Online Appendix for: Modeling the Evolution of Rates of Continuous Trait Evolution

B. S. Martin[1,*], G. S. Bradburd[2], L. J. Harmon[3], and M. G. Weber[1]

[1] *Department of Plant Biology, Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, MI 48824, USA*

[2] *Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA*

[3] *Department of Biological Sciences, Institute for Bioinformatics and Evolutionary Studies (IBEST), University of Idaho, Moscow, ID 83843, USA*

[*] *Correspondence to be sent to: bruce.stagg.martin@gmail.com*

Supplemental Tables and Figures

Table S1. *Cetacean body length data and associated references used for empirical example*

species	length (m)	reference
Balaena mysticetus	18.0	Slater et al., 2010
Balaenoptera acutorostrata	10.7	Slater et al., 2010
Balaenoptera bonaerensis	10.2	Konishi et al., 2008
Balaenoptera borealis	16.1	Slater et al., 2010
Balaenoptera edeni	15.4	Slater et al., 2010
Balaenoptera musculus	33.6	Slater et al., 2010
Balaenoptera omurai	10.7	Slater et al., 2010
Balaenoptera physalus	21.2	Slater et al., 2010
Berardius arnuxii	8.9	Slater et al., 2010
Berardius bairdii	12.0	Slater et al., 2010
Caperea marginata	6.2	Slater et al., 2010
Cephalorhynchus commersoni	1.5	Slater et al., 2010
Cephalorhynchus eutropia	1.5	Molina and Oporto, 1993
Cephalorhynchus heavisidii	1.7	Slater et al., 2010
Cephalorhynchus hectori	1.5	Slater et al., 2010
Delphinapterus leucas	3.8	Slater et al., 2010
Delphinus capensis	2.5	Plön et al., 2012
Delphinus delphis	2.3	Slater et al., 2010
Eschrichtius robustus	14.6	Slater et al., 2010
Eubalaena australis	13.9	Slater et al., 2010
Eubalaena glacialis	13.7	Slater et al., 2010
Eubalaena japonica	17.4	Fortune et al., 2021
Feresa attenuata	2.4	Slater et al., 2010
Globicephala macrorhynchus	4.8	Slater et al., 2010
Globicephala melas	5.1	Slater et al., 2010
Grampus griseus	3.7	Slater et al., 2010
Hyperoodon ampullatus	7.9	Slater et al., 2010
Hyperoodon planifrons	7.5	Slater et al., 2010
Indopacetus pacificus	7.2	Slater et al., 2010
Inia geoffrensis	2.0	Slater et al., 2010
Kogia breviceps	3.4	Slater et al., 2010
Kogia sima	2.4	Slater et al., 2010
Lagenodelphis hosei	2.6	Slater et al., 2010
Lagenorhynchus albirostris	3.0	Slater et al., 2010
Leucopleurus acutus	2.4	Slater et al., 2010
Lipotes vexillifer	2.0	Slater et al., 2010
Lissodelphis borealis	2.3	Slater et al., 2010
Lissodelphis peronii	2.3	Baker, 1981

species	length (m)	reference
Megaptera novaeangliae	18.0	Slater et al., 2010
Mesoplodon bidens	5.1	Slater et al., 2010
Mesoplodon bowdoini	4.5	Slater et al., 2010
Mesoplodon carlhubbsi	5.3	Mead et al., 1982
Mesoplodon densirostris	4.7	Slater et al., 2010
Mesoplodon europaeus	5.2	Slater et al., 2010
Mesoplodon ginkgodens	4.9	Slater et al., 2010
Mesoplodon grayi	5.3	Slater et al., 2010
Mesoplodon hectori	4.4	Slater et al., 2010
Mesoplodon hotaula	4.8	Dalebout et al., 2014
Mesoplodon layardii	6.2	Slater et al., 2010
Mesoplodon mirus	5.1	Slater et al., 2010
Mesoplodon perrini	4.4	Dalebout et al., 2002
Mesoplodon peruvianus	3.7[a]	Reyes et al., 1991
Mesoplodon stejnegeri	5.7	Slater et al., 2010
Mesoplodon traversii	5.3	Thompson et al., 2012
Monodon monoceros	4.3	Slater et al., 2010
Neophocaena phocaenoides	1.4	Slater et al., 2010
Orcaella brevirostris	2.2	Slater et al., 2010
Orcaella heinsohni	2.2	Arnold and Heinsohn, 1996
Orcinus orca	7.9	Slater et al., 2010
Peponocephala electra	2.8	Lodi et al., 1990
Phocoena dioptrica	1.9	Slater et al., 2010
Phocoena phocoena	1.9	Slater et al., 2010
Phocoena sinus	1.1	Slater et al., 2010
Phocoena spinipinnis	1.7	Slater et al., 2010
Phocoenoides dalli	1.9	Slater et al., 2010
Physeter macrocephalus	11.0	Slater et al., 2010
Platanista gangetica	2.5	Slater et al., 2010
Pontoporia blainvillii	1.5	Slater et al., 2010
Pseudorca crassidens	5.1	Slater et al., 2010
Sagmatias australis	2.1	Slater et al., 2010
Sagmatias cruciger	1.8	Slater et al., 2010
Sagmatias obliquidens	2.4	Slater et al., 2010
Sagmatias obscurus	1.9	Slater et al., 2010
Sotalia fluviatilis	1.5	Slater et al., 2010
Sotalia guianensis	2.1	Barros, 1991
Sousa chinensis	2.4	Slater et al., 2010
Sousa teuszii	2.5	Jefferson and Rosenbaum, 2014
Stenella attenuata	2.1	Slater et al., 2010
Stenella clymene	1.9	Slater et al., 2010
Stenella coeruleoalba	2.3	Slater et al., 2010
Stenella frontalis	2.1	Slater et al., 2010

species	length (m)	reference
Stenella longirostris	2.0	Slater et al., 2010
Steno bredanensis	2.6	Slater et al., 2010
Tasmacetus shepherdi	6.5	Slater et al., 2010
Tursiops aduncus	2.1	Slater et al., 2010
Tursiops australis	2.8[b]	Charlton-Robb et al., 2011
Tursiops truncatus	2.4	Slater et al., 2010
Ziphius cavirostris	6.4	Slater et al., 2010

[a] from male specimen because no mature females were measured

[b] sex not reported

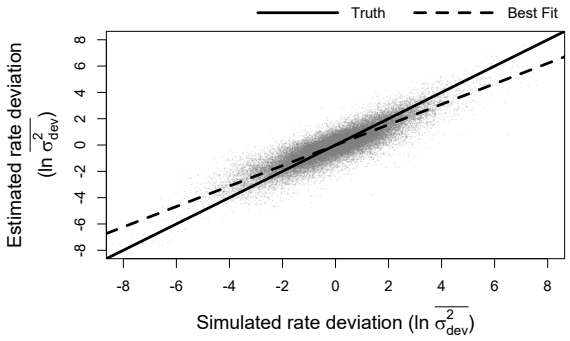

Figure S1. Relationship between simulated and estimated branchwise rate deviation parameters (ln $\overline{\sigma^2_{dev}}$). The solid line represents the position of the true branchwise rate deviations, while the shallower, dashed line represents the observed line of best fit for these data.

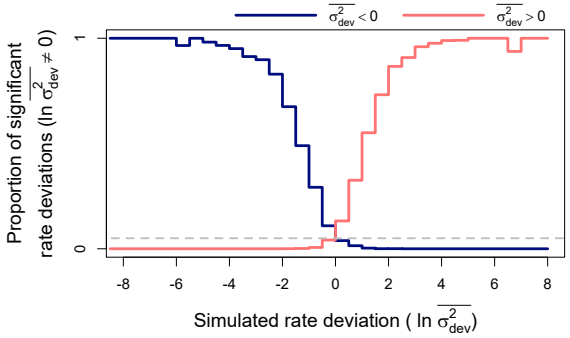

Figure S2. Power and error rates for branchwise rate parameters (ln $\overline{\sigma^2}$) under relaxed significance thresholds (posterior probability < 0.1 or > 0.9). Lines depict changes in proportions of branchwise rates considered anomalously slow (in dark blue) or fast (in light red) as a function of simulated rate deviations (ln $\overline{\sigma^2_{dev}}$). These results combine all fits to simulated data that detected rate variance ($\sigma^2_{\sigma^2}$) significantly greater than 0. The proportions are equivalent to power when the detected rate deviation is of the same sign as the true, simulated deviation (left of 0 for anomalously slow rates in dark blue and right for anomalously fast rates in light red), and to error rate when the detected and true rate deviations are of opposite signs. Here, significant rate deviations for simulated rate deviations that are exactly 0 are considered errors regardless of sign.

Approximating Geometric Brownian Motion Time-Averages

Our model seeks to model rates (σ^2) as "evolving" under a trended Geometric Brownian Motion (GBM)-like process, whereby the natural log of rates evolve in a trended Brownian Motion (BM)-like manner. Unfortunately, this requires an expression for the probability distribution of GBM time-averages along each branch in the phylogeny. Expressions for such distributions are infamously intractable, necessitating approximate solutions (Dufresne, 2004; Lepage et al., 2007). For our model, we use a multivariate log-normal approximation to model rate time-averages along each branch (branchwise averages, $\bar{\sigma^2}$) based on two observations. First, as the rate variance parameter ($\sigma^2_{\sigma^2}$) approaches 0, rates (σ^2) will converge to following a simple exponential function with respect to time, $\sigma^2 = \sigma_0^2 \exp[\mu_{\sigma^2} t]$, where σ_0^2 is the starting rate, μ_{σ^2} is the trend, and t is time. In this case, the branchwise averages can be derived through integration and are equivalent to the time-averaged rates expected under a conventional "early/late burst" (EB/LB) model (Blomberg et al., 2003). Second, over short amounts of time and/or with low rate variance, the arithmetic and geometric time-averages of a GBM process approach one another. The geometric time-average of a GBM process is simply the exponentiated arithmetic time-average of the GBM process on the natural log scale, which has a straight-forward and tractable log-normal distribution (Devreese et al., 2010). Thus, assuming that branch lengths in a phylogeny are typically short and rate variance is relatively low, we can approximate the distribution of the natural log of branchwise averages by adding multivariate normal "noise", γ, to the natural log of branchwise averages expected under a conventional EB/LB model, β. In other words:

$$\ln(\overline{\sigma^2}) \approx \beta + \gamma \tag{1}$$

$$\beta = \ln(\sigma_0^2) + \begin{cases} 0 & \text{if } \mu_{\sigma^2} = 0 \\ \ln(|\exp[\mu_{\sigma^2}\tau_2] - \exp[\mu_{\sigma^2}\tau_1]|) - \ln(|\mu_{\sigma^2}|) - \ln(t) & \text{if } \mu_{\sigma^2} \neq 0 \end{cases} \tag{2}$$

$$\gamma \sim MVN(0, \sigma^2_{\sigma^2} D) \tag{3}$$

as in the main text. Here, t is a vector of branch lengths, τ_1 and τ_2 are vectors of the start and end times of each branch (i.e., $\tau_2 - \tau_1 = t$), and D is the variance-covariance matrix of branchwise averages for a value evolving under an untrended BM process on a phylogeny. Let $\bar{x}$ and t be vectors of time-averaged trait values and edge lengths, respectively, for three edges: two sister edges, i and j, with ancestral edge, k. If traits evolve under an untrended BM process and the ancestral trait value of k is fixed, the variances of $\bar{x_i}$ and $\bar{x_j}$ are $t_i/3 + t_k$ and $t_j/3 + t_k$, respectively. The covariance between $\bar{x_i}$ and $\bar{x_j}$ is simply t_k, and the covariances between either $\bar{x_i}$ or $\bar{x_j}$ and $\bar{x_k}$ is $t_k/2$ (Devreese et al., 2010). From this, we can derive an expression for the variance-covariance matrix of branchwise averages given an arbitrary phylogeny, as shown in the main text:

$$D_{i,j} = \sum_{k \in \text{anc}(i,j)} t_k - \begin{cases} 2t_i/3 & \text{if } i = j \\ t_i/2 & \text{if } i \in \text{anc}(j,j) \\ t_j/2 & \text{if } j \in \text{anc}(i,i) \\ 0 & \text{if } i \neq j, i \notin \text{anc}(j,j), j \notin \text{anc}(i,i) \end{cases} \tag{4}$$

While this multivariate log-normal approximation is rough, we demonstrate here that it is largely sufficient for our purposes. Notably, we are not the first to approximate GBM time-averages using log-normal distributions in the context of comparative phylogenetics (Welch and Waxman, 2008). There are two other tractable strategies for approximating these distributions given in the comparative phylogenetics literature. Both of these strategies use the fact that values at the nodes of a phylogeny evolving under a GBM process follow an exact multivariate log-normal distribution, and instead focus on estimating nodewise values. Branchwise averages are then approximated by either averaging ancestral and descendant nodewise values for each edge (e.g., Thorne et al., 1998) or via the maximum likelihood estimate of branchwise averages given the ancestral and descendant nodewise values (e.g., Lartillot and Poujol, 2011; Revell, 2021). We term these strategies "endpoint averaging" and "endpoint integration", respectively. We prefer the log-normal approximation due to its convenient formulation and direct focus on estimating branchwise, rather than nodewise, quantities. In the spirit of thoroughness,

however, we conducted three simulation experiments to investigate the relative performance of these different approximation strategies.

We first conducted a simple experiment where we simulated 100,000 GBM time-averages on the natural log scale under each approximation strategy. We also estimated a "true" branchwise average distribution for comparison by simulating 100,000 fine-grained GBM sample paths (1,000 time points) and taking the natural log of each sample path's average. We repeated these simulations for each combination of trend (μ_{σ^2}) and rate variance ($\sigma^2_{\sigma^2}$) parameter values used in the main text's simulation study (Fig. S3). All simulations were standardized to occur over a time interval of 1, just as each phylogeny in our simulation study was rescaled to have a total height of 1. The results below thus represent how "off" each approximation would be for a single branch spanning the entire height of a phylogeny in our simuation study. The log-normal approximation notably lacks a right skew characteristic of the true distribution and other approximations. The log-normal approximation also appears to overestimate the variance of branchwise averages when trends are decreasing and underestimates variance when trends are increasing, particularly with high rate variance. On the other hand, the endpoint average approximation exhibits notable upward bias and consistently underestimates branchwise average variance. Additionally, this approximation fails to converge to the correct branchwise average when rate variance is 0. Lastly, the endpoint integration approximation exhibits no notable bias but underestimates branchwise average variance in the case of no or decreasing trends. The accuracy of branchwise average variance under the log-normal approximation might be improved by adapting the Fenton-Wilkinson approximation of log-normal sums for GBM processes (Safak and Safak, 2002), but we did not explore this here.

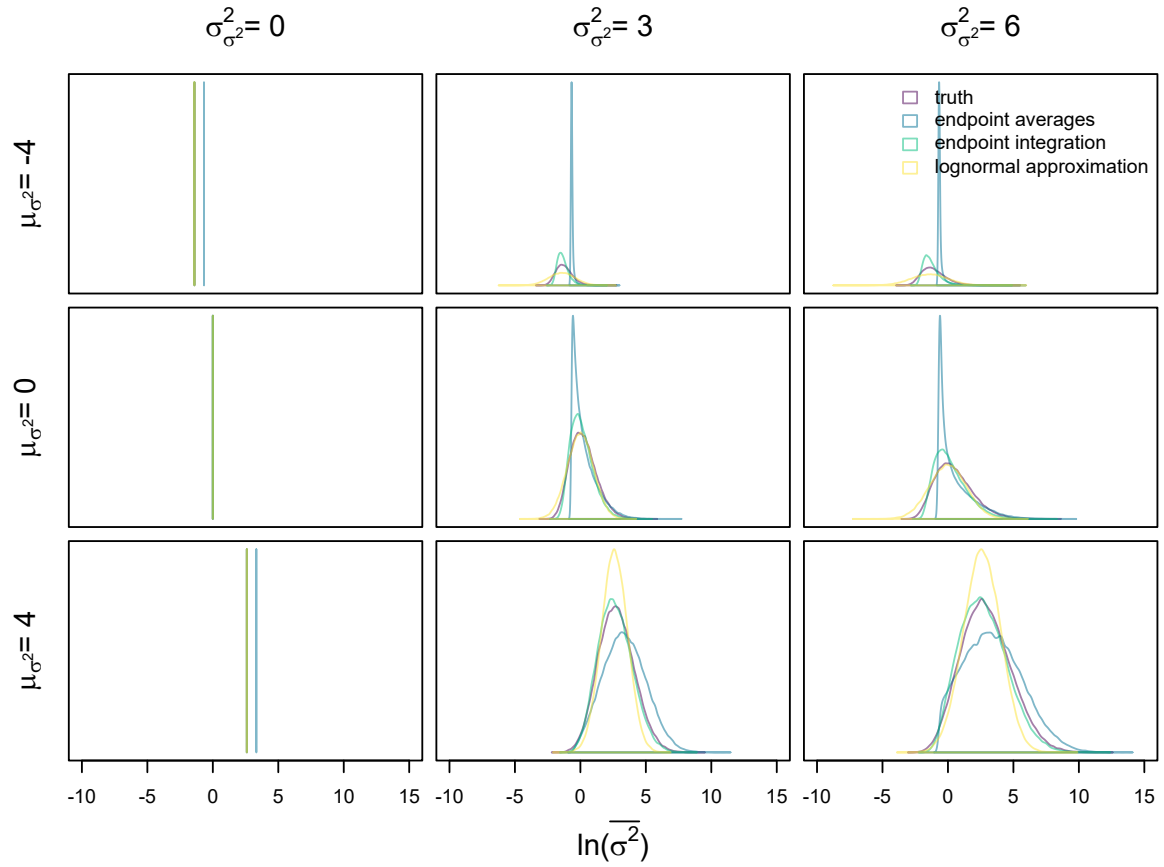

Figure S3. Distributions of simulated branchwise averages under different approximation strategies and the true distribution given parameter combinations used in the main text's simulation study. All simulations were run on single branches of length 1.

The above results help give a sense of where each approximation breaks down in parameter space, yet poorly represent the practical behavior of each approximation. In the context of our model, these approximations take place on individual branches of a phylogeny, which typically span relatively short intervals of time. For our next simulation experiment, we scaled up to simulating sets of branchwise averages on entire phylogenies. For each parameter combination (excluding combinations where rate variance is 0), we repeated the same simulations on 100 pure birth phylogenies with either 50, 100, or 200 species (generated using the R package *phytools*; Revell, 2012) standardized to a height of 1. For each phylogeny, we simulated 1,000 sets of branchwise averages under each approximation strategy, as well as fine-grained GBM sample paths (1,000 time points

across entire phylogeny's height) representing the true distribution. Because these samples have a high number of dimensions (one for each branch in a phylogeny), we visualized how well these multivariate distributions match one another using summary statistics. Specifically, for each tree, we recorded the correlation coefficients between the means/(co)variances of branchwise averages simulated under each approximation strategy and the true distribution (Figs. S4-9). To have a null expectation for these correlation coefficients, we also simulated a second true distribution and estimated correlation coefficients for means/(co)variances between replicate true distributions.

Overall, the results indicate that all approximations do a fairly good job at recapitulating the means and (co)variances expected under the true distribution. The log-normal approximation notably exhibits uncorrelated means in the case of no trend, in contrast to other approximations. This is due to the log-normal approximation lacking the right skew of the true distribution and other approximations (Fig. 3), which naturally inflates the means of branchwise average distributions along long branches. In the case of any trend, the endpoint average approximation exhibits somewhat less strong correlations between branchwise average means compared to other approximations. When rate variance is high, the log-normal approximation exhibits performance intermediate between the endpoint average approximation and endpoint integration approximation/null distribution. However, even the worst performing simulations nearly always exhibit strong correlations in branchwise average means above 0.98. In contrast to means, correlations for branchwise average (co)variances consistently varied between about 0.98-0.99 regardless of simulation parameters or approximation strategy, closely matching the null distribution.

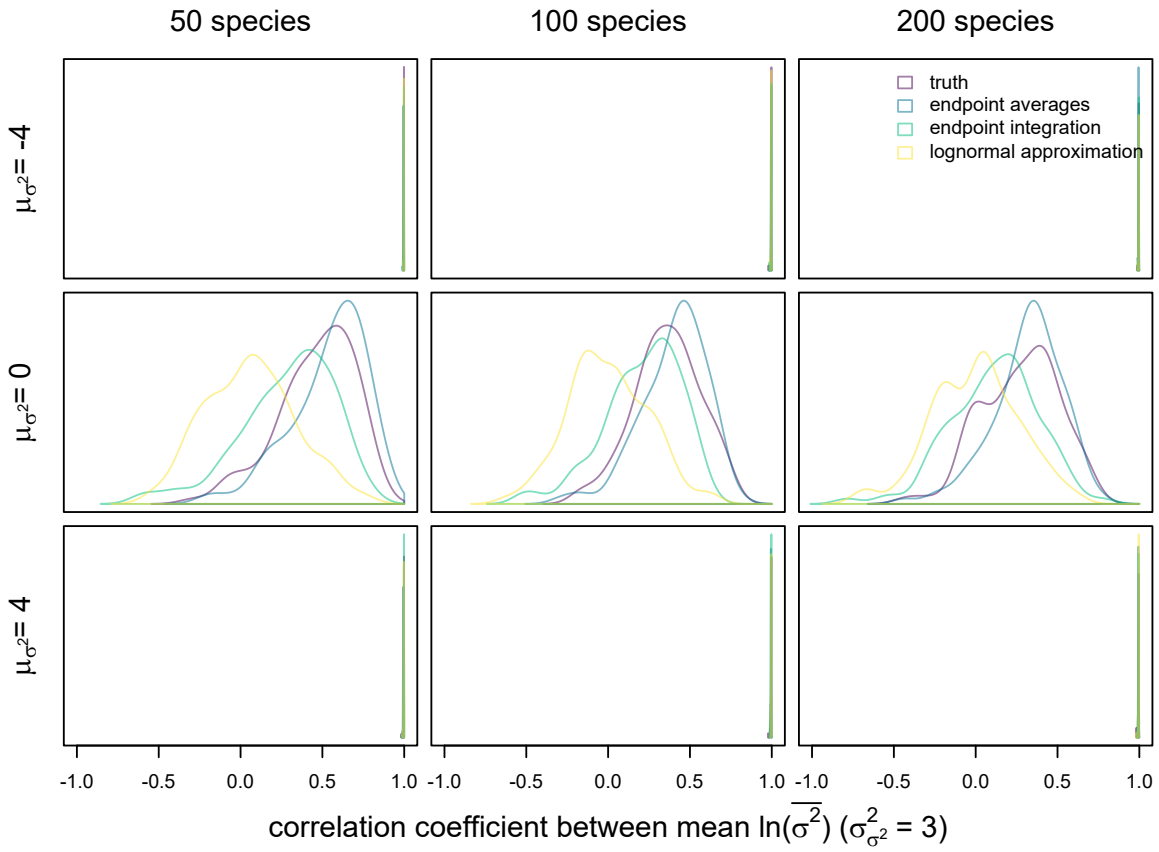

Figure S4. Distributions of correlation coefficients between mean simulated branchwise averages under different approximation strategies and the true distribution with rate variance ($\sigma^2_{\sigma^2}$) set to 3. All simulations were run on pure-birth phylogenies of height 1.

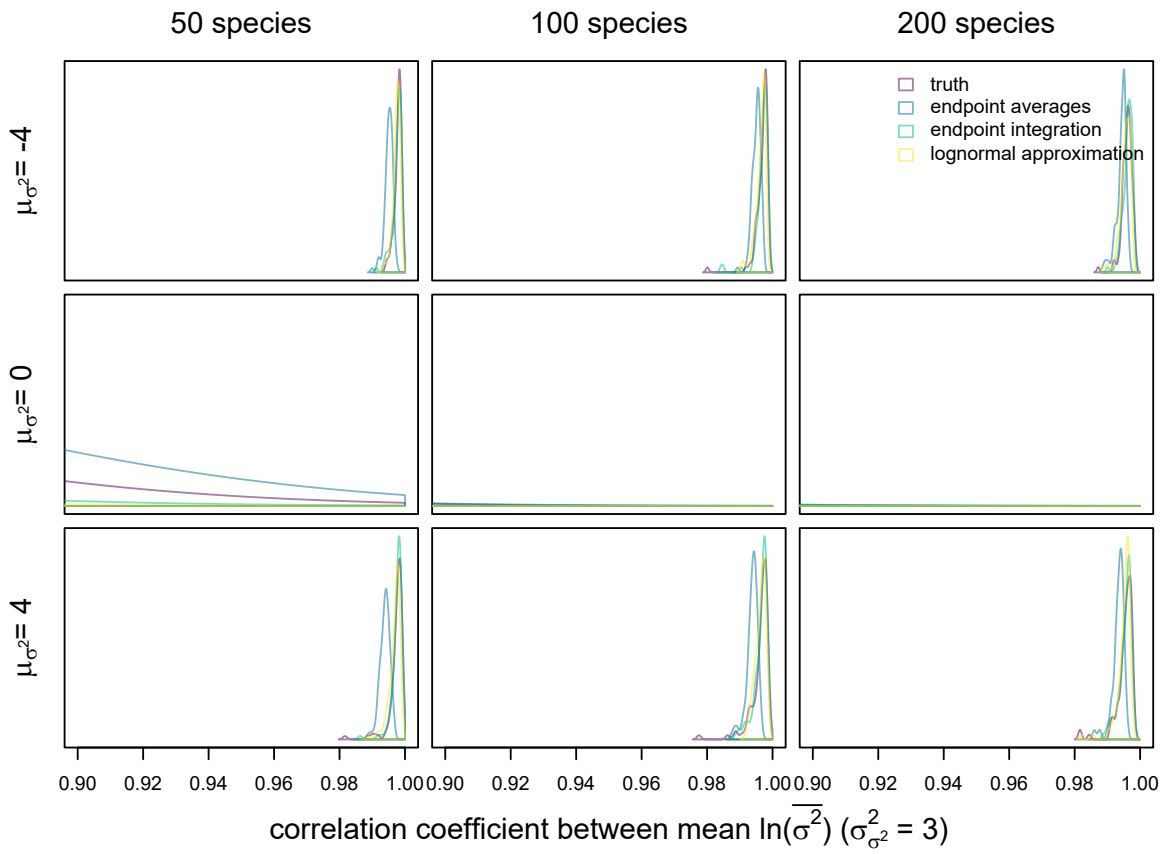

Figure S5. Distributions of correlation coefficients between mean simulated branchwise averages under different approximation strategies and the true distribution with rate variance ($\sigma^2_{\sigma^2}$) set to 3. All simulations were run on pure-birth phylogenies of height 1. Plots are zoomed in on distributions close to 1.

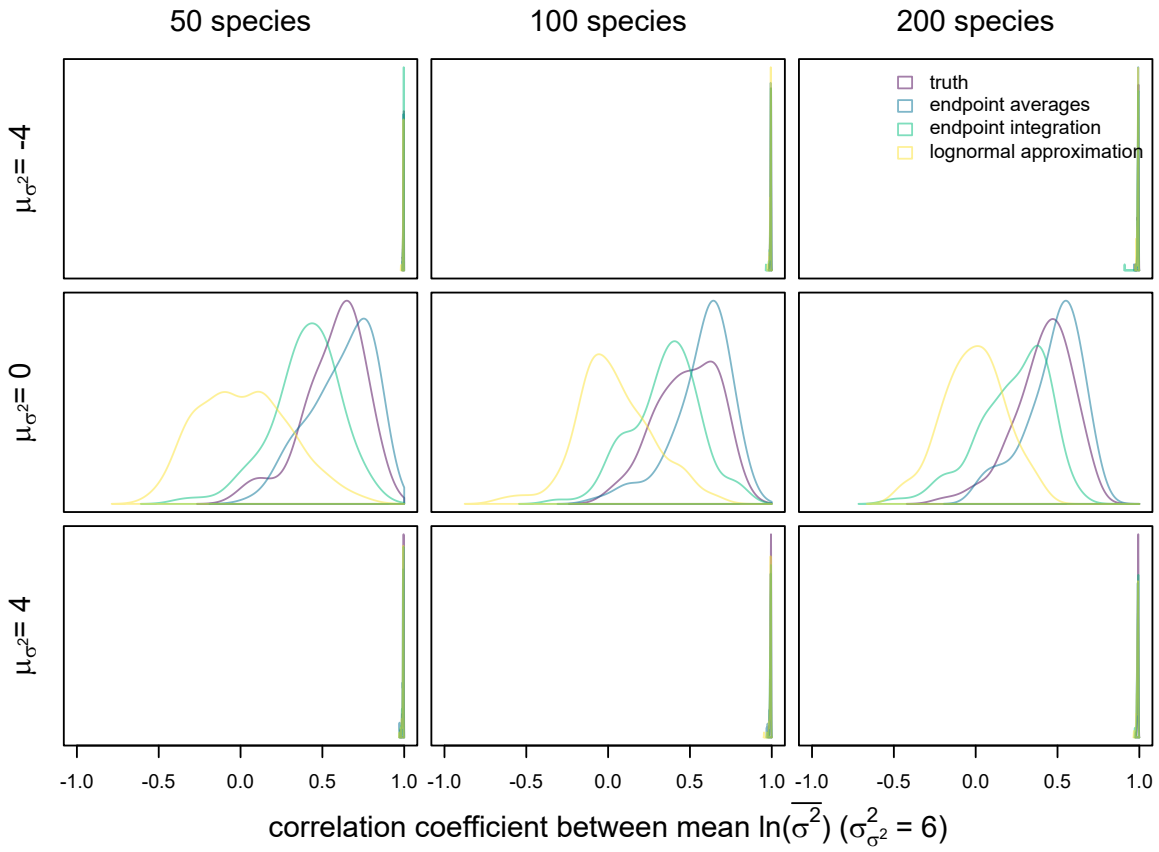

Figure S6. Distributions of correlation coefficients between mean simulated branchwise averages under different approximation strategies and the true distribution with rate variance ($\sigma^2_{\sigma^2}$) set to 6. All simulations were run on pure-birth phylogenies of height 1.

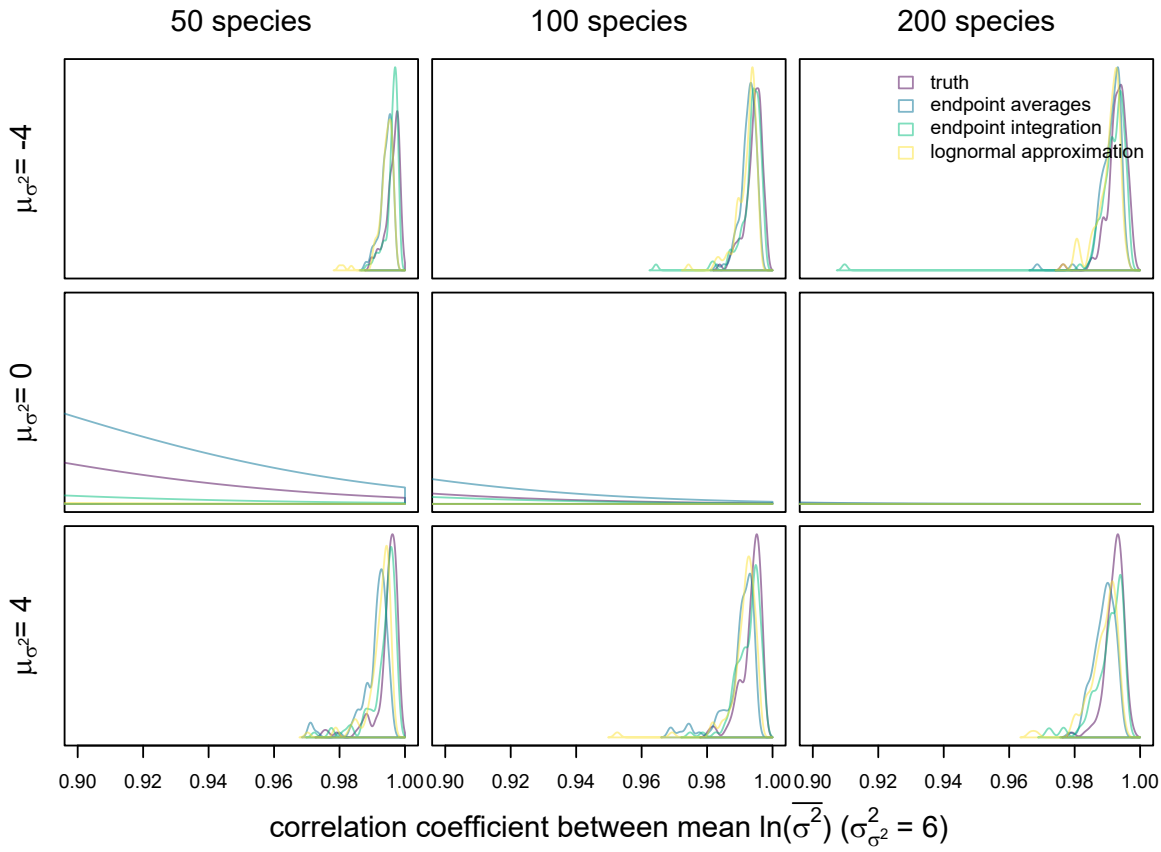

Figure S7. Distributions of correlation coefficients between mean simulated branchwise averages under different approximation strategies and the true distribution with rate variance ($\sigma^2_{\sigma^2}$) set to 6. All simulations were run on pure-birth phylogenies of height 1. Plots are zoomed in on distributions close to 1.

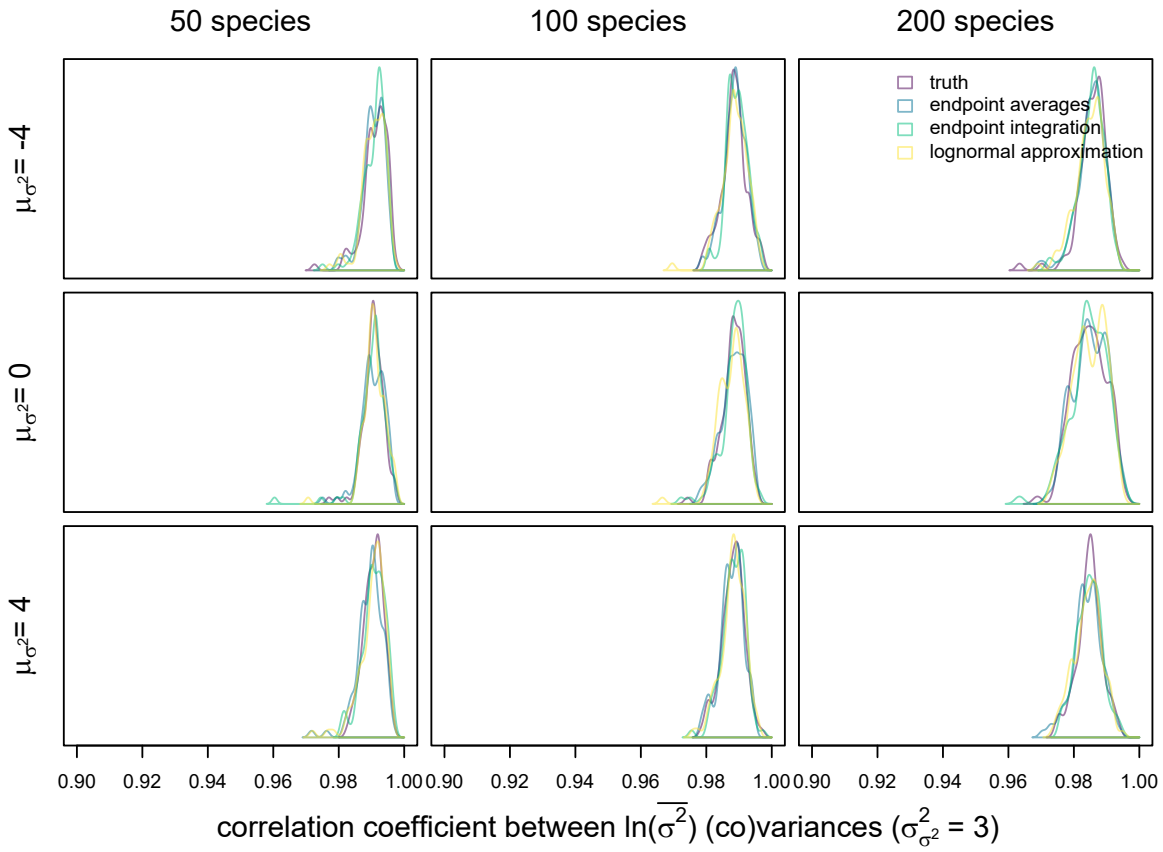

Figure S8. Distributions of correlation coefficients between simulated branchwise average (co)variances under different approximation strategies and the true distribution with rate variance ($\sigma^2_{\sigma^2}$) set to 3. All simulations were run on pure-birth phylogenies of height 1.

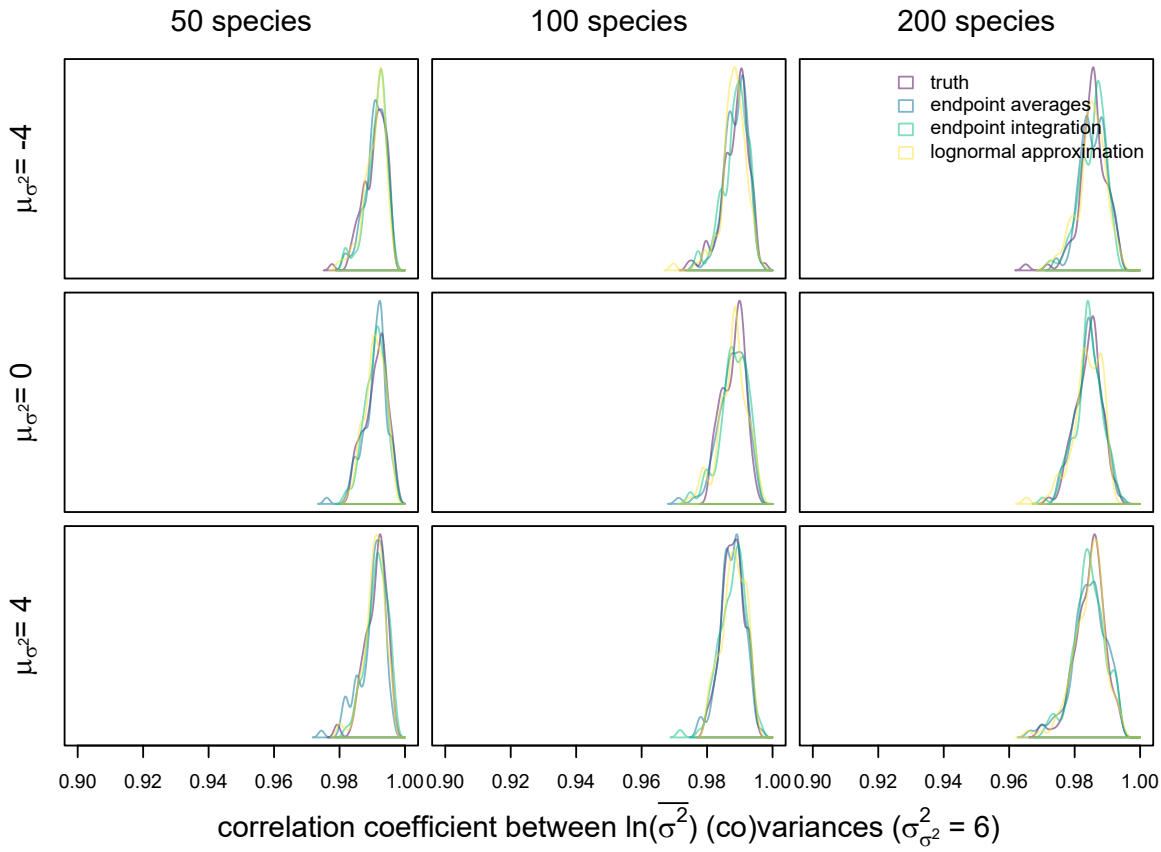

Figure S9. Distributions of correlation coefficients between simulated branchwise average (co)variances under different approximation strategies and the true distribution with rate variance ($\sigma^2_{\sigma^2}$) set to 6. All simulations were run on pure-birth phylogenies of height 1.

Because GBM time-averages are non-normally distributed, we also sought a non-parametric method of comparing samples from the approximations and true distributions. For this, we attempted to use the R package *FNN* (Beygelzimer et al., 2019) to estimate Kullback-Leibler (KL) divergence from each approximation to the true distribution. However, this estimator exhibited severe numerical issues, like negative KL divergence estimates. Thus, we instead implemented a crude K nearest neighbor probability density estimator (Zhao and Lai, 2021). For each tree in the simulation experiment above, we used this estimator to calculate local probability densities under each approximation and the true distribution around samples from a replicate true distribution. We then calculated log ratios of the true densities to densities under each

approximation and averaged the distances between these log ratios and 0 (i.e., equal densities). These averaged distances give a rough sense of how well the probability density of each approximation matches that of the true distribution, with increased sampling in higher-density regions of the true distribution (Figs. S10-11). Overall, the average log density ratio distances under each approximation matches the null distribution well. The endpoint average and log-normal approximations exhibit marginally elevated distances in the case of non-zero trends and decreasing trends, respectively, likely due to these approximations' under/overestimation of branchwise average variance in certain regions of parameter space (Fig. S3).

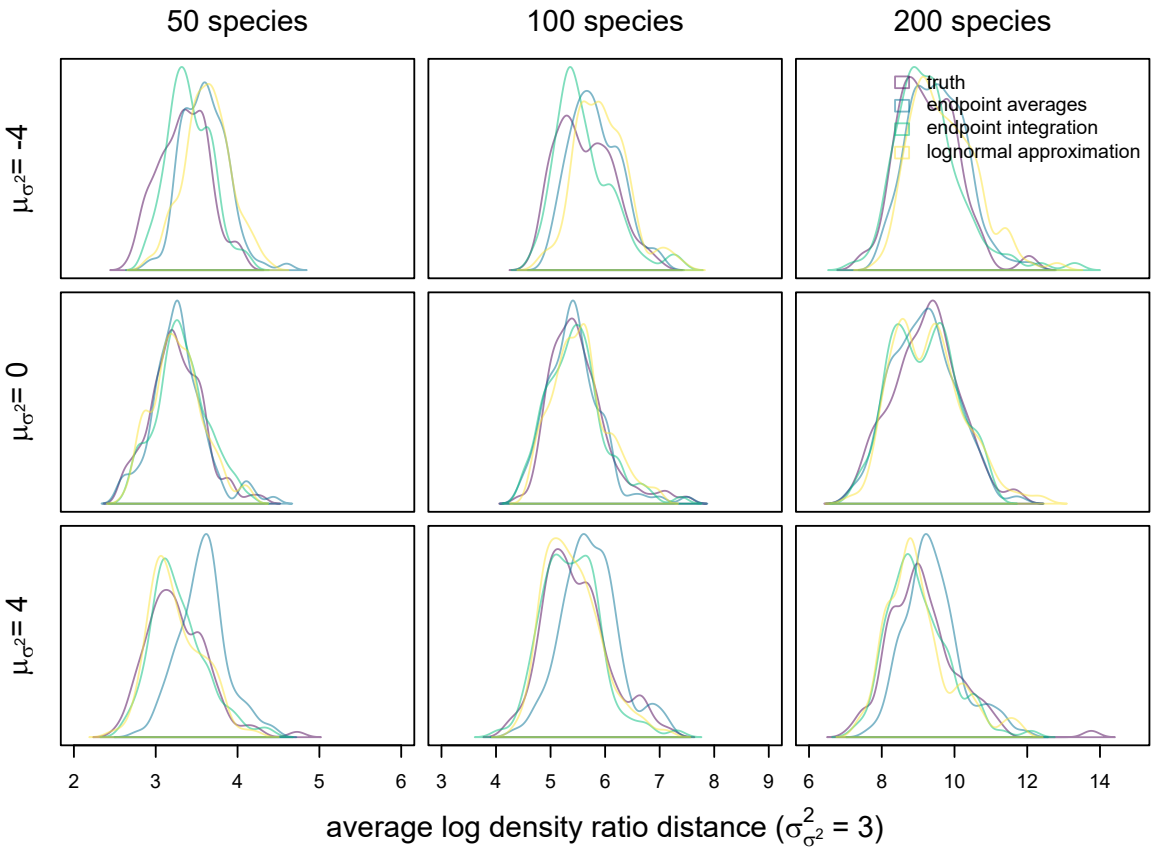

Figure S10. Distributions of average log density ratio distances between simulated branchwise average distributions under different approximation strategies and the true distribution with rate variance ($\sigma^2_{\sigma^2}$) set to 3. Probability densities were estimated via K nearest neighbors. All simulations were run on pure-birth phylogenies of height 3.

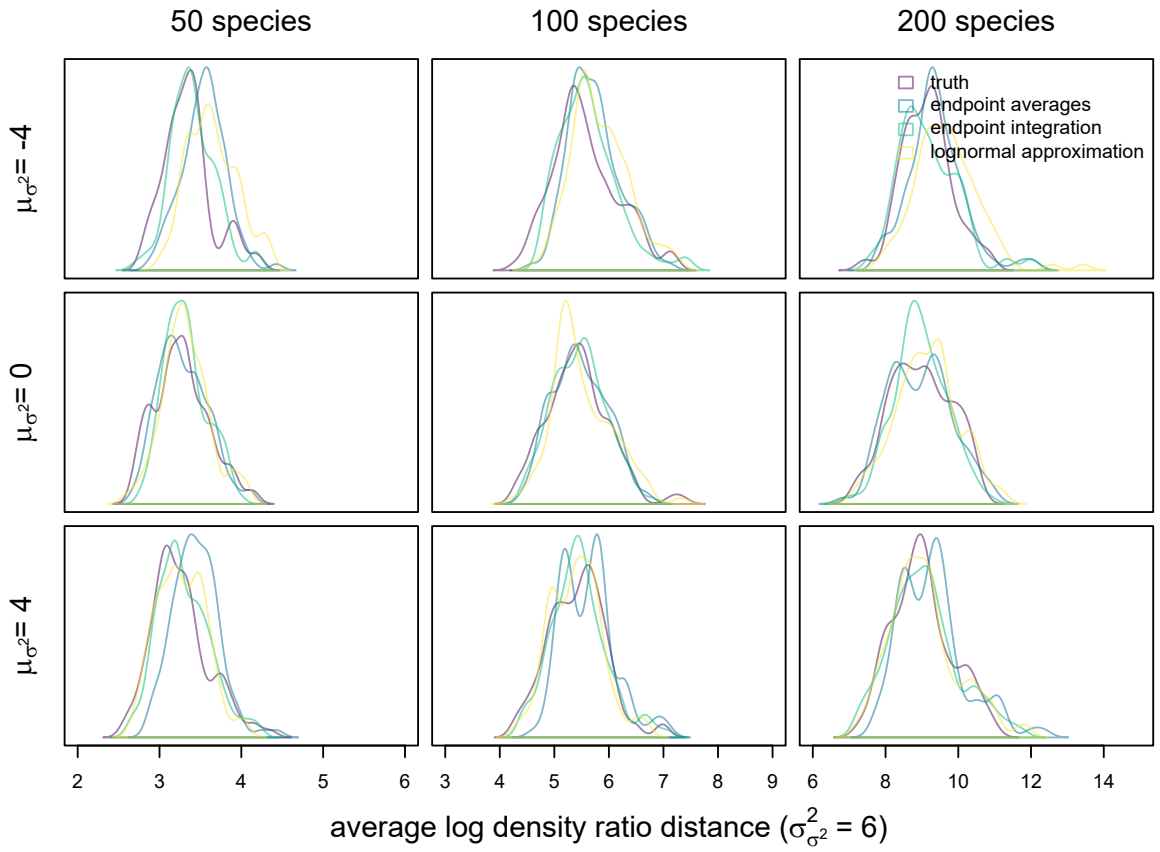

Figure S11. Distributions of average log density ratio distances between simulated branchwise average distributions under different approximation strategies and the true distribution with rate variance ($\sigma^2_{\sigma^2}$) set to 6. Probability densities were estimated via K nearest neighbors. All simulations were run on pure-birth phylogenies of height 1.

Lastly, we redid our entire simulation study with trait evolution rates simulated as evolving under a fine-grained GBM process (∼500 time points across entire phylogeny's height). We present all figures and tables for this simulation study below (Figs. S12-16; Tables S2-4). In general, the results qualitatively match those of the simulation study presented in the main text, and we feel confident that the log-normal approximation of branchwise averages is sufficient for our model. While there is some discrepancy in the statistical power of trend detection compared to results in the main text, it is unlikely such discrepancies result from systematic bias. Notably, statistical power for trend detection even under conventional EB/LB models in this simulation study also differs from the main text results, suggesting that any discrepancies are attributable to variation in the

simulated data.

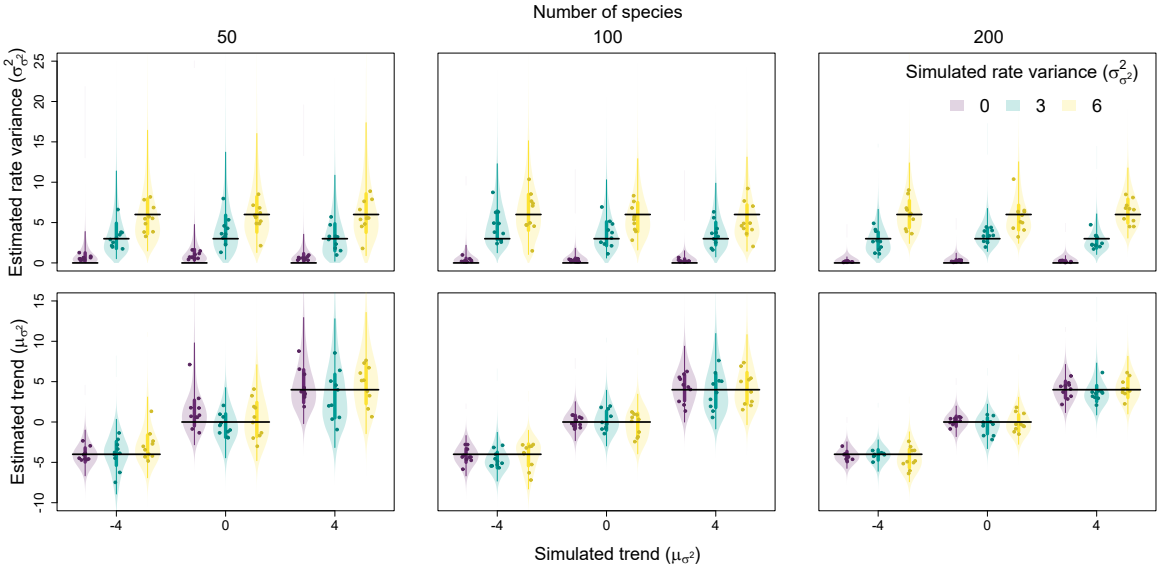

Figure S12. Relationship between simulated and estimated rate variance ($\sigma^2_{\sigma^2}$) and trend (μ_{σ^2}) parameters. Each point is the posterior median from a single fit, while the violins are combined posterior distributions from all fits for a given trait evolution scenario. Vertical lines represent the 50% (thicker lines) and 95% equal-tailed intervals (thinner lines) of these combined posteriors, while horizontal lines represent positions of true, simulated values.

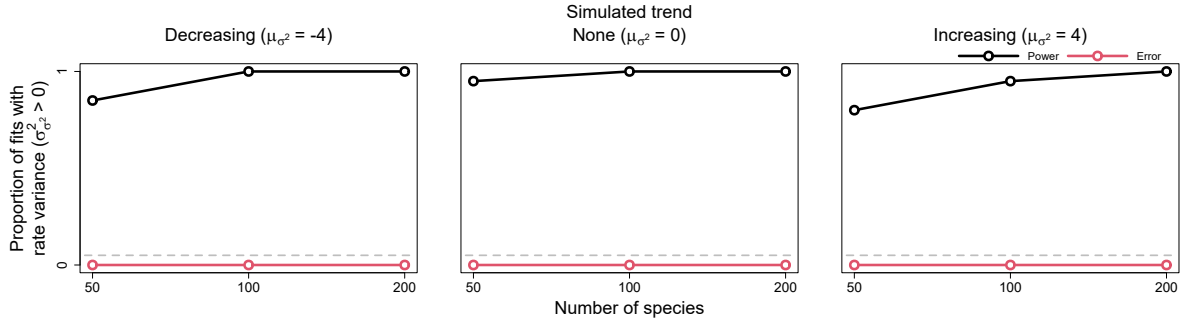

Figure S13. Power and error rates for the rate variance parameter ($\sigma^2_{\sigma^2}$). Lines depict changes in the proportion of model fits that correctly showed evidence for rate variance significantly greater than 0 (i.e., power, in black) and incorrectly showed evidence (i.e., error, in red) as a function of tree size.

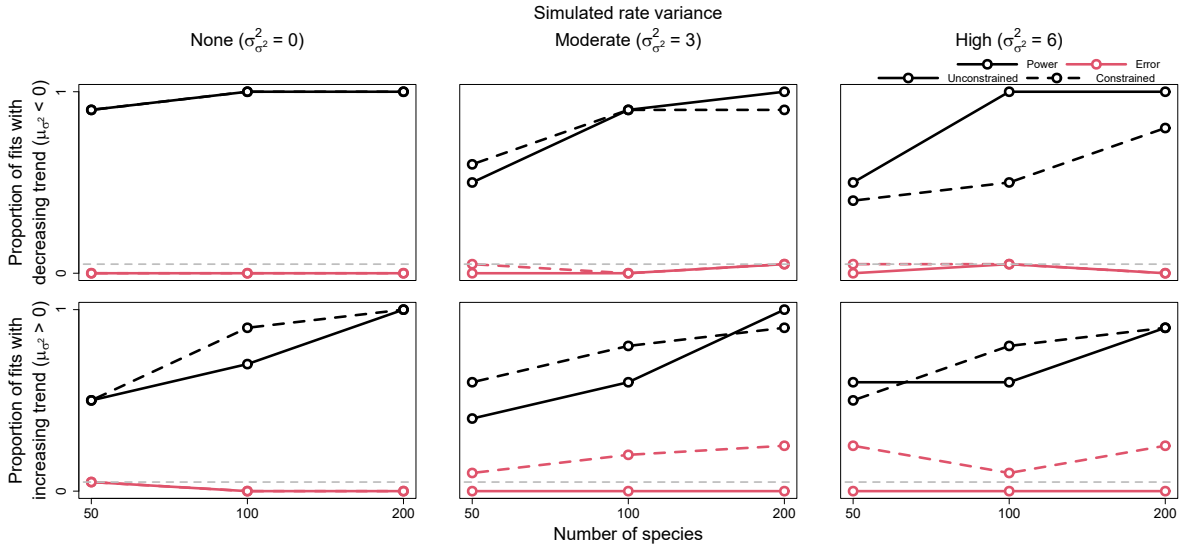

Figure S14. Power and error rates for the trend parameter (μ_{σ^2}). Lines depict changes in the proportion of model fits that correctly showed evidence for trends significantly less and greater than 0 (i.e., power, in black) and incorrectly showed evidence (i.e., error, in light red) as a function of tree size. Results are shown for both models allowed to freely estimate rate variance ($\sigma^2_{\sigma^2}$) (i.e., unconstrained models, solid lines) and models with rate variance constrained to 0 (i.e., constrained models, dashed lines). The latter models are identical to conventional early/late burst models.

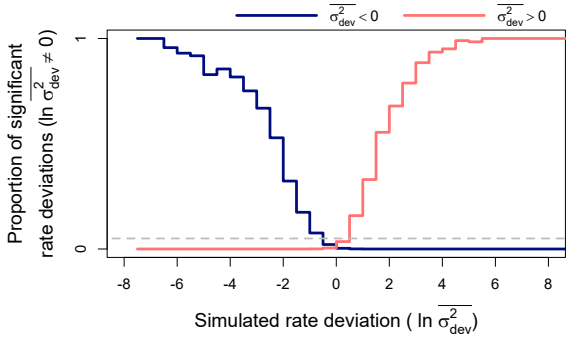

Figure S15. Power and error rates for branchwise rate parameters (ln $\overline{\sigma^2}$). Lines depict changes in proportions of branchwise rates considered anomalously slow (in dark blue) or fast (in light red) as a function of simulated rate deviations (ln $\overline{\sigma^2_{dev}}$). These results combine all fits to simulated data that detected rate variance ($\sigma^2_{\sigma^2}$) significantly greater than 0. The proportions are equivalent to power when the detected rate deviation is of the same sign as the true, simulated deviation (left of 0 for anomalously slow rates in dark blue and right for anomalously fast rates in light red), and to error rate when the detected and true rate deviations are of opposite signs. Here, significant rate deviations for simulated rate deviations that are exactly 0 are considered errors regardless of sign.

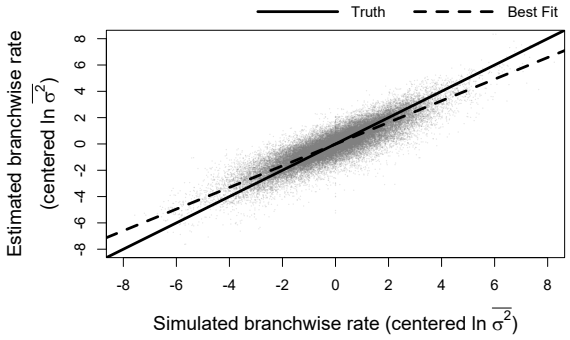

Figure S16. Relationship between simulated and estimated branchwise rate parameters (ln $\overline{\sigma^2}$). For each simulation and posterior sample, branchwise rates were first centered by subtracting their mean. We estimated centered branchwise rates by taking the median of the centered posterior samples. The solid line represents the position of the true centered branchwise rates, while the shallower, dashed line represents the observed line of best fit for these data.

Table S2. *Median absolute errors of rate variance, trend, and branchwise rate posteriors (i.e., median absolute difference between posterior samples and their true, simulated values, a measure of posterior distribution accuracy), averaged across replicates for each simulated trait evolution scenario and tree size.* $\sigma^2_{\sigma^2}$ *and* μ_{σ^2} *indicate the true, simulated values of rate variance and trend parameters, respectively.*

		rate variance			trend			branchwise rates		
$\sigma^2_{\sigma^2}$ =		0	3	6	0	3	6	0	3	6
					50 species					
μ_{σ^2} =	-4	0.61	1.58	2.26	0.94	1.68	1.78	0.42	0.80	0.96
	0	0.89	1.89	2.23	2.09	1.56	2.22	0.62	0.82	1.04
	4	0.58	1.68	2.41	2.15	2.98	2.62	0.63	0.92	0.98
					100 species					
μ_{σ^2} =	-4	0.31	2.11	2.37	0.91	1.22	1.43	0.32	0.77	0.86
	0	0.31	1.59	1.95	0.81	1.26	1.47	0.32	0.82	0.93
	4	0.26	1.49	2.21	1.67	2.16	2.02	0.41	0.85	0.94
					200 species					
μ_{σ^2} =	-4	0.14	1.23	1.79	0.62	0.66	1.29	0.23	0.68	0.80
	0	0.21	0.93	1.82	0.65	1.09	1.10	0.24	0.72	0.84
	4	0.18	0.98	1.50	1.09	1.17	1.27	0.28	0.73	0.84

Table S3. *Breadths of rate variance, trend, and branchwise rate posteriors (i.e., the difference between the 97.5% and 2.5% quantiles of posterior samples, a measure of posterior distribution precision), averaged across replicates for each simulated trait evolution scenario and tree size.* $\sigma^2_{\sigma^2}$ *and* μ_{σ^2} *indicate the true, simulated values of rate variance and trend parameters, respectively.*

		rate variance			trend			branchwise rates		
$\sigma^2_{\sigma^2} =$		0	3	6	0	3	6	0	3	6
					50 species					
$\mu_{\sigma^2} =$	-4	3.67	9.11	12.98	4.66	6.02	6.81	2.28	3.24	3.65
	0	4.38	10.67	12.60	7.28	7.09	8.00	2.60	3.41	3.89
	4	3.35	9.00	13.88	10.34	10.95	12.09	2.81	3.50	4.10
					100 species					
$\mu_{\sigma^2} =$	-4	1.77	7.96	9.58	3.53	4.56	4.72	1.71	3.22	3.46
	0	1.64	6.72	9.15	4.04	5.09	5.67	1.76	3.12	3.42
	4	1.36	6.77	8.13	6.74	8.08	7.86	1.87	3.31	3.55
					200 species					
$\mu_{\sigma^2} =$	-4	0.71	3.97	7.20	2.64	3.58	4.06	1.24	2.50	3.12
	0	1.04	4.26	6.52	3.34	3.98	4.15	1.36	2.77	3.25
	4	0.79	3.62	6.89	4.53	4.88	5.69	1.39	2.70	3.37

Table S4. *Coverage of rate variance, trend, and branchwise rate posteriors (i.e., proportion of times the true, simulated value is greater than the 2.5% posterior distribution quantile and less than the 97.5% quantile) for each simulated trait evolution scenario and tree size.* $\sigma^2_{\sigma^2}$ *and* μ_{σ^2} *indicate the true, simulated values of rate variance and trend parameters, respectively.*

		rate variance			trend			branchwise rates		
$\sigma^2_{\sigma^2}$ =		0	3	6	0	3	6	0	3	6
					50 species					
μ_{σ^2} =	-4	—	1.00	1.00	1.00	0.80	0.90	1.00	0.95	0.94
	0	—	1.00	1.00	0.90	1.00	1.00	0.97	0.98	0.94
	4	—	1.00	1.00	1.00	0.80	1.00	0.95	0.94	0.96
					100 species					
μ_{σ^2} =	-4	—	0.70	0.90	0.90	1.00	0.90	1.00	0.96	0.96
	0	—	1.00	1.00	1.00	1.00	0.90	1.00	0.94	0.94
	4	—	1.00	0.90	0.90	0.90	1.00	0.99	0.95	0.93
					200 species					
μ_{σ^2} =	-4	—	0.90	1.00	1.00	1.00	0.90	0.99	0.93	0.95
	0	—	1.00	0.80	1.00	0.90	1.00	1.00	0.95	0.95
	4	—	1.00	1.00	0.90	1.00	1.00	0.99	0.93	0.96

Average Changes in Trait Evolution Rates

Conventional early/late burst (EB/LB) models of trait evolution assume that rates follow a homogeneous, exponential declines or increases with respect to time (Blomberg et al., 2003). The definition of EBs/LBs under such models is thus straight-forward–any given time slice in a clade's history is associated with a single trait evolution rate, and these rates can only decrease, increase or stay the same. On the other hand, allowing for rate heterogeneity independent of overall temporal trends means that any given time slice in a clade's history is associated with a *distribution* of trait evolution rates. Because of this, our new method allows for alternative definitions of EBs/LBs, depending on how one summarizes these distributions. In the current study, we mainly consider a definition based on whether the medians, or geometric means, of these distributions decrease or increase over time (change per unit time given by μ_{σ^2}, hereafter the "trend" parameter, as in the

main text). Alternatively, one could use a definition based on whether the average, or arithmetic means, of these distributions decrease or increase over time (change per unit time given by $\mu_{\sigma^2} + \sigma^2_{\sigma^2}/2$, hereafter the "average change" parameter, δ_{σ^2}).

We chose to focus on trend over average change estimation and define EBs/LBs based on the trend parameter for a few reasons. First, average change is a composite parameter of both the trend and rate variance parameters, posing some interpretational challenges. In general, it seems more intuitive to consider the magnitude of deterministic changes in trait evolution rates (the trend component) apart from the magnitude of stochastic changes (the rate variance component). Second, because rates evolve in an approximately log-normal manner under our model, medians are a natural, reliable way of summarizing their distributions, corresponding to the exponentiated average of rates on the natural log scale. In contrast, the right skew of log-normal distributions causes raw averages of trait evolution rates to be highly influenced by few, extreme outliers, particularly when rate variance is high. For this reason, our model can produce trait evolution scenarios whereby rates exhibit declines in the majority of lineages (directly related to changes in median rates) while increasing on average (Figs. S17-18). Lastly, many macroevolutionary biologists consider "accounting" for lineages/subclades exhibiting unusual trait evolution rates critical to elucidating and understanding changes in rates over time (Lloyd et al., 2012; Slater and Pennell, 2014; Benson et al., 2014; Hopkins and Smith, 2015; Wright, 2017; Puttick, 2018). This implies that many empiricists intuitively define EBs/LBs based on majority changes in rates rather than changes in average rates. Additionally, by log-transforming traits prior to analysis, many macroevolutionary biologists implicitly use GBM processes to model trait evolution, just as we use a (approximate) GBM process to model rate evolution here. In the context of trait evolution, the analogous trend parameter is widely considered by empiricists and method developers alike to determine whether a clade exhibits a directional "evolutionary trend" in traits, regardless the estimated variance parameter (Hunt, 2006; Raj Pant et al., 2014; Sookias

173 et al., 2012; Gill et al., 2017).

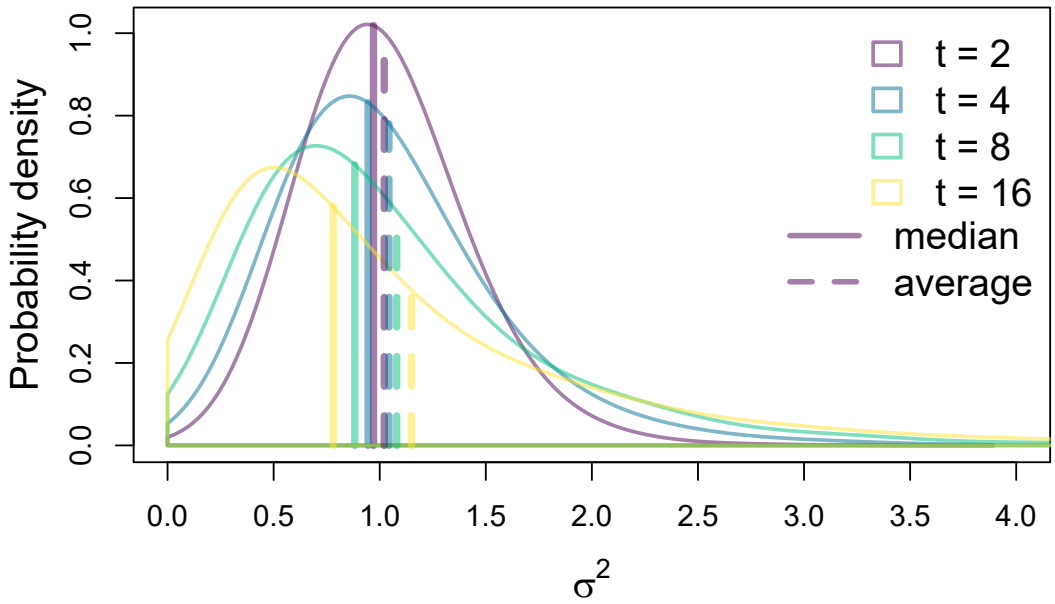

Figure S17. Distributions of 6,000 rates simulated as evolving under a GBM process with trend of -0.015 and rate variance of 0.05 at various time points, with starting rate of 1 at time $t = 0$. Parameter values were chosen to clearly illustrate how rates under our model may exhibit majority declines while increasing on average due to the skewed nature of rate change. Solid and dashed vertical lines represent the positions of median and average rate values, respectively, for each time point.

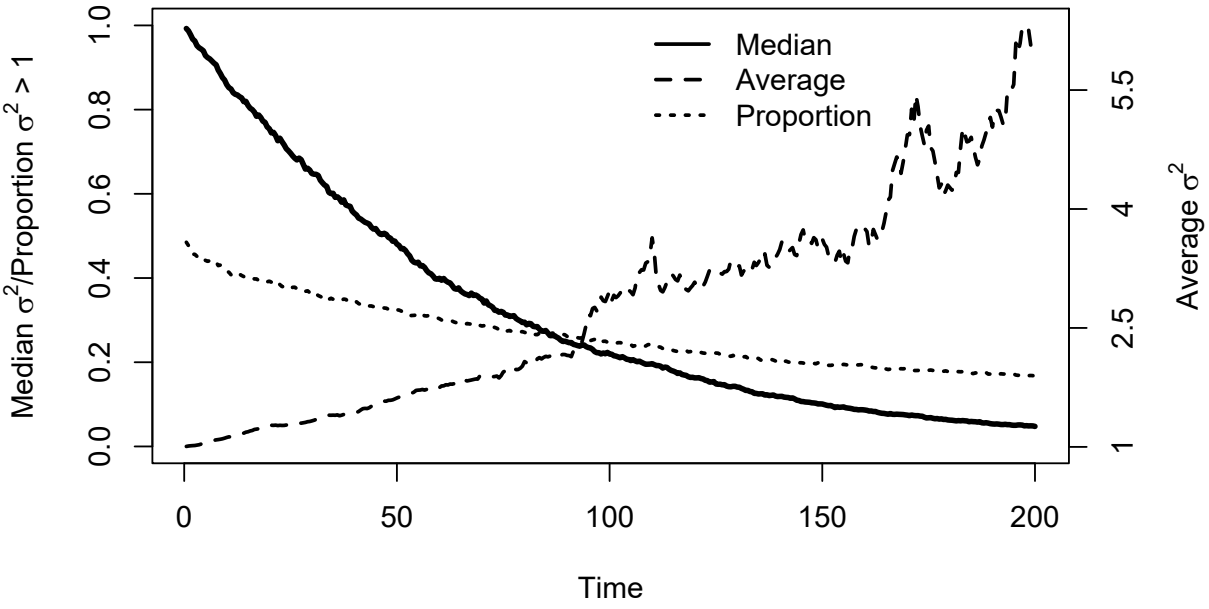

Figure S18. Changes over time in the median and average of 6,000 rates simulated as evolving under a GBM process with trend of -0.015 and rate variance of 0.05, with starting rate of 1 at time $t = 0$. Parameter values were chosen to clearly illustrate how rates under our model may exhibit majority declines while increasing on average due to the skewed nature of rate change. Solid and dashed lines depict changes in median and average rate values, respectively, while the dotted line depicts changes in the proportion of rates greater than the starting rate of 1.

Here, we briefly consider our new method's performance with respect to estimating and detecting average changes in trait evolution rates. Interestingly, our simulation study results revealed that, in the presence of time-independent rate heterogeneity, conventional EB/LB models (equivalent to our new models with rate variance constrained to 0) appear to estimate average change, rather than trend parameters, as defined under our model (Figs. S19-20). We are not aware of any previous research explicitly demonstrating this phenomenon. When comparing performance of constrained to unconstrained models with respect to detecting significant average change (i.e., 95% equal-tailed interval lies entirely below or above 0), we generally see only a modest reduction in error rates and greatly reduced power to detect negative average change under the full, unconstrained model (Fig. S21). Nonetheless, inference of the average change parameter seems substantially improved under unconstrained models (Tables S5-7). In the presence of time-independent rate

heterogeneity, constrained models tend to exhibit less accurate, overly-narrow posterior estimates of average change, particularly when the rate variance and trend parameters are high, resulting in low posterior coverage. This warrants caution in interpreting the results of conventional EB/LB models fitted to comparative data exhibiting substantial time-independent rate heterogeneity, and we recommend estimating rate variance even when one's only goal is to estimate changes in average trait evolution rates over time.

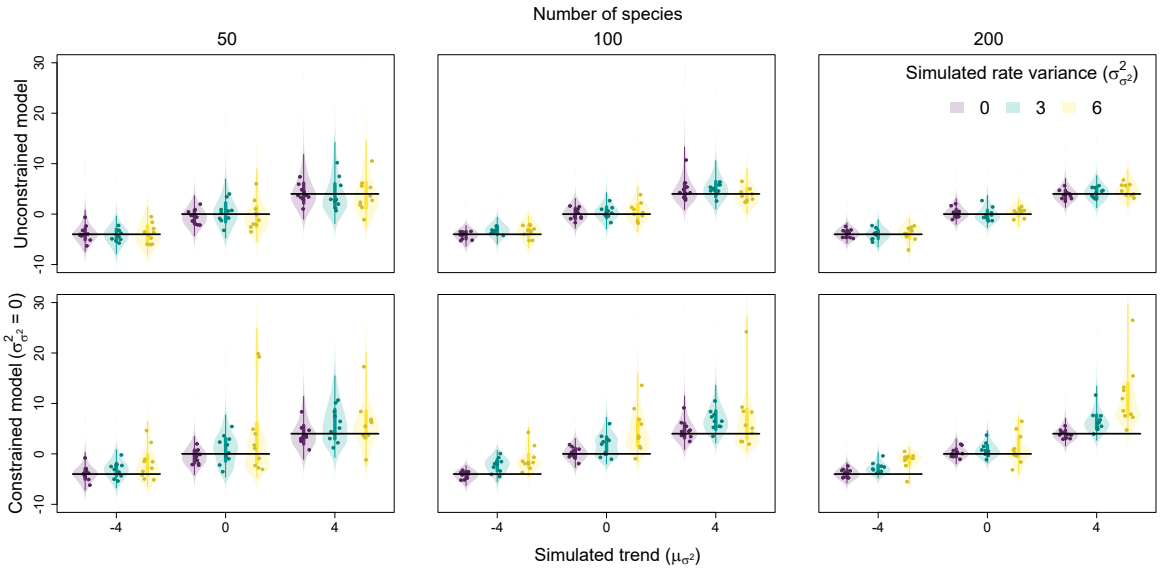

Figure S19. Relationship between simulated rate variance ($\sigma^2_{\sigma^2}$)/trend (μ_{σ^2}) and estimated trend parameters. Each point is the posterior median from a single fit, while the violins are combined posterior distributions from all fits for a given trait evolution scenario. Vertical lines represent the 50% (thicker lines) and 95% equal-tailed intervals (thinner lines) of these combined posteriors, while horizontal lines represent positions of true, simulated values. Results for models with estimated rate variance unconstrained and constrained to 0 are shown on top and bottom, respectively.

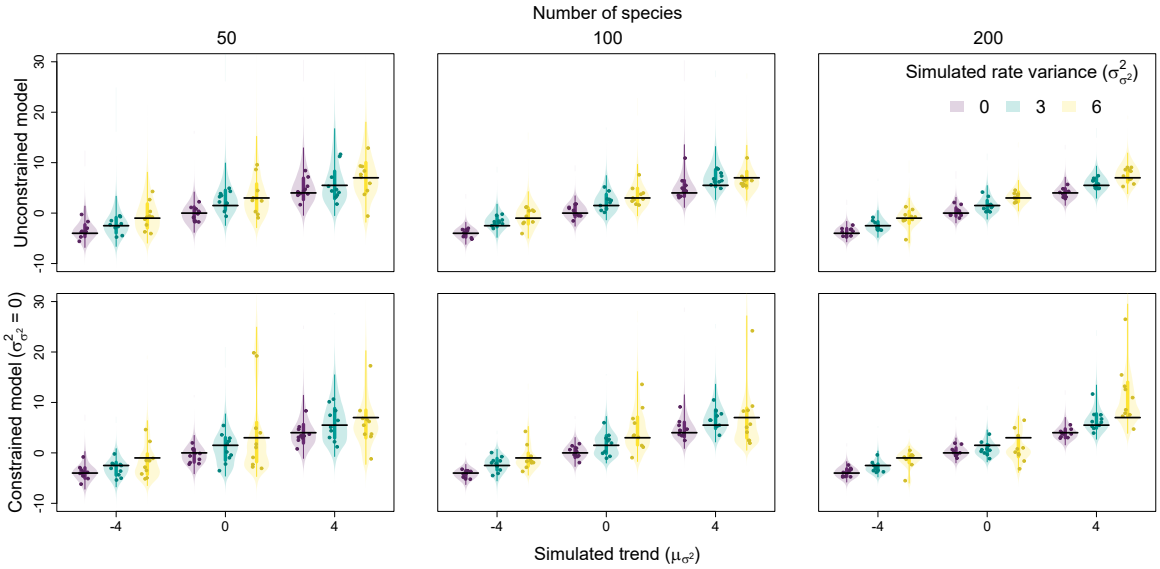

Figure S20. Relationship between simulated rate variance ($\sigma^2_{\sigma^2}$)/trend (μ_{σ^2}) and estimated average change (δ_{σ^2}) parameters. Each point is the posterior median from a single fit, while the violins are combined posterior distributions from all fits for a given trait evolution scenario. Vertical lines represent the 50% (thicker lines) and 95% equal-tailed intervals (thinner lines) of these combined posteriors, while horizontal lines represent positions of true, simulated values. Results for models with estimated rate variance unconstrained and constrained to 0 are shown on top and bottom, respectively.

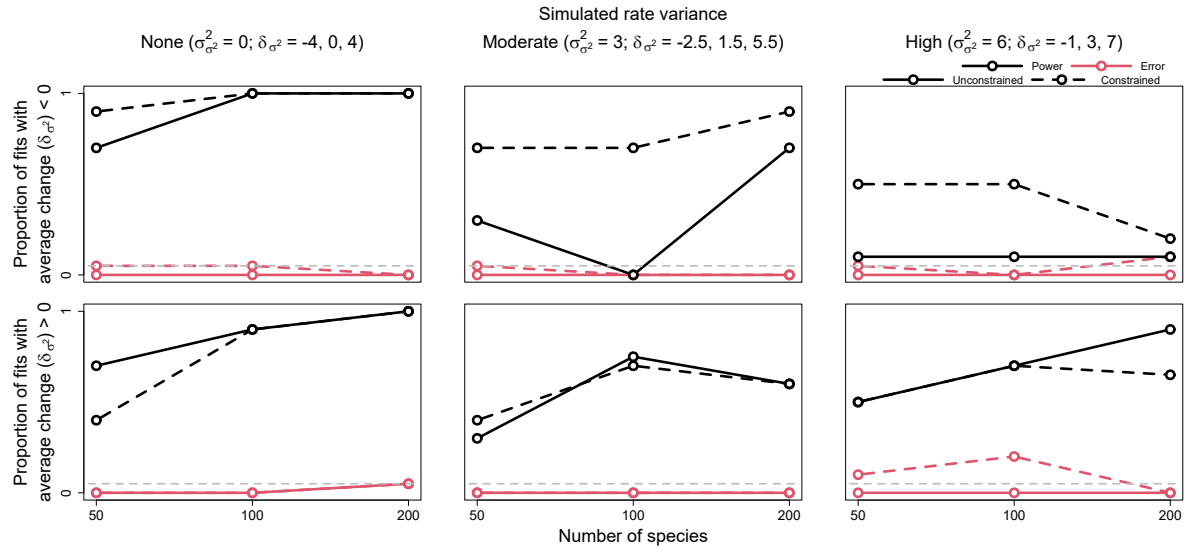

Figure S21. Power and error rates for the average parameter (δ_{σ^2}). Lines depict changes in the proportion of model fits that correctly showed evidence for average change significantly less and greater than 0 (i.e., power, in black) and incorrectly showed evidence (i.e., error, in light red) as a function of tree size. Results are shown for both models allowed to freely estimate rate variance ($\sigma^2_{\sigma^2}$) (i.e., unconstrained models, solid lines) and models with rate variance constrained to 0 (i.e., constrained models, dashed lines). The latter models are identical to conventional early/late burst models.

Table S5. *Median absolute errors of average change posteriors (i.e., median absolute difference between posterior samples and their true, simulated values, a measure of posterior distribution accuracy) under models with rate variance unconstrained and constrained to 0, averaged across replicates for each simulated trait evolution scenario and tree size.* $\sigma^2_{\sigma^2}$ *and* μ_{σ^2} *indicate the true, simulated values of rate variance and trend parameters, respectively.*

		unconstrained			constrained		
$\sigma^2_{\sigma^2} =$		0	3	6	0	3	6
		50 species					
$\mu_{\sigma^2} =$	-4	1.41	1.61	2.50	1.23	1.50	2.74
	0	1.43	2.10	3.08	1.45	2.45	6.08
	4	2.22	3.04	3.34	2.05	3.05	3.87
		100 species					
$\mu_{\sigma^2} =$	-4	0.78	1.27	1.70	0.74	1.28	1.78
	0	1.15	1.65	1.72	1.08	1.88	3.35
	4	1.92	1.98	1.85	1.70	2.08	4.39
		200 species					
$\mu_{\sigma^2} =$	-4	0.79	0.92	1.44	0.78	0.96	1.19
	0	0.92	1.21	1.01	0.90	1.35	3.05
	4	0.97	1.15	1.51	0.94	1.80	5.06

Table S6. *Breadths of average change posteriors (i.e., the difference between the 97.5% and 2.5% quantiles of posterior samples, a measure of posterior distribution precision) under models with rate variance unconstrained and constrained to 0, averaged across replicates for each simulated trait evolution scenario and tree size.* $\sigma^2_{\sigma^2}$ *and* μ_{σ^2} *indicate the true, simulated values of rate variance and trend parameters, respectively.*

		unconstrained			constrained		
$\sigma^2_{\sigma^2} =$		0	3	6	0	3	6
		50 species					
$\mu_{\sigma^2} =$	-4	5.56	7.95	10.27	4.50	4.65	5.63
	0	6.25	9.89	11.46	5.47	6.82	8.45
	4	11.04	11.69	12.45	9.48	10.81	10.60
		100 species					
$\mu_{\sigma^2} =$	-4	3.42	5.48	6.54	3.07	3.49	3.84
	0	4.46	6.27	7.44	3.97	4.58	6.60
	4	7.64	8.97	8.56	7.12	8.41	8.40
		200 species					
$\mu_{\sigma^2} =$	-4	2.82	4.14	5.07	2.70	2.87	3.26
	0	3.40	4.50	5.13	3.25	3.30	3.37
	4	4.51	5.45	6.29	4.38	5.78	8.73

Table S7. *Coverage of average change posteriors (i.e., proportion of times the true, simulated value is greater than the 2.5% posterior distribution quantile and less than the 97.5% quantile) under models with rate variance unconstrained and constrained to 0 for each simulated trait evolution scenario and tree size.* $\sigma^2_{\sigma^2}$ *and* μ_{σ^2} *indicate the true, simulated values of rate variance and trend parameters, respectively.*

		unconstrained			constrained		
$\sigma^2_{\sigma^2} =$		0	3	6	0	3	6
		50 species					
$\mu_{\sigma^2} =$	-4	0.90	1.00	0.90	0.80	0.60	0.50
	0	1.00	1.00	1.00	0.90	0.80	0.40
	4	1.00	1.00	0.90	1.00	1.00	0.80
		100 species					
$\mu_{\sigma^2} =$	-4	1.00	1.00	0.90	1.00	0.80	0.60
	0	1.00	0.90	0.90	0.90	0.60	0.60
	4	0.90	1.00	1.00	0.90	0.90	0.60
		200 species					
$\mu_{\sigma^2} =$	-4	1.00	1.00	0.90	1.00	0.90	0.80
	0	0.90	1.00	1.00	0.90	0.60	0.10
	4	1.00	1.00	1.00	1.00	0.90	0.50

Prior Sensitivity Study

To see how sensitive our method is to alternate prior specifications, we refit models to our smallest simulations (50 tips) while varying prior settings. We focus on the smallest simulations because the priors are more influential when there is less data. In addition to refitting models with default priors to each simulation (see *Priors* subsection of *Materials and Methods* section in main text), we also refit models with "tight" and "loose" prior settings, whereby the priors for rate variance ($\sigma^2_{\sigma^2}$), trend (μ_{σ^2}), and root rate (σ^2_0) parameters were made more or less informative, respectively. We did this by either reducing the prior scale parameter (i.e., standard deviation in the case of normal distributions) 5-fold for more precise, informative priors or increasing 3-fold for more relaxed, uninformative priors (i.e., prior scales of $1/T$ for rate variance, $2/T$ for trend, and

2 for root rate under the tight settings and $15/T$, $30/T$, and 30 under the loose settings, where T is the height of the phylogeny). Within each of these three prior settings (tight, default, or loose), we additionally shifted the location of the root rate prior by either -3, 0, or 3, yielding a total of 9 prior settings. These shifts correspond to ∼20-fold changes in the expected root rate.

Because this simulation study design requires many more model fits compared to the main text's simulation study (9 trait evolution scenarios with 10 replicates refit under 9 different prior settings, yielding 810 model fits), we only ran 2 Hamiltonian Monte Carlo chains consisting of 1,500 iterations for each model fit and discarded the first 750 iterations as warmup. Chains still mixed relatively well despite the shorter chains (greatest $\hat{R} \approx 1.021$), though effective sample sizes were unsurprisingly lower compared to results in the main text. Nonetheless, bulk effective sample sizes always exceeded the minimum recommended 100 per chain (Vehtari et al., 2021), and all tail effective sizes exceeded 100. Divergent transitions remained relatively rare, with 18 fits exhibiting a single divergent transition and another 4 with 2-5 each. Most low tail effective sample sizes and divergent transitions were associated with loose prior settings, likely reflecting difficulty in sampling the fat tails of posteriors under such priors.

Overall results suggest that *evorates* is robust to alternate prior specifications unless the priors are overly informative (Figs. S22-24; Tables S8-19). In particular, shifting the root rate prior location had little effect on posterior distributions provided the prior's scale is larger than the shift magnitude (as in the case of default and loose prior settings). Unsurprisingly, posterior precision generally decreased with more uninformative priors, and loose priors thus tended to yield less accurate posteriors with higher median absolute errors. Counterintuitively, however, default prior settings often resulted in more accurate posteriors than tight prior settings. In the case of branchwise rates, this is likely due to lower estimates of rate variance under tight priors, increasing the shrinkage of branchwise rate estimates (Fig. S25). In the case of trend and root rate inference, this phenomenon

mostly occurred when the root rate prior and simulated trend "conflict" by implying different patterns of rate change over time (e.g., a root rate prior shifted by -3 suggests rates must have increased over time to yield the observed trait data, while a decreasing trend implies the opposite). Accordingly, posterior coverage remained essentially constant at ∼95% under default and loose prior settings, but dropped significantly–sometimes as low as 10%–under tight prior settings when the root rate prior and simulated trend conflicted in this manner.

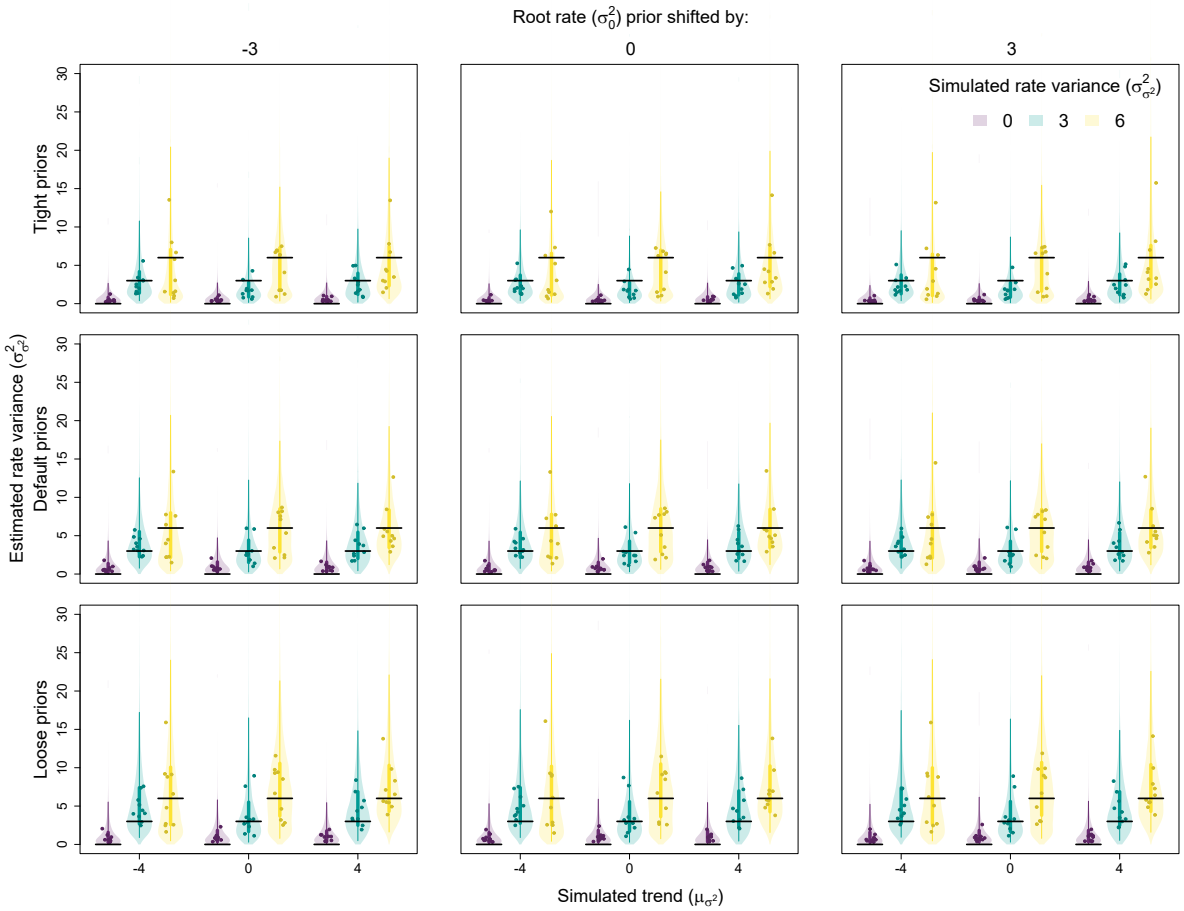

Figure S22. The effect of trait evolution scenario and prior settings on inference of the rate variance parameter ($\sigma^2_{\sigma^2}$). Each point is the posterior median from a single fit, while the violins are combined posterior distributions from all fits for a given trait evolution scenario and prior setting. Vertical lines represent the 50% (thicker lines) and 95% equal-tailed intervals (thinner lines) of these combined posteriors, while horizontal lines represent positions of true, simulated values.

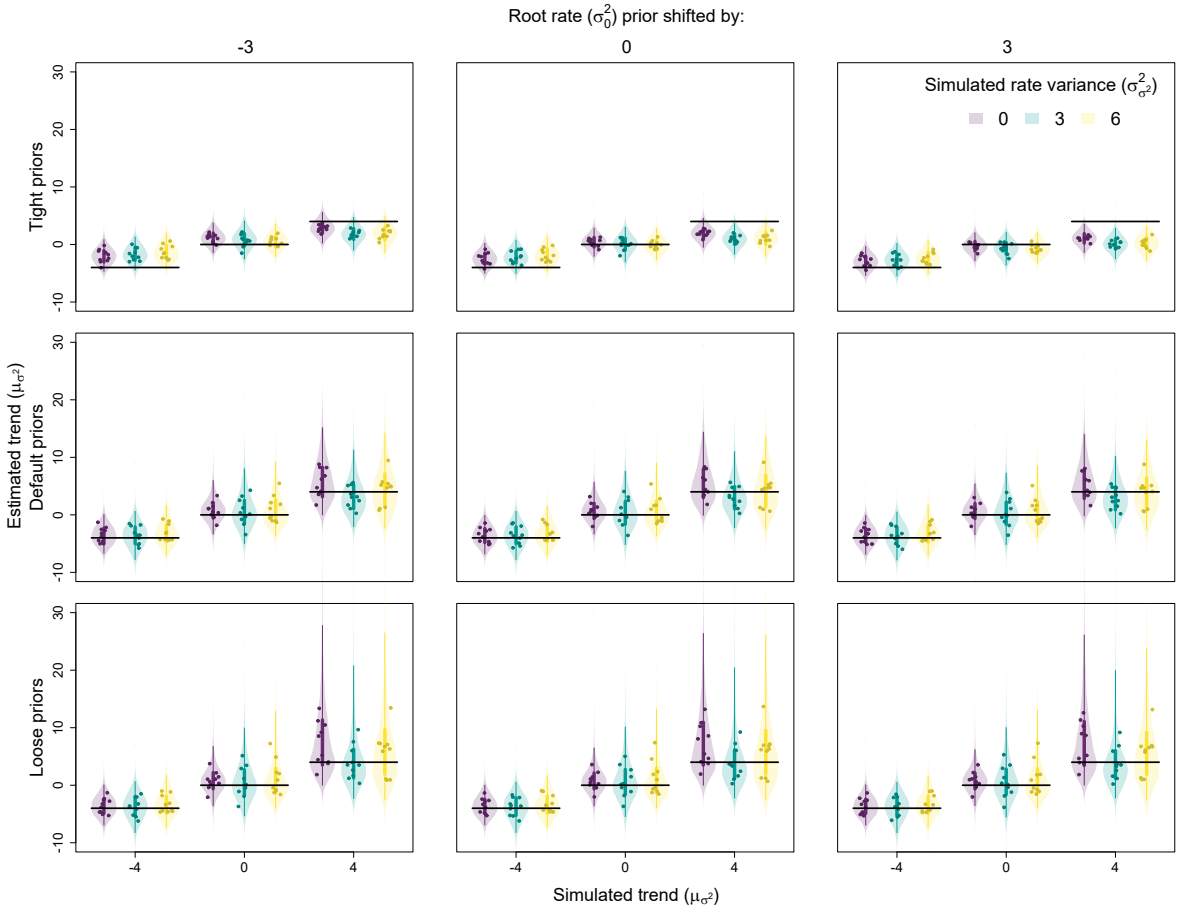

Figure S23. The effect of trait evolution scenario and prior settings on inference of the trend parameter (μ_{σ^2}). Each point is the posterior median from a single fit, while the violins are combined posterior distributions from all fits for a given trait evolution scenario and prior setting. Vertical lines represent the 50% (thicker lines) and 95% equal-tailed intervals (thinner lines) of these combined posteriors, while horizontal lines represent positions of true, simulated values.

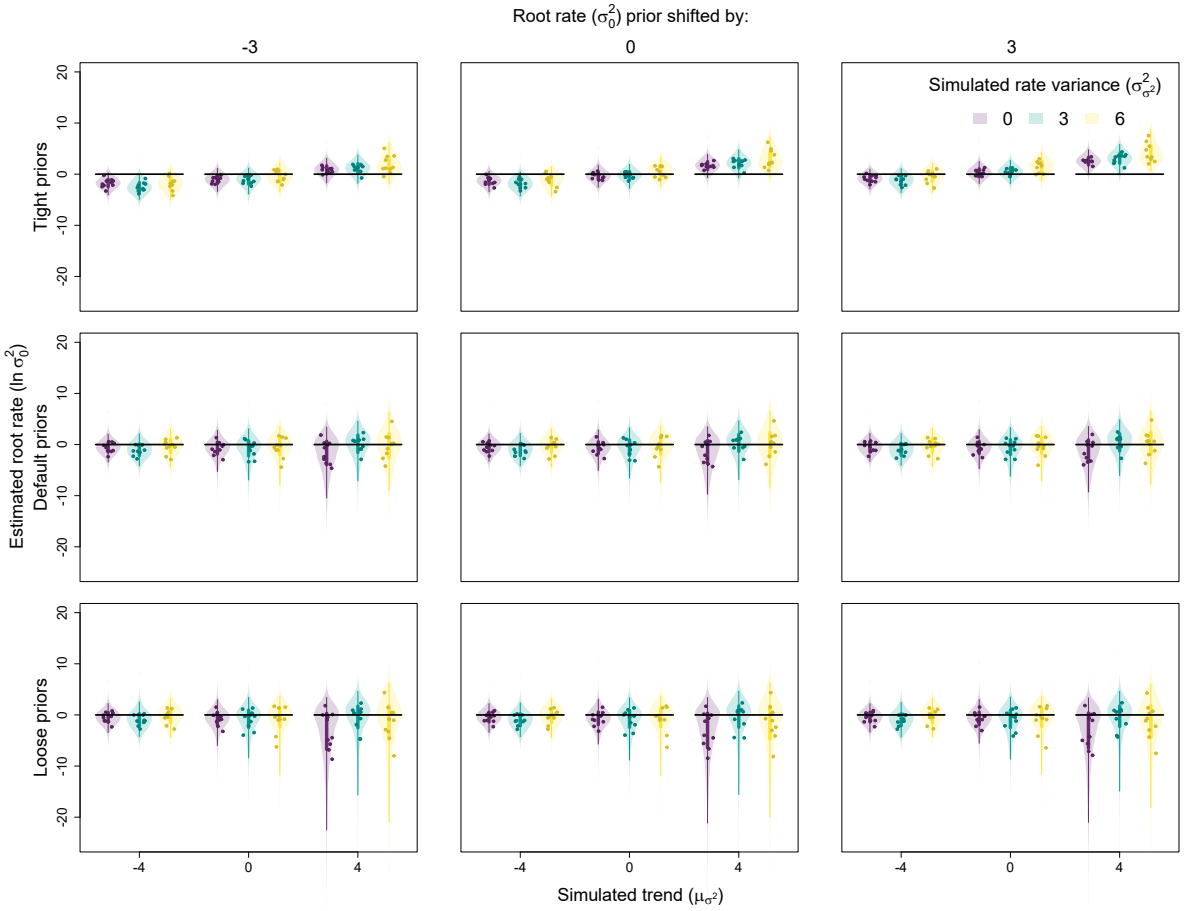

Figure S24. The effect of trait evolution scenario and prior settings on inference of the root rate parameter (σ_0^2). Each point is the posterior median from a single fit, while the violins are combined posterior distributions from all fits for a given trait evolution scenario and prior setting. Vertical lines represent the 50% (thicker lines) and 95% equal-tailed intervals (thinner lines) of these combined posteriors, while horizontal lines represent positions of true, simulated values.

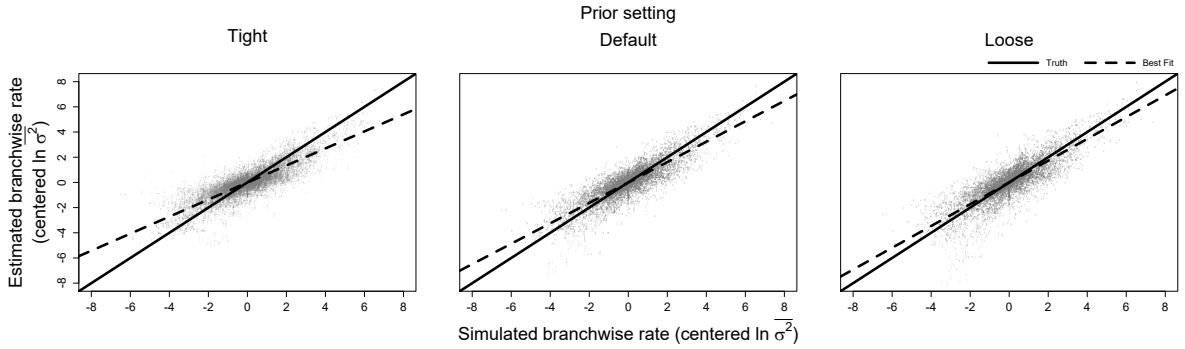

Figure S25. Relationship between simulated and estimated branchwise rate parameters (ln $\overline{\sigma^2}$) under different prior settings, with tight priors being the most informative and loose priors the least. For each simulation and posterior sample, branchwise rates were first centered by subtracting their mean. We estimated centered branchwise rates by taking the median of the centered posterior samples. The solid line represents the position of the true centered branchwise rates, while the shallower, dashed line represents the observed line of best fit for the data under each prior setting. Note that tighter, more informative priors result in shallower best fit lines due to increased shrinkage of branchwise rate estimates.

Table S8. *Median absolute errors of rate variance posteriors (i.e., median absolute difference between posterior samples and their true, simulated values, a measure of posterior distribution accuracy), averaged across replicates for each simulated trait evolution scenario and prior settings.* $\sigma^2_{\sigma^2}$ *and* μ_{σ^2} *indicate the true, simulated values of rate variance and trend parameters, respectively, while* σ^2_0 *prior shifts refer to alteration of the root rate prior location.*

		tight priors			default priors			loose priors		
$\sigma^2_{\sigma^2}$ =		0	3	6	0	3	6	0	3	6
					σ^2_0 prior shifted by -3					
μ_{σ^2} =	-4	0.46	1.58	4.14	0.70	1.54	3.49	0.79	2.30	3.97
	0	0.48	1.70	3.07	0.84	1.67	2.81	0.94	2.21	3.40
	4	0.52	1.62	3.39	0.82	1.72	2.79	0.93	2.27	3.08
					σ^2_0 prior shifted by 0					
μ_{σ^2} =	-4	0.43	1.52	3.98	0.68	1.53	3.51	0.79	2.33	4.04
	0	0.45	1.71	3.04	0.81	1.65	2.80	0.95	2.20	3.34
	4	0.51	1.63	3.50	0.83	1.72	2.88	0.94	2.34	3.00
					σ^2_0 prior shifted by 3					
μ_{σ^2} =	-4	0.40	1.52	4.14	0.69	1.53	3.66	0.79	2.28	3.95
	0	0.47	1.74	3.10	0.84	1.69	2.73	0.96	2.20	3.46
	4	0.52	1.65	3.73	0.83	1.72	2.82	0.94	2.23	3.09

Table S9. *Median absolute errors of trend posteriors (i.e., median absolute difference between posterior samples and their true, simulated values, a measure of posterior distribution accuracy), averaged across replicates for each simulated trait evolution scenario and prior settings.* $\sigma^2_{\sigma^2}$ *and* μ_{σ^2} *indicate the true, simulated values of rate variance and trend parameters, respectively, while* σ^2_0 *prior shifts refer to alteration of the root rate prior location.*

		tight priors			default priors			loose priors		
$\sigma^2_{\sigma^2}$ =		0	3	6	0	3	6	0	3	6
					σ^2_0 prior shifted by -3					
μ_{σ^2} =	-4	2.03	2.34	2.68	1.32	1.57	1.65	1.35	1.64	1.64
	0	1.28	1.32	1.07	1.55	2.22	2.11	1.65	2.45	2.53
	4	1.34	2.22	2.18	2.75	2.33	2.78	4.24	2.88	3.86
					σ^2_0 prior shifted by 0					
μ_{σ^2} =	-4	1.63	1.88	2.16	1.30	1.60	1.61	1.32	1.61	1.66
	0	0.91	1.07	0.94	1.54	2.21	2.05	1.64	2.43	2.51
	4	2.06	3.04	2.97	2.61	2.30	2.77	4.07	2.85	3.82
					σ^2_0 prior shifted by 3					
μ_{σ^2} =	-4	1.28	1.50	1.69	1.32	1.58	1.62	1.34	1.62	1.62
	0	0.88	1.16	1.12	1.51	2.14	1.98	1.64	2.50	2.50
	4	2.94	3.91	3.77	2.51	2.35	2.68	4.15	2.78	3.68

Table S10. *Median absolute errors of branchwise rate posteriors (i.e., median absolute difference between posterior samples and their true, simulated values, a measure of posterior distribution accuracy), averaged across replicates for each simulated trait evolution scenario and prior settings.* $\sigma^2_{\sigma^2}$ *and* μ_{σ^2} *indicate the true, simulated values of rate variance and trend parameters, respectively, while* σ^2_0 *prior shifts refer to alteration of the root rate prior location.*

		tight priors			default priors			loose priors		
$\sigma^2_{\sigma^2}$ =		0	3	6	0	3	6	0	3	6
					σ^2_0 prior shifted by -3					
μ_{σ^2} =	-4	0.53	0.87	0.98	0.48	0.83	0.90	0.50	0.86	0.91
	0	0.44	0.76	0.94	0.52	0.83	1.01	0.54	0.87	1.07
	4	0.46	0.83	0.93	0.64	0.87	1.02	0.82	0.95	1.16
					σ^2_0 prior shifted by 0					
μ_{σ^2} =	-4	0.47	0.83	0.94	0.48	0.83	0.90	0.49	0.86	0.91
	0	0.40	0.73	0.95	0.51	0.82	1.01	0.53	0.87	1.07
	4	0.52	0.88	0.99	0.63	0.87	1.01	0.80	0.95	1.16
					σ^2_0 prior shifted by 3					
μ_{σ^2} =	-4	0.43	0.79	0.92	0.48	0.82	0.90	0.50	0.86	0.91
	0	0.39	0.73	0.97	0.51	0.82	1.01	0.54	0.88	1.06
	4	0.61	0.95	1.06	0.62	0.87	1.00	0.81	0.94	1.14

Table S11. *Median absolute errors of root rate posteriors (i.e., median absolute difference between posterior samples and their true, simulated values, a measure of posterior distribution accuracy), averaged across replicates for each simulated trait evolution scenario and prior settings.* $\sigma^2_{\sigma^2}$ *and* μ_{σ^2} *indicate the true, simulated values of rate variance and trend parameters, respectively, while* σ^2_0 *prior shifts refer to alteration of the root rate prior location.*

		tight priors			default priors			loose priors		
$\sigma^2_{\sigma^2}$ =		0	3	6	0	3	6	0	3	6
					σ^2_0 prior shifted by -3					
μ_{σ^2} =	-4	1.78	2.54	1.97	1.06	1.41	1.39	1.07	1.45	1.40
	0	1.18	1.24	1.18	1.38	1.66	1.82	1.45	1.84	2.20
	4	1.08	1.36	2.16	2.49	1.83	2.49	3.87	2.39	3.43
					σ^2_0 prior shifted by 0					
μ_{σ^2} =	-4	1.38	1.88	1.40	1.04	1.41	1.36	1.06	1.41	1.41
	0	0.82	0.82	1.18	1.34	1.64	1.79	1.45	1.84	2.17
	4	1.71	2.17	3.05	2.39	1.81	2.50	3.73	2.33	3.36
					σ^2_0 prior shifted by 3					
μ_{σ^2} =	-4	1.02	1.33	1.09	1.06	1.38	1.39	1.06	1.41	1.38
	0	0.79	0.88	1.64	1.33	1.56	1.76	1.44	1.87	2.21
	4	2.54	3.07	4.15	2.27	1.83	2.41	3.80	2.27	3.27

Table S12. *Breadths of rate variance posteriors (i.e., the difference between the 97.5% and 2.5% quantiles of posterior samples, a measure of posterior distribution precision), averaged across replicates for each simulated trait evolution scenario and prior settings.* $\sigma^2_{\sigma^2}$ *and* μ_{σ^2} *indicate the true, simulated values of rate variance and trend parameters, respectively, while* σ^2_0 *prior shifts refer to alteration of the root rate prior location.*

		tight priors			default priors			loose priors		
$\sigma^2_{\sigma^2}$ =		0	3	6	0	3	6	0	3	6
					σ^2_0 prior shifted by -3					
μ_{σ^2} =	-4	2.31	8.74	11.61	3.83	10.48	13.13	4.86	14.17	15.74
	0	2.42	6.13	10.76	4.24	8.76	12.73	5.33	11.29	16.48
	4	2.40	7.14	11.94	3.89	9.60	13.33	4.81	12.18	16.82
					σ^2_0 prior shifted by 0					
μ_{σ^2} =	-4	2.20	7.73	11.23	3.84	10.45	13.05	4.60	14.78	15.98
	0	2.23	6.34	10.18	4.21	8.42	13.17	5.14	11.20	16.64
	4	2.40	7.04	11.75	3.93	9.44	13.65	4.78	12.35	16.75
					σ^2_0 prior shifted by 3					
μ_{σ^2} =	-4	2.10	7.82	11.03	3.82	10.31	12.66	4.74	14.44	16.06
	0	2.26	6.10	10.49	4.02	8.34	12.68	5.21	11.50	16.57
	4	2.50	6.90	12.57	4.04	9.28	13.47	4.90	12.18	17.02

Table S13. *Breadths of trend posteriors (i.e., the difference between the 97.5% and 2.5% quantiles of posterior samples, a measure of posterior distribution precision), averaged across replicates for each simulated trait evolution scenario and prior settings.* $\sigma^2_{\sigma^2}$ *and* μ_{σ^2} *indicate the true, simulated values of rate variance and trend parameters, respectively, while* σ^2_0 *prior shifts refer to alteration of the root rate prior location.*

		tight priors			default priors			loose priors		
$\sigma^2_{\sigma^2} =$		0	3	6	0	3	6	0	3	6
					σ^2_0 prior shifted by -3					
$\mu_{\sigma^2} =$	-4	3.62	4.34	4.61	4.87	6.18	6.68	4.99	6.70	7.08
	0	4.42	4.95	4.80	6.77	8.67	8.60	7.53	9.72	10.23
	4	4.90	5.19	5.38	12.14	10.74	12.73	21.26	15.61	21.57
					σ^2_0 prior shifted by 0					
$\mu_{\sigma^2} =$	-4	3.63	4.40	4.64	4.81	6.25	6.73	4.97	6.72	7.04
	0	4.23	4.84	4.76	6.77	8.46	8.52	7.44	9.90	10.47
	4	4.64	5.00	5.18	11.57	10.51	12.32	19.99	15.26	20.95
					σ^2_0 prior shifted by 3					
$\mu_{\sigma^2} =$	-4	3.64	4.37	4.65	4.85	6.23	6.69	4.90	6.64	6.94
	0	4.22	4.66	4.66	6.81	8.23	8.52	7.36	10.13	10.61
	4	4.63	4.90	5.20	11.56	10.28	11.82	19.62	15.68	19.45

Table S14. *Breadths of branchwise rate posteriors (i.e., the difference between the 97.5% and 2.5% quantiles of posterior samples, a measure of posterior distribution precision), averaged across replicates for each simulated trait evolution scenario and prior settings.* $\sigma^2_{\sigma^2}$ *and* μ_{σ^2} *indicate the true, simulated values of rate variance and trend parameters, respectively, while* σ^2_0 *prior shifts refer to alteration of the root rate prior location.*

		tight priors			default priors			loose priors		
$\sigma^2_{\sigma^2}$ =		0	3	6	0	3	6	0	3	6
					σ^2_0 prior shifted by -3					
μ_{σ^2} =	-4	2.01	3.06	3.21	2.33	3.36	3.49	2.41	3.61	3.67
	0	2.05	2.73	3.37	2.52	3.28	3.85	2.65	3.52	4.20
	4	2.15	2.91	3.39	3.10	3.59	4.14	4.32	4.25	5.26
					σ^2_0 prior shifted by 0					
μ_{σ^2} =	-4	1.98	3.00	3.18	2.33	3.36	3.49	2.41	3.61	3.66
	0	2.04	2.72	3.33	2.51	3.26	3.83	2.66	3.54	4.21
	4	2.14	2.89	3.37	3.03	3.57	4.10	4.11	4.24	5.23
					σ^2_0 prior shifted by 3					
μ_{σ^2} =	-4	1.97	2.98	3.17	2.33	3.36	3.49	2.41	3.61	3.66
	0	2.04	2.70	3.36	2.49	3.22	3.83	2.63	3.57	4.21
	4	2.15	2.90	3.41	3.03	3.53	4.06	4.07	4.26	5.03

Table S15. *Breadths of root rate posteriors (i.e., the difference between the 97.5% and 2.5% quantiles of posterior samples, a measure of posterior distribution precision), averaged across replicates for each simulated trait evolution scenario and prior settings.* $\sigma^2_{\sigma^2}$ *and* μ_{σ^2} *indicate the true, simulated values of rate variance and trend parameters, respectively, while* σ^2_0 *prior shifts refer to alteration of the root rate prior location.*

		tight priors			default priors			loose priors		
$\sigma^2_{\sigma^2}$ =		0	3	6	0	3	6	0	3	6
					σ^2_0 prior shifted by -3					
μ_{σ^2} =	-4	2.91	3.76	3.85	4.03	5.10	5.54	4.13	5.69	5.86
	0	3.86	4.21	4.34	5.94	7.34	7.55	6.52	8.47	9.18
	4	4.33	4.56	4.83	10.91	9.50	11.60	19.65	14.21	19.87
					σ^2_0 prior shifted by 0					
μ_{σ^2} =	-4	2.92	3.65	3.84	4.03	5.22	5.51	4.17	5.52	5.80
	0	3.71	4.10	4.20	5.98	7.21	7.37	6.61	8.51	9.39
	4	4.17	4.43	4.78	10.40	9.38	11.22	18.51	13.81	19.42
					σ^2_0 prior shifted by 3					
μ_{σ^2} =	-4	3.01	3.63	3.94	4.06	5.14	5.59	4.12	5.66	5.76
	0	3.73	4.09	4.29	5.87	7.10	7.40	6.41	8.71	9.39
	4	4.13	4.41	4.80	10.25	9.10	10.72	18.05	14.09	17.82

Table S16. *Coverage of rate variance posteriors (i.e., proportion of times the true, simulated value is greater than the 2.5% posterior distribution quantile and less than the 97.5% quantile) for each simulated trait evolution scenario and prior settings.* $\sigma^2_{\sigma^2}$ *and* μ_{σ^2} *indicate the true, simulated values of rate variance and trend parameters, respectively, while* σ^2_0 *prior shifts refer to alteration of the root rate prior location.*

		tight priors			default priors			loose priors		
$\sigma^2_{\sigma^2} =$		0	3	6	0	3	6	0	3	6
					σ^2_0 prior shifted by -3					
$\mu_{\sigma^2} =$	-4	—	1.00	0.70	—	1.00	1.00	—	1.00	0.90
	0	—	0.90	0.80	—	1.00	1.00	—	1.00	1.00
	4	—	1.00	0.90	—	1.00	1.00	—	1.00	0.90
					σ^2_0 prior shifted by 0					
$\mu_{\sigma^2} =$	-4	—	1.00	0.70	—	1.00	1.00	—	1.00	0.90
	0	—	0.90	0.70	—	1.00	1.00	—	1.00	1.00
	4	—	1.00	0.80	—	1.00	0.90	—	1.00	0.90
					σ^2_0 prior shifted by 3					
$\mu_{\sigma^2} =$	-4	—	1.00	0.60	—	1.00	1.00	—	1.00	0.90
	0	—	0.90	0.70	—	1.00	1.00	—	1.00	1.00
	4	—	1.00	0.80	—	1.00	1.00	—	1.00	0.90

Table S17. *Coverage of trend posteriors (i.e., proportion of times the true, simulated value is greater than the 2.5% posterior distribution quantile and less than the 97.5% quantile) for each simulated trait evolution scenario and prior settings.* $\sigma^2_{\sigma^2}$ *and* μ_{σ^2} *indicate the true, simulated values of rate variance and trend parameters, respectively, while* σ^2_0 *prior shifts refer to alteration of the root rate prior location.*

		tight priors			default priors			loose priors		
$\sigma^2_{\sigma^2} =$		0	3	6	0	3	6	0	3	6
					σ^2_0 prior shifted by -3					
$\mu_{\sigma^2} =$	-4	0.50	0.40	0.40	1.00	0.90	1.00	1.00	1.00	1.00
	0	1.00	1.00	1.00	1.00	1.00	0.90	1.00	1.00	0.90
	4	1.00	0.70	0.80	1.00	1.00	1.00	0.90	1.00	0.90
					σ^2_0 prior shifted by 0					
$\mu_{\sigma^2} =$	-4	0.50	0.70	0.60	1.00	1.00	1.00	1.00	1.00	1.00
	0	1.00	1.00	1.00	1.00	1.00	0.90	1.00	0.90	0.90
	4	0.80	0.10	0.30	1.00	1.00	1.00	1.00	1.00	0.90
					σ^2_0 prior shifted by 3					
$\mu_{\sigma^2} =$	-4	0.70	0.70	0.80	1.00	0.90	1.00	1.00	1.00	1.00
	0	1.00	1.00	1.00	1.00	0.90	0.90	1.00	0.90	0.90
	4	0.10	0.00	0.10	1.00	1.00	1.00	1.00	1.00	0.90

Table S18. *Coverage of branchwise rate posteriors (i.e., proportion of times the true, simulated value is greater than the 2.5% posterior distribution quantile and less than the 97.5% quantile) for each simulated trait evolution scenario and prior settings.* $\sigma^2_{\sigma^2}$ *and* μ_{σ^2} *indicate the true, simulated values of rate variance and trend parameters, respectively, while* σ^2_0 *prior shifts refer to alteration of the root rate prior location.*

		tight priors			default priors			loose priors		
$\sigma^2_{\sigma^2}$ =		0	3	6	0	3	6	0	3	6
					σ^2_0 prior shifted by -3					
μ_{σ^2} =	-4	0.91	0.89	0.86	0.98	0.95	0.94	0.98	0.96	0.95
	0	0.99	0.93	0.91	1.00	0.96	0.94	1.00	0.97	0.94
	4	1.00	0.91	0.92	0.99	0.96	0.96	0.97	0.97	0.95
					σ^2_0 prior shifted by 0					
μ_{σ^2} =	-4	0.94	0.91	0.87	0.98	0.95	0.94	0.98	0.96	0.95
	0	1.00	0.93	0.90	1.00	0.96	0.93	1.00	0.97	0.95
	4	0.97	0.87	0.89	0.99	0.96	0.95	0.98	0.97	0.96
					σ^2_0 prior shifted by 3					
μ_{σ^2} =	-4	0.97	0.93	0.88	0.98	0.95	0.94	0.98	0.96	0.95
	0	1.00	0.94	0.89	1.00	0.97	0.94	1.00	0.97	0.95
	4	0.89	0.83	0.86	0.99	0.96	0.95	0.98	0.97	0.96

Table S19. *Coverage of root rate posteriors (i.e., proportion of times the true, simulated value is greater than the 2.5% posterior distribution quantile and less than the 97.5% quantile) for each simulated trait evolution scenario and prior settings.* $\sigma^2_{\sigma^2}$ *and* μ_{σ^2} *indicate the true, simulated values of rate variance and trend parameters, respectively, while* σ^2_0 *prior shifts refer to alteration of the root rate prior location.*

		tight priors			default priors			loose priors		
$\sigma^2_{\sigma^2}$ =		0	3	6	0	3	6	0	3	6
					σ^2_0 prior shifted by -3					
μ_{σ^2} =	-4	0.40	0.30	0.60	1.00	0.80	1.00	0.90	0.90	1.00
	0	0.90	0.90	1.00	1.00	1.00	1.00	1.00	1.00	0.90
	4	1.00	1.00	0.60	1.00	1.00	1.00	1.00	1.00	1.00
					σ^2_0 prior shifted by 0					
μ_{σ^2} =	-4	0.60	0.60	0.80	1.00	0.80	1.00	1.00	1.00	1.00
	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	4	0.80	0.40	0.60	1.00	1.00	1.00	1.00	1.00	1.00
					σ^2_0 prior shifted by 3					
μ_{σ^2} =	-4	0.90	0.80	0.90	1.00	0.80	1.00	1.00	0.80	1.00
	0	1.00	1.00	0.60	1.00	1.00	1.00	1.00	1.00	1.00
	4	0.10	0.10	0.10	1.00	1.00	1.00	1.00	1.00	1.00

Despite the relatively inaccurate inferences of branchwise rate, root rate, and trend parameters under overly informative priors, hypothesis testing was still largely reliable, albeit sometimes underpowered, under all prior settings we considered. Across the board, error rates remained conservative at around 5% or lower, with decreasing trends never mistaken for increasing trends and vice versa. Error rates for detecting significant rate variance may be slightly inflated under tight priors (Fig. S26), perhaps due to tighter constraints on trend estimation forcing the model to instead attribute apparent rate heterogeneity to rate variance. Nonetheless, power to detect significant rate variance appears consistent regardless of prior settings. Notably, the same is true for anomalous rate detection, despite the increasing shrinkage of branchwise rate estimation under tighter priors (Fig. S28). On the other hand, prior settings had considerable influence on power to detect trends (Fig. S27), with generally increasing power under looser priors – particularly

when the root rate prior shift and simulated trend both imply similar patterns of rate change over time (e.g., a root rate prior shifted by 3 and decreasing trend).

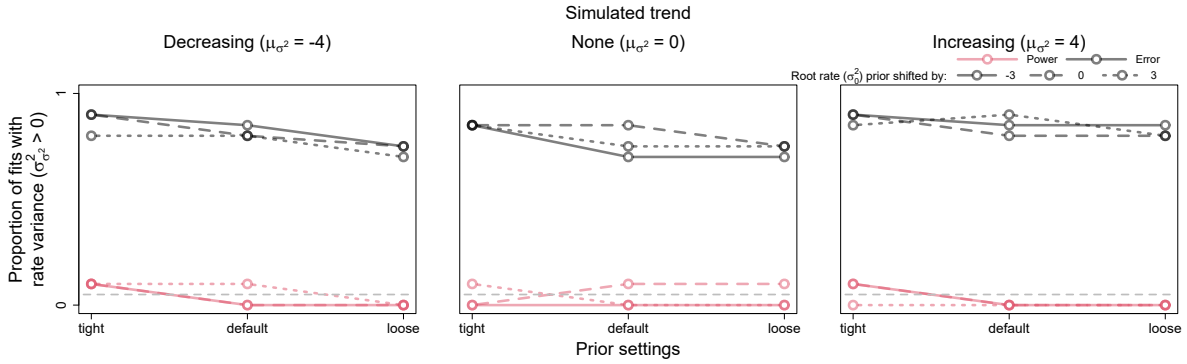

Figure S26. Power and error rates for the rate variance parameter ($\sigma^2_{\sigma^2}$). Lines depict changes in the proportion of model fits that correctly showed evidence for rate variance significantly greater than 0 (i.e., power, in black) and incorrectly showed evidence (i.e., error, in light red) as a function of prior settings, with tight priors being the most informative and loose priors the least. Results are also shown for fits with the location of the root rate (σ_0^2) prior shifted by -3 (solid lines), 0 (dashed lines), and 3 (dotted lines) from the default setting.

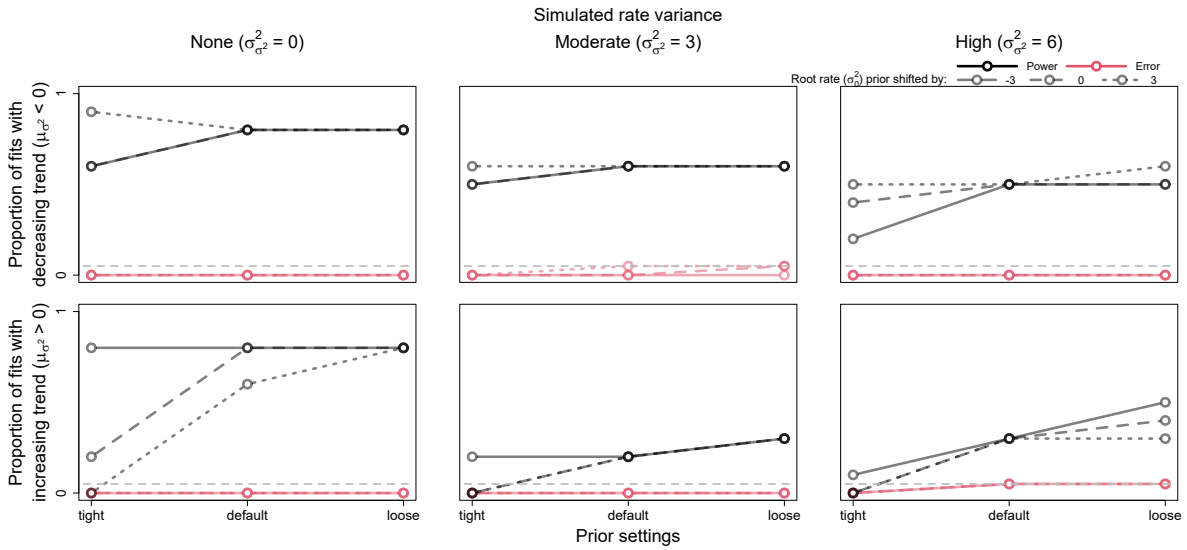

Figure S27. Power and error rates for the trend parameter (μ_{σ^2}). Lines depict changes in the proportion of model fits that correctly showed evidence for trends significantly less and greater than 0 (i.e., power, in black) and incorrectly showed evidence (i.e., error, in light red) as a function of prior settings, with tight priors being the most informative and loose priors the least. Results are also shown for fits with the location of the root rate (σ_0^2) prior shifted by -3 (solid lines), 0 (dashed lines), and 3 (dotted lines) from the default setting.

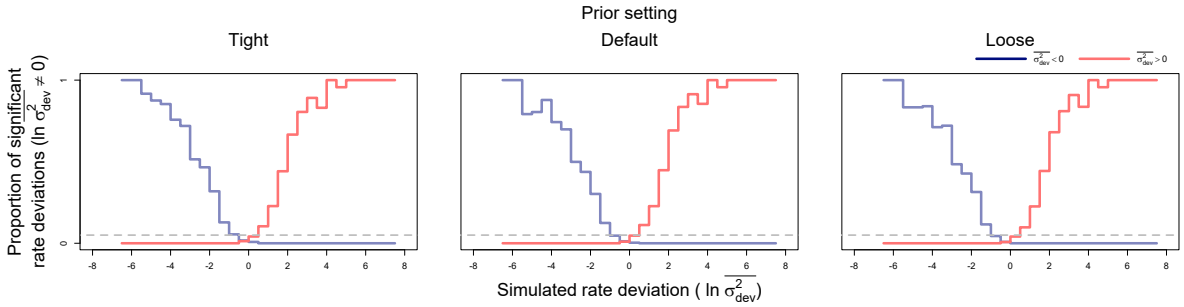

Figure S28. Power and error rates for branchwise rate parameters (ln $\overline{\sigma^2}$) under different prior settings. Lines depict changes in proportions of branchwise rates considered anomalously slow (in dark blue) or fast (in light red) as a function of simulated rate deviations (ln $\overline{\sigma^2_{dev}}$). These results combine all fits to simulated data that detected rate variance ($\sigma^2_{\sigma^2}$) significantly greater than 0. The proportions are equivalent to power when the detected rate deviation is of the same sign as the true, simulated deviation (left of 0 for anomalously slow rates in dark blue and right for anomalously fast rates in light red), and to error rate when the detected and true rate deviations are of opposite signs. Here, significant rate deviations for simulated rate deviations that are exactly 0 are considered errors regardless of sign.

References

Arnold P.W. and Heinsohn G.E. 1996. Phylogenetic status of the Irrawaddy dolphin *Orcaella brevirostris* (Owen in Gray): a cladistic analysis. Mem. Queensl. Mus. 39:143–204.

Baker A.N. 1981. The southern right whale dolphin *Lissodelphis peronii* (Lacépède) in Australasian waters. Natl. Mus. NZ Rec. 2:17–34.

Barros N.B. 1991. Recent cetacean records for southeastern brazil. Mar. Mamm. Sci. 7:296–306.

Benson R.B.J., Campione N.E., Carrano M.T., Mannion P.D., Sullivan C., Upchurch P., and Evans D.C. 2014. Rates of dinosaur body mass evolution indicate 170 million years of sustained ecological innovation on the avian stem lineage. PLoS Biol. 12:e1001853.

Beygelzimer A., Kakadet S., Langford J., Arya S., Mount D., and Li S. 2019. FNN: fast nearest neighbor search algorithms and applications. Version 1.1.3.

Blomberg S.P., Garland T. Jr, and Ives A.R. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57:717–745.

Charlton-Robb K., Gershwin L.A., Thompson R., Austin J., Owen K., and McKechnie S. 2011. A new dolphin species, the Burrunan dolphin *Tursiops australis sp. nov.*, endemic to southern Australian coastal waters. PLoS One 6:e24047.

Dalebout M.L., Mead J.G., Baker C.S., Baker A.N., and Helden A.L. 2002. A new species of beaked whale *Mesoplodon perrini sp. n.* (cetacea: ziphiidae) discovered through phylogenetic analyses of mitochondrial DNA sequences. Mar. Mamm. Sci. 18:577–608.

Dalebout M.L., Scott Baker C., Steel D., Thompson K., Robertson K.M., Chivers S.J., Perrin W.F., Goonatilake M., Charles Anderson R., Mead J.G., Potter C.W., Thompson L., Jupiter D., and Yamada T.K. 2014. Resurrection of *Mesoplodon hotaula*

Deraniyagala 1963: a new species of beaked whale in the tropical Indo-Pacific. Mar. Mamm. Sci. 30:1081–1108.

Devreese J.P.A., Lemmens D., and Tempere J. 2010. Path integral approach to Asian options in the Black-Scholes model. Phys. A: Stat. Mech. Appl. 389:780–788.

Dufresne D. 2004. The log-normal approximation in financial and other computations. Adv. Appl. Probab. 36:747–773.

Fortune S.M.E., Moore M.J., Perryman W.L., and Trites A.W. 2021. Body growth of North Atlantic right whales (*Eubalaena glacialis*) revisited. Mar. Mamm. Sci. 37:433–447.

Gill M.S., Tung Ho L.S., Baele G., Lemey P., and Suchard M.A. 2017. A relaxed directional random walk model for phylogenetic trait evolution. Syst. Biol. 66:299–319.

Hopkins M.J. and Smith A.B. 2015. Dynamic evolutionary change in post-Paleozoic echinoids and the importance of scale when interpreting changes in rates of evolution. Proc. Natl. Acad. Sci. U. S. A. 112:3758–3763.

Hunt G. 2006. Fitting and comparing models of phyletic evolution: random walks and beyond. Paleobiology 32:578–601.

Jefferson T.A. and Rosenbaum H.C. 2014. Taxonomic revision of the humpback dolphins (*Sousa spp.*), and description of a new species from Australia. Mar. Mamm. Sci. 30:1494–1541.

Konishi K., Tamura T., Zenitani R., Bando T., Kato H., and Walløe L. 2008. Decline in energy storage in the Antarctic minke whale (*Balaenoptera bonaerensis*) in the southern ocean. Polar Biol. 31:1509–1520.

Lartillot N. and Poujol R. 2011. A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. Mol. Biol. Evol. 28:729–744.

Lepage T., Bryant D., Philippe H., and Lartillot N. 2007. A general comparison of relaxed molecular clock models. Mol. Biol. Evol. 24:2669–2680.

Lloyd G.T., Wang S.C., and Brusatte S.L. 2012. Identifying heterogeneity in rates of morphological evolution: discrete character change in the evolution of lungfish (sarcopterygii; dipnoi). Evolution 66:330–348.

Lodi L., Sicilian S., and Capistran L. 1990. Mass stranding of *Peponocephala electra* (cetacea, globicephalinae) on Piracange Beach, Baria, northeastern Brazil. Sci. Rep. Cetacean Res. 1:79–84.

Mead J.G., Walker W.A., and Houck W.J. 1982. Biological observations on *Mesoplodon carlhubbsi* (cetacea, ziphiidae). Smithson. Contrib. Zool. Pages 1–25.

Molina D.M. and Oporto J.A. 1993. Comparative study of dentine staining techniques to estimate age in the Chilean dolphin, *Cephalorhynchus eutropia* (Gray, 1846). Aquat. Mamm. 19:45–48.

Plön S., Albrecht K.H., Cliff G., and Froneman P.W. 2012. Organ weights of three dolphin species (*Sousa chinensis*, *Tursiops aduncus* and *Delphinus capensis*) from South Africa: implications for ecological adaptation? J. Cetacean Res. Manag. 12:265–276.

Puttick M.N. 2018. Mixed evidence for early bursts of morphological evolution in extant clades. J. Evol. Biol. 31:502–515.

Raj Pant S., Goswami A., and Finarelli J.A. 2014. Complex body size trends in the evolution of sloths (xenarthra: pilosa). BMC Evol. Biol. 14:184.

Revell L.J. 2012. *phytools*: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3:217–223.

Revell L.J. 2021. A variable-rate quantitative trait evolution model using penalized-likelihood. PeerJ 9:e11997.

Reyes J.C., Mead J.G., and Van Waerebeek K. 1991. A new species of beaked whale *Mesoplodon peruvianus sp. n.* (cetacea: ziphiidae) from Peru. Mar. Mamm. Sci. 7:1–24.

Safak A. and Safak M. 2002. Moments of the sum of correlated log-normal random variables. *in* Proceedings of IEEE Vehicular Technology Conference (VTC).

Slater G.J. and Pennell M.W. 2014. Robust regression and posterior predictive simulation increase power to detect early bursts of trait evolution. Syst. Biol. 63:293–308.

Slater G.J., Price S.A., Santini F., and Alfaro M.E. 2010. Diversity versus disparity and the radiation of modern cetaceans. Proc. Biol. Sci. 277:3097–3104.

Sookias R.B., Butler R.J., and Benson R.B.J. 2012. Rise of dinosaurs reveals major body-size transitions are driven by passive processes of trait evolution. Proc. Biol. Sci. 279:2180–2187.

Thompson K., Baker C.S., van Helden A., Patel S., Millar C., and Constantine R. 2012. The world's rarest whale. Curr. Biol. 22:R905–6.

Thorne J.L., Kishino H., and Painter I.S. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15:1647–1657.

Vehtari A., Gelman A., Simpson D., Carpenter B., and Bürkner P.C. 2021. Rank-normalization, folding, and localization: an improved $\hat{R}$ for assessing convergence of MCMC (with discussion). Bayesian Anal. 16:667–718.

Welch J.J. and Waxman D. 2008. Calculating independent contrasts for the comparative study of substitution rates. J. Theor. Biol. 251:667–678.

Wright D.F. 2017. Phenotypic innovation and adaptive constraints in the evolutionary radiation of Palaeozoic crinoids. Sci. Rep. 7:13745.

Zhao P. and Lai L. 2021. On the convergence rates of KNN density estimation. Pages 2840–2845 *in* Proceedings of IEEE International Symposium on Information Theory (ISIT).