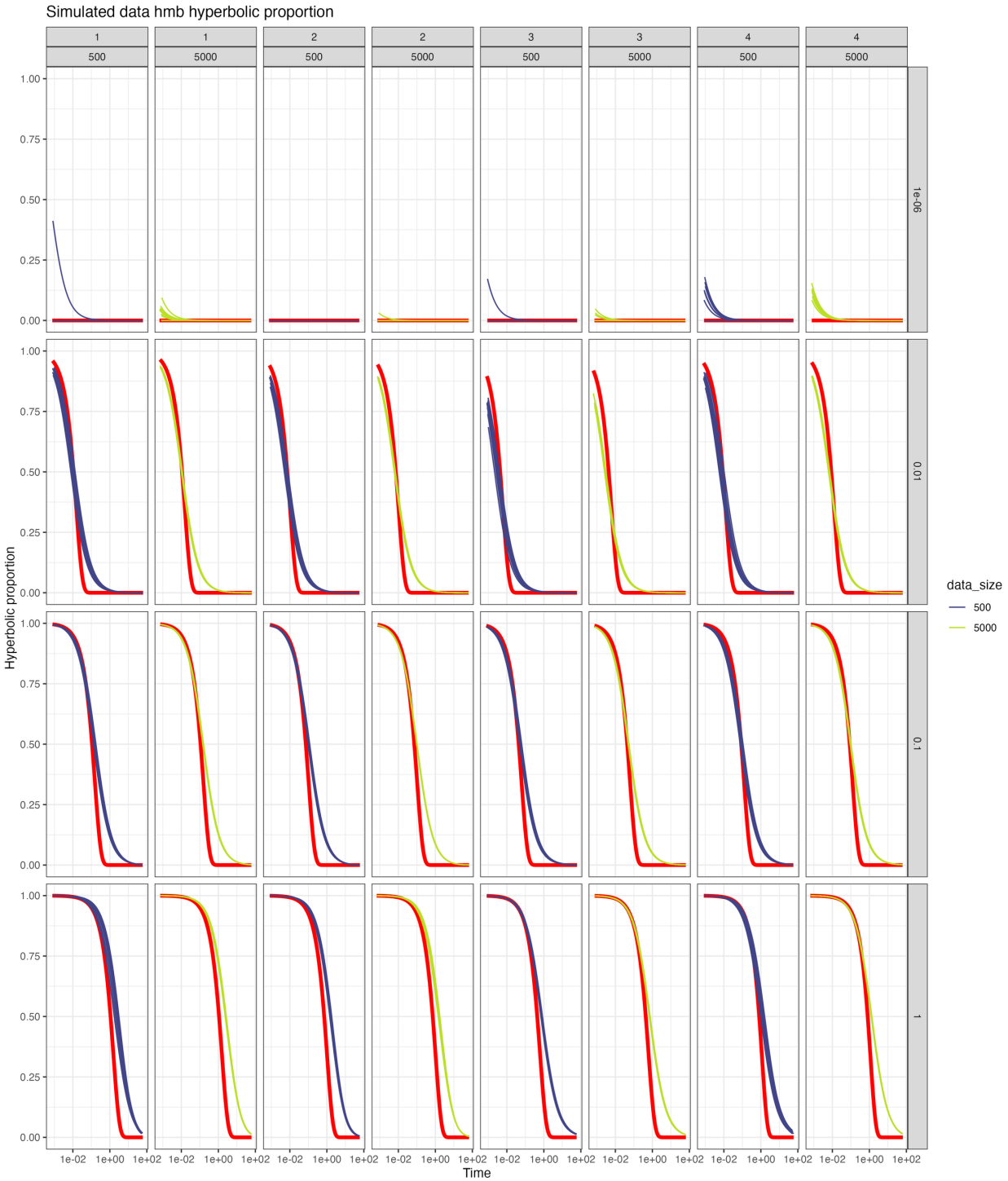

Fig. S13. Plot arranged as in Figure S12 but showing the estimate of how much of the empirical rate stemmed from the hyperbolic process (equivalent to measurement error) as solid purple or green lines versus the median across the simulations (red line).
The match is not perfect, but *hmb* generally performs well despite in some cases (like model 4) not being able to match the complexity of the generating model.

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.

See biorxiv supplemental files; also
https://github.com/bomeara/hyperbolic_rates/blob/main/hyperanalysis/supplemental_table.csv

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

See biorxiv supplemental files; also
https://github.com/bomeara/hyperbolic_rates/blob/main/hyperanalysis/hyperr8_animation.gif