



HAL
open science

The birth-death diffusion leading to present-day Mammal diversity

Ignacio Quintero, Nicolas Lartillot, H el ene Morlon

► **To cite this version:**

Ignacio Quintero, Nicolas Lartillot, H el ene Morlon. The birth-death diffusion leading to present-day Mammal diversity. 2022. hal-03866088

HAL Id: hal-03866088

<https://hal.science/hal-03866088>

Preprint submitted on 22 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.

1

Version dated: August 9, 2022

2

The birth-death diffusion leading to present-day Mammal diversity

3

4

IGNACIO QUINTERO¹, NICOLAS LARTILLOT² AND HÉLÈNE MORLON¹

5

¹ *Institut de Biologie de l'ENS (IBENS), Département de biologie, École normale supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France*

6

7

² *Laboratoire de Biométrie et Biologie Évolutive UMR, CNRS, Université de Lyon, Université Lyon 1, 69622 Villeurbanne, France*

8

9

Corresponding author: Ignacio Quintero, Institut de Biologie de l'ENS (IBENS), Département de biologie, École normale supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France. E-mail: ignacioquinterom@gmail.com.

10

11

12

Abstract

13

Dramatic spatial, temporal and taxonomic variation in biodiversity is ultimately explained by differences in speciation and extinction rates. Mammals represent a ~200 My old radiation that resulted in over 6500 extant species, with stark temporal, spatial and taxonomic heterogeneity in biodiversity. Throughout their history, every mammal lineage is expected to have undergone diversification rates that vary instantaneously in time resulting from the complex interplay of context-specific extrinsic factors (*e.g.*, K-Pg mass extinction event, rise of angiosperms) with their evolving ecologies (*e.g.*, body size, diet). When studying the diversification history of a clade, however, mathematical and computational limitations have hindered inference of such a flexible birth-death model where speciation and extinction rates evolve continuously along a phylogenetic tree. Here we overcome these challenges by implementing a series of phylogenetic models in which speciation and extinction rates are inherited and diffuse following a latent Geometric Brownian motion process. We enable full Bayesian inference using data augmentation techniques to sample from the posterior distribution of model parameters, including augmented phylogenetic trees and validate using simulations. Using a genome-informed time-calibrated tree for over 4000 Mammals species, we are able to estimate a complete and fine-grained picture of the variation in diversification rates that captures both global and lineage specific effects. We find that, contrary to the idea of a suppressed mammalian diversification before the K-Pg mass extinction event (*i.e.*, explosive- or delayed-rise), mammal speciation rates dramatically increased around 10-20 My before the K-Pg. Our new model opens exciting possibilities in disentangling the drivers behind variation in diversification and assaying how small-scale processes scale-up to macroevolutionary dynamics.

30

31 Introduction

32 Understanding the tempo and mode in which lineages diversify is fundamental in explaining the origin and
33 maintenance of biodiversity. The rates at which species originate or go extinct result from the interplay
34 between their intrinsic traits and their specific abiotic and biotic environment (Benton, 2009). For instance,
35 environmental oscillations and landscape heterogeneity are commonly posited as major drivers of diversifica-
36 tion by precipitating population dispersal and fragmentation as well as opening new opportunities (Barnosky,
37 2001; Jablonski, 2008). Similarly, the evolution of both intrinsic (*e.g.*, species phenotype, their niche, and
38 their evolutionary rate) and extrinsic biotic factors (*e.g.*, competition and other inter-specific interactions)
39 are thought to affect the pace of evolutionary radiations (Van Valen, 1973; Benton, 2009; Quintero and
40 Landis, 2019). Therefore, the interrelation of fluctuating context-specific dynamics with species' intrinsic
41 evolving ecologies are expected to result in lineage-specific diversification rates that themselves evolve and
42 can vary at any point in time.

43 The mammal radiation started at around 200 Mya (Upham et al., 2019; Álvarez-Carretero et al., 2021) and
44 resulted in an estimated present-day diversity of *ca.* 6450 currently recognized species (Mammal Diversity
45 Database, 2022). Distinct mammalian evolutionary routes led to marked differences in ecomorphologies,
46 reflected in body size (10^8 fold differences), generation time, litter size and habitat (aquatic, arboreal,
47 terrestrial, fossorial, etc.) variation, concomitant with an uneven distribution of richness across clades (Davies
48 et al., 2008; Meredith et al., 2011; Grossnickle et al., 2019). Furthermore, throughout their long evolutionary
49 history, lineages were differentially impacted by environmental factors such as the radiation of flowering
50 plants (*i.e.*, the Cretaceous Terrestrial Revolution), the K-Pg extinction event, the Paleocene-Eocene thermal
51 maximum, and other dramatic environmental oscillations, which likely spurred widespread distribution shifts
52 and extinctions together with novel ecological opportunities that impacted diversification rates (Meredith
53 et al., 2011; Grossnickle et al., 2019; Upham et al., 2021). Indeed, a major unsolved debate in mammalian
54 evolution revolves around understanding the timing at which the orders with living representatives originated
55 and diversified (Bininda-Emonds et al., 2007; Stadler, 2011a; Meredith et al., 2011; Grossnickle et al., 2019;
56 Springer et al., 2019; Upham et al., 2021). Standing hypotheses posit that crown orders increased their
57 diversification either before, at or after the K-Pg event, dubbed as 'early-', 'explosive-' or 'delayed-' rise
58 of extant mammals, respectively (Meredith et al., 2011; Grossnickle et al., 2019). The complex interplay
59 of mammal species-specific ecomorphologies with their particular environments that fluctuate throughout
60 lineage's duration translate into lineage- and time- specific rates of diversification along their evolutionary
61 history, that is, a given lineage is expected to undergo diversification rate changes at any moment in time.
62 Nonetheless, extreme external events, such as the rise of angiosperms or the K-Pg extinction event, are
63 thought to transcend lineage-specific diversification dynamics, and leave a common signature across lineages
64 during that particular period (Barnosky, 2001). Therefore, to explore the temporal dynamics that led to
65 extant mammal diversity, an idealized model of diversification that enable the reconstruction of overarching
66 temporal dynamics while incorporating rates of speciation and extinction that change instantaneously along
67 time for any lineage is needed.

68 Several phylogenetic approaches have been developed to account for heterogeneity in time or across

69 lineages in characterizing diversification rate variation across taxa. Most available methods that incorporate
70 rate heterogeneity across lineages assume a (usually few) number of independent shifts across the tree that
71 partition the tree into separate ‘regimes’ (or states) wherein lineages undergo constant rates: ‘BAMM’
72 (Rabosky, 2014), ‘birth-death-shift’ process (Höhna et al., 2019), ‘MTBD’ (Barido-Sottani et al., 2020), and
73 ‘MiSSE’ (Vasconcelos et al., 2022). Instead of identifying large diversification shifts along a phylogenetic
74 tree, ‘ClaDS’ assumes that shifts occur at each cladogenetic event, with daughter lineages undergoing rate
75 constancy within each lineage (Maliet et al., 2019). The only method we are aware of that allows lineage-
76 specific speciation and extinction rates to vary instantaneously through time, ‘QuaSSE’, assumes that this
77 variation is completely explained by the evolution of a trait under Brownian motion (FitzJohn, 2010). While
78 these methods have proven very useful in reconstructing diversification dynamics on phylogenetic trees, they
79 remain restrictive in assuming rate constancy either across ‘regimes’ or along a given branch, and a more
80 flexible model is required.

81 Here we assess the tempo and mode in which surviving mammal orders originated and diversified by
82 applying the first phylogenetic diversification model in which lineage-specific diversification rates vary in-
83 stantaneously in time by assuming that speciation and extinction rates follow a Geometric Brownian motion:
84 the Birth-Death Diffusion (BDD) model. We show that the BDD and other simpler diffusion models, in-
85 cluding no extinction, constant-extinction and constant-turnover, exhibit good statistical properties in most
86 scenarios and are identifiable even when restricted to extant taxa alone. Considering that the temporal
87 dynamics of mammalian diversification will rely on the accuracy of the phylogenetic tree, we use the latest
88 time-calibrated molecular tree for mammals, which incorporated 72 species genomes enhancing the time-
89 dating robustness (Álvarez-Carretero et al., 2021). We validate our method and perform Bayesian inference
90 of our diffusion models on this mammalian tree using data augmentation techniques. On top of posterior
91 probabilities for the main process parameters, our model also returns for free posterior sampled histories of
92 diversification across trees with unobserved speciation events enabling post-hoc analyses and visualizations.

93 Model

94 We assume that, at some time t , each lineage l has an instantaneous rate of producing new species of $\lambda_l(t)$
95 (*i.e.*, speciation rate) and an instantaneous rate of going extinct of $\mu_l(t)$ (*i.e.*, extinction rate). This general
96 birth-death process generates a bifurcating phylogenetic tree with some lineages dying out and others giving
97 rise to daughter species after some time. Probabilistic inference is complicated since, in practice, we do not
98 observe the whole process, but rather the evolutionary relationships among those lineages that were able to be
99 sampled, that is, the “reconstructed” phylogenetic tree (Nee et al., 1994). Our goal is to perform inference
100 on speciation and extinction rates that are inherited and stochastically diffuse through time following a
101 Geometric Brownian motion, given that we only observe the reconstructed tree. Because there is no available
102 analytical solution to estimate the likelihood, we use Bayesian data augmentation techniques to perform full
103 posterior inference on birth-death diffusion models. We start by describing the data augmented approach for
104 a simple constant rate birth-death (‘CBD’) model, which we then expand on to enable inference on models
105 with rate diffusion.

106 Let Ψ be an ultrametric rooted phylogenetic tree under a general birth-death process that starts at some
107 time T_Ψ (*i.e.*, tree height) in the past and continues to time 0 in the present with n surviving tips and per-
108 lineage speciation rate $\lambda_l(t)$ and extinction-rate $\mu_l(t)$ for lineage l at time t . Furthermore, each clade c has
109 a specific probability for extant lineages to be represented in the observed tree, specified by $\rho_c \in [0, 1]$. We
110 propagate this probability throughout all the branches in the tree by specifying a branch-specific sampling
111 fraction ρ_b for branch b . For terminal branches in clade c we simply assign $\rho_b = \rho_c \forall b \in c$. Let a_b be the
112 number of alive tips descending from branch b in the observed tree, then, for internal branches, we calculate
113 $\rho_b = (a_{d1} + a_{d2}) / (a_{d1} / \rho_{d1} + a_{d2} / \rho_{d2})$, where $d1$ and $d2$ are the daughter branches. Thus, the probability of
114 sampling exactly 1 extant species for a terminal branch b is $a_b \rho_b (1 - \rho_b)^{a_b - 1}$ and sampling no extant species
115 for internal branch b is $(1 - \rho_b)^{a_b}$, following Maliet and Morlon (2021).

116 We use Bayesian data augmentation (DA) to sample unobserved lineages that either went extinct in the
117 past or were not sampled at the present during inference. For clarity let Ψ_o represent only the reconstructed
118 (observed) tree, Ψ_a represent some unobserved speciation and extinction events and let $\Psi (= \Psi_o \cup \Psi_a)$
119 represent the complete tree. Given that we do not observe Ψ_a , we treat it as a random variable to integrate
120 over using Markov Chain Monte Carlo (MCMC).

121 Constant rate Birth-Death

122 In the CBD model, at any time, all lineages share the same speciation and extinction rate λ and μ , respectively
123 (*i.e.*, $\lambda_l(t) = \lambda$ and $\mu_l(t) = \mu$).

124 **Likelihood** The likelihood for a complete unordered crown tree under a CBD process is simply

$$\ell(\Psi|\lambda, \mu) = \lambda^{s-1} \mu^z e^{-(\lambda+\mu)L},$$

125 where s is the number of speciation events, z is the number of extinction events, L is the tree length (sum
126 of all branches). For a stem tree, we do consider all s speciation events.

127 To integrate over Ψ during MCMC, we developed two alternative approaches for data augmentation. A
128 first approach samples Ψ_a directly from its conditional distribution, given Ψ_o and the model parameters (*i.e.*,
129 Gibbs sampling) using forward simulation. A second approach is based on accept-reject sampling, based on
130 proposals of grafting and pruning. Each can be useful on different contexts, but we relied mostly on the
131 forward simulation approach, which we describe below, and leave the description of the grafting/pruning
132 approach for the Appendix.

133 **Forward simulation** Here we follow Maliet and Morlon (2021) in our forward simulation approach but
134 describe it more generally for the CBD process. First, we uniformly sample a branch in the tree and simulate
135 a birth-death process forward in time throughout the branch length t_b for branch b , given the current model
136 parameters. For any branch, if the process becomes extinct before t_b , we reject the proposal (*i.e.*, it requires
137 that at least 1 species is alive at time t_b). Note that for a terminal branch t_b is the present. If the branch
138 is internal, and there is more than one surviving lineage by t_b , we pick one tip at random as the one that
139 gave rise to the observed speciation event and continue the simulation in forward fashion for the others. All

140 other unobserved lineages should go extinct before the present or, if there are unobserved data augmented
 141 lineages at the present, then the proposal is rejected or accepted following the branch's sampling fraction,
 142 ρ_b .

143 Let ψ_b denote the forward simulated tree in branch b , $n_b(t_b)$ denote the number of lineages alive at time
 144 t_b (*i.e.*, at the end of the branch) and $n_b(0)$ denote the number of lineages alive at the present for ψ_b , then
 145 the acceptance ratio in the Metropolis-Hastings (MH) step for the proposed forward simulation ψ'_b if b is a
 146 terminal branch is

$$a = \min \left\{ 1, \frac{\ell(\Psi'|\lambda, \mu) \ell(\psi_b|\lambda, \mu)}{\ell(\Psi|\lambda, \mu) \ell(\psi'_b|\lambda, \mu)} \times \frac{n'_b(0)\rho_b(1-\rho_b)^{n'_b(0)-1}}{n_b(0)\rho_b(1-\rho_b)^{n_b(0)-1}} \right\},$$

147 and if b is an internal branch

$$a = \min \left\{ 1, \frac{\ell(\Psi'|\lambda, \mu) \ell(\psi_b|\lambda, \mu)}{\ell(\Psi|\lambda, \mu) \ell(\psi'_b|\lambda, \mu)} \times \frac{1/n_b(t_b)}{1/n'_b(t_b)} \times \frac{(1-\rho_b)^{n'_b(0)-1}}{(1-\rho_b)^{n_b(0)-1}} \right\},$$

where the $1/n_b(t_b)$ factor comes from the proposal density of randomly and uniformly choosing one of the
 extant lineages at time t_b as the one sampled in the reconstructed tree. Because the full tree likelihood for
 the proposal only differs on the specific branch, the acceptance ratio simplifies in terminal branches to

$$a = \min \left\{ 1, \frac{\ell(\psi'_b|\lambda, \mu) \ell(\psi_b|\lambda, \mu)}{\ell(\psi_b|\lambda, \mu) \ell(\psi'_b|\lambda, \mu)} \times \frac{n'_b(0)\rho_b(1-\rho_b)^{n'_b(0)-1}}{n_b(0)\rho_b(1-\rho_b)^{n_b(0)-1}} \right\} = \min \left\{ 1, \frac{n'_b(0)}{n_b(0)} (1-\rho_b)^{n'_b(0)-n_b(0)} \right\} \quad (1)$$

and in internal branches to

$$a = \min \left\{ 1, \frac{\lambda \ell(\psi'_b|\lambda, \mu) \ell(\psi_b|\lambda, \mu) 1/n_b(t_b) (1-\rho_b)^{n'_b(0)-n_b(0)}}{\lambda \ell(\psi_b|\lambda, \mu) \ell(\psi'_b|\lambda, \mu) 1/n_b(t_b)'} \right\} = \min \left\{ 1, \frac{n_b(t_b)'}{n_b(t_b)} (1-\rho_b)^{n'_b(0)-n_b(0)} \right\}. \quad (2)$$

148 The acceptance probability for the sampling fraction is common to all further models, so, for simplicity, we
 149 suppress it in the rest of the manuscript.

150 **Conditioning on survival** To condition on survival of the process we use the pseudo-marginal principle
 151 from Andrieu and Roberts (2009). We describe here the approach in a general manner since it applies to all
 152 models that involve extinction. Assuming crown conditioning, that is, that both crown lineages survive to
 153 the present and letting $\ell(\Psi|\theta)$ be the likelihood of the tree and $p(\theta)$ the prior for parameters θ we target the
 154 unnormalized density

$$q(\theta) = \frac{\ell(\Psi|\theta)p(\theta)}{(1-S(\theta))^2}$$

155 where $S(\theta)$ is the probability that a lineage at T_Ψ survives until the present.

Following Ronquist et al. (2021),

$$z(\theta) = \frac{1}{(1-S(\theta))^2} = \sum_{K=0}^{\infty} K g(K|\theta)$$

156 where $g(K|\theta) = (1-S(\theta))^{K-1} S(\theta)^2$. Thus, $z(\theta)$ is the expectation of a geometric distribution of parameter
 157 $S(\theta)^2$. We can then sample K from $g(K|\theta)$ (for which we do not have an analytical solution: see Appendix),
 158 by simulating two lineages under θ starting at T_Ψ and counting the number of attempts until both survive.

159 To sample from density $q(\theta)$ we perform MCMC on the pair of variables (θ, K) targeting instead

$$q_2(\theta, K) = \ell(\Psi|\theta)p(\theta)g(K|\theta)K$$

160 marginally obtaining our target density

$$\sum_{K=0}^{\infty} q_2(\theta, K) = \ell(\Psi|\theta)p(\theta)z(\theta) = q(\theta).$$

161 Effectively, our algorithm proposes a new value for θ , $\theta' \sim s(\theta, d\theta')$, and then, given θ' , proposes K' by
162 drawing from $K' \sim g(K'|\theta')$ and accept this with probability $a = \min\{1, r\}$, where

$$r = \frac{s(\theta', d\theta)\ell(\Psi|\theta')p(\theta')K'}{s(\theta, d\theta')\ell(\Psi|\theta)p(\theta)K};$$

163 circumventing the need to analytically compute $z(\theta)$ while providing exact MCMC inference. Of note, in
164 this equation for r , $\ell(\Psi|\theta)p(\theta)K$ is an unbiased estimate of $q(\theta)$ in the denominator, and similarly for $q(\theta')$ in
165 the numerator. Replacing the target by an unbiased stochastic estimate of it in the acceptance ratio, while
166 preserving exact MCMC sampling, is the fundamental idea of the pseudo-marginal principle (Andrieu and
167 Roberts, 2009). For stem conditioning, the probability of survival of one lineage until the present, we simply
168 estimate K as the number of times until a single lineage at time T_Ψ survives to the present (*i.e.*, $z(\theta)$ is the
169 expectation of a geometric distribution of parameter $S(\theta)$).

170 **Mixed Gibbs sampling for λ and μ** Given the complete tree, Ψ , λ and μ follow a Poisson distribution
171 and thus one can sample (almost) directly from their full conditional posterior via Gibbs sampling. We
172 use the conjugate Gamma prior, $\Gamma(\kappa, \varsigma)$ for both λ and μ , which results in the following full conditional
173 distribution from which we can sample directly

$$p(\lambda|\cdot) \sim \Gamma(\kappa + s, \varsigma + L),$$

174 and similarly with μ but using the number of extinction events. The conditioning on survival, however,
175 requires adding a MH step, such that the proposed speciation and extinction rates λ' and μ' are accepted
176 with probability $a = \min\{1, K'/K\}$.

177 We validate our data augmentation implementation of the constant rate birth-death by comparing the
178 rate posteriors with the analytical solution from (Nee et al., 1994) implemented in a Bayesian framework in
179 the ‘diversitree’ package (FitzJohn, 2010) for R (R Core Team, 2022) (see Fig S1).

180 Pure-Birth Diffusion

181 We now relax the condition that speciation rates are constant through time and across taxa by rather defining
182 $\lambda_l(t)$ to be the result of a stochastic diffusion process. In other words, we consider the observed phylogenetic
183 tree as the outcome of an unobserved latent process of speciation. For simplicity, we start by assuming
184 that there is no extinction, and call this model the Pure-Birth Diffusion (‘PBD’). Specifically, we assume
185 that speciation rates for lineage l evolve anagenetically following the exponential of a Brownian motion (*i.e.*,
186 Geometric Brownian motion, GBM), such that

$$d\ln(\lambda_l(t)) = \alpha dt + \sigma_\lambda dW(t), \tag{3}$$

187 where α represents the drift and σ_λ the diffusion rate for speciation rates and $W(t)$ is the Wiener process
 188 (*i.e.*, standard BM). This diffusion rate has units $\frac{\text{spp}}{(\text{unit time})^{3/2}}$. Under this model, the drift, α , determines
 189 the median or geometric expectation of the rates, but not its arithmetic mean expectation. That is, if $\alpha = 0$,
 190 after some time half of the GBM processes will be lower than the starting rate and the other half will have
 191 increased. Given the non-negative nature and log-normal expectation, the arithmetic mean of will be larger
 192 than the median, which is why we perform all rate transformations and aggregations using the geometric
 193 mean (see below).

194 At cladogenesis, the current value for speciation rates are inherited identically, that is, for any cladogenic
 195 event at time t_s and an ancestral lineage rate of $\lambda_a(t_s)$, then $\lambda_{d_1}(t_s) = \lambda_{d_2}(t_s) = \lambda_a(t_s)$, where d_1 and d_2
 196 represent each of the daughter lineages. At the start of the tree (*i.e.*, time T_Ψ) $\lambda_i(T_\Psi) = \lambda_r$.

197 Note that adding the drift parameter α is important in some contexts to restrain the “run-away species
 198 selection” that our model and others with inherited speciation rates (Beaulieu and O’Meara, 2015; Maliet
 199 et al., 2019), produce. Specifically, lineages with higher speciation rates will generate daughters with a higher
 200 expected rate, which will themselves generate even more daughters, and so on. Including a drift parameter
 201 and/or extinction constraints allows avoiding such a run-away.

202 **Data augmentation** We use a data augmentation approach to approximate the likelihood of rates that
 203 follow GBM diffusion (Horvillour and Lartillot, 2014; Quintero and Landis, 2019). Namely, we generate
 204 unobserved stochastic paths of GBM for all lineages across the tree to approximate the likelihood of a
 205 phylogenetic tree generated under Eq. 3. We determine a small time step, δ , such that $\delta_i = t_{i+1} - t_i$, where
 206 $t_i < t_{i+1}$, and divide each branch of the tree into m small time steps such that $m = \lfloor t_b/\delta \rfloor + 1$, where t_b
 207 represents the edge length of branch b . Note that the last time step for branch b is always smaller than δ
 208 assuming that the probability that t_b is a multiple of δt is 0. Thus, for any branch b , we sample the data
 209 augmented diffusion process at times $t = \{t_1 = 0, t_2 = t_1 + \delta, \dots, t_{m+1} = t_b\}$, obtaining the stochastic process
 210 $\Lambda_b = \{\lambda_b(t_1), \dots, \lambda_b(t_b)\}$ sampled in a discrete time grid. For clarity and conciseness we denote $\lambda_i(t_i)$ as λ_i
 211 from now on.

212 **Likelihood** For any time step $[t_i, t_{i+1}]$, we sample λ_i and λ_{i+1} at the endpoints. Given a sufficiently small
 213 time step, δ_i , a good approximation for the likelihood of no event happening during time $[t_i, t_{i+1}]$ is:

$$p(\text{no event in } [t_i, t_{i+1}] | \lambda_i, \lambda_{i+1}, \sigma_\lambda, \mathcal{M}_{\text{PBD}}) = \exp\{-\bar{\lambda}_{i,i+1}\delta_i\} \times \frac{1}{\sigma_\lambda \sqrt{2\pi\delta_i}} \exp\left\{-\frac{(\Delta\lambda_i - \alpha\delta_i)^2}{2\delta_i\sigma_\lambda^2}\right\}$$

214 where $\bar{\lambda}_{i,i+1}$ represents the geometric mean for $\{\lambda_i, \lambda_{i+1}\}$, $\Delta\lambda_i = \ln(\lambda_{i+1}) - \ln(\lambda(t_i))$, and \mathcal{M}_{PBD} represents
 215 the PBD model given by Eq. 3. The first part of the equation is the probability that there is no speciation
 216 events during δ_i and the second part is the probability that the GBM diffused from λ_i to λ_{i+1} . For internal
 217 branches we then simply multiply the speciation events likelihoods. Given the data augmented stochastic
 218 diffusion for speciation rates across the tree, Λ , the likelihood for the full tree under the Yule Diffusion
 219 process can be then straightforwardly approximated by

$$p(\Psi | \Lambda, \lambda_r, \sigma_\lambda, \mathcal{M}_{\text{PBD}}) = \prod_{b \in \Psi_I} \lambda(t_b) \prod_b \prod_{i=1}^m p(\text{no event in } [t_i, t_{i+1}] | \lambda_i, \lambda_{i+1}, \sigma_\lambda, \mathcal{M}_{\text{PBD}}), \quad (4)$$

220 where Ψ_I are the set of internal branches. Note that topology matters, and for an ancestral lineage a with
 221 daughter branches d_1 and d_2 , we have that $\lambda_a(t_b) = \lambda_{d_1}(t_1) = \lambda_{d_2}(t_1)$.

222 **Diffusion updates** We integrate over possible rate diffusion histories, Λ , using Metropolis-Hastings diffu-
 223 sion path updates. We use a path diffusion update based on internal nodes that are connected to a ‘triad’
 224 of branches in Ψ , that is, given a parent and its daughter branches, pr , d_1 and d_2 , respectively. We con-
 225 sider the following four cases to update the triad diffusion paths Λ_{pr,d_1,d_2} : *i*) If all branches are internal
 226 and the parent is not the root, we make a GBM proposal for the node conditioned on the end points (i.e.,
 227 $\lambda_{pr}(t_1), \lambda_{d_1}(t_b), \lambda_{d_2}(t_b)$) and respective branch lengths. We then propose diffusion paths using Brownian
 228 bridges for each branch with the new node value as an endpoint. *ii*) If all branches are internal and the par-
 229 ent is the root, we make a GBM node proposal conditioned on the daughter end points, from which we then
 230 backwardly propose a new root value λ_r . We then use Brownian bridges to sample their respective diffusion
 231 paths. *iii*) If one of the daughter branches is terminal, we make a GBM node proposal conditioned on the
 232 endpoints, from which we propose a new GBM value for the endpoint (i.e., tip) of the terminal branch. We
 233 then use Brownian bridges to sample their respective diffusion paths. *iv*) If both of the daughter branches
 234 are terminal, we simply make a GBM node proposal given the parent branch, and then forwardly simulate
 235 both terminal daughter diffusion paths. Note that the drift term, α , cancels out when proposing Brownian
 236 bridges with drift, as it is fully determined by the endpoints and time elapsed.

237 Then, the acceptance probability, a , for a new diffusion path proposal, Λ' is

$$a = \min \left\{ 1, \frac{p(\Lambda'|\Psi, \alpha, \sigma_\lambda, \mathcal{M}_{\text{PBD}})}{p(\Lambda|\Psi, \alpha, \sigma_\lambda, \mathcal{M}_{\text{PBD}})} \times \frac{p(\Lambda|\alpha, \sigma_\lambda, \mathcal{M}_{\text{GBM}})}{p(\Lambda'|\alpha, \sigma_\lambda, \mathcal{M}_{\text{GBM}})} \right\},$$

238 where \mathcal{M}_{GBM} represents the GBM model. Note that this ratio simplifies because the Brownian motion part
 239 in the likelihood cancels out with the proposal probability. We specify a uniform prior on the the speciation
 240 rates at the root, λ_r .

241 **α and σ_λ updates** We use a Normal conjugate prior, $p(\alpha) \sim \text{N}(\nu, \tau)$ to directly sample from the full
 242 conditional posterior distribution (see Appendix for derivation):

$$p(\alpha|\Lambda, \sigma_\lambda, \nu, \tau, \mathcal{M}_{\text{PBD}}) \sim \text{N} \left(\frac{(\frac{\sigma_\lambda}{\tau})^2 \nu + \sum_b (\lambda_{m+1} - \lambda_1)}{(\frac{\sigma_\lambda}{\tau})^2 + L}, \frac{\sigma_\lambda^2}{(\frac{\sigma_\lambda}{\tau})^2 + L} \right), \quad (5)$$

243 where L is the tree length. Similarly, we specify the Inverse Gamma conjugate prior $p(\sigma_\lambda^2) \sim \Gamma^{-1}(\kappa, \varsigma)$, and
 244 we obtain the following posterior conditional distribution (see Appendix for derivation):

$$p(\sigma_\lambda^2|\Lambda, \alpha, \kappa, \varsigma, \mathcal{M}_{\text{PBD}}) \sim \Gamma^{-1} \left(\kappa + \frac{N}{2}, \varsigma + \sum_b \sum_{i=1}^m \frac{(\Delta\lambda_i - \alpha\delta_i)^2}{2\delta_i} \right), \quad (6)$$

245 where $N = \sum_b \sum_{i=1}^m 1$.

246 While the Inverse Gamma prior for σ_λ is advantageous by allowing full Gibbs sampling, substantially
 247 enhancing the properties of our MCMC approach, it has some drawbacks, particularly with small phylogenetic
 248 trees. As expected, small datasets (e.g., 50 tip trees) are particularly affected by the choice of prior: a
 249 $\Gamma^{-1}(0.05, 0.05)$, for instance, will enforce very low probabilities to small values (e.g., $\sigma_\lambda < 0.1$), overestimating

250 the diffusion rate. Conversely, a $\Gamma^{-1}(0.1, 0.001)$ prior, for instance, will capture small values of σ_λ but will
 251 underestimate higher diffusion rates (Fig S2). While the prior choice becomes asymptotically irrelevant as
 252 more data is included, we emphasize that, for inference, some rationale should be taken for the choice of
 253 prior (Fig S2).

254 Birth-Death Diffusion

255 We now relax the condition that extinction rates are non-existent and rather define three diffusion models
 256 that incorporate extinction: *i*) constant extinction, $\mu(t) = \mu$ ('CED'), *ii*) constant turnover, $\mu(t) = \epsilon\lambda(t)$
 257 ('CTD'), and *iii*) extinction also being the result of a stochastic diffusion process, $\mu(t)$, ('BDD'). For
 258 conciseness, we now describe only the latter, the Birth-Death Diffusion (BDD), which is the most general
 259 among these models, with straightforward simplifications towards the CED and CTD. Thus, we assume
 260 that extinction rates for lineage l evolve anagenetically following the exponential of a Brownian Motion (i.e.,
 261 Geometric Brownian motion, GBM) with no drift, such that

$$d\ln(\mu_l(t)) = \sigma_\mu dW(t), \quad (7)$$

262 where σ_μ represents the diffusion rate for extinction rates and $W(t)$ is the Wiener process (i.e., standard
 263 BM). At cladogenesis, the current value for extinction rates is inherited identically, in the same manner as
 264 are speciation rates. At the start of the tree, with time T_Ψ , $\lambda_l(T_\Psi) = \lambda_r$ and $\mu_l(T_\Psi) = \mu_r$. Thus, the model
 265 has three hyper-parameters: α , σ_λ , and σ_μ . As we did for λ_i , we denote $\mu_l(t_i)$ as μ_i for conciseness.

266 **Likelihood** In the same way of the PBD process, for any time step $[t_i, t_{i+1}]$, we sample λ_i , λ_{i+1} , μ_i and
 267 μ_{i+1} at the endpoints. Given a sufficiently small time step, δ_i , a good approximation for the likelihood of
 268 no events during time $[t_i, t_{i+1}]$ is:

$$p(\text{no event in } [t_i, t_{i+1}] | \lambda_i, \lambda_{i+1}, \mu_i, \mu_{i+1}, \alpha, \sigma_\lambda, \sigma_\mu, \mathcal{M}_{\text{BDD}}) = \\ \exp\{-\bar{\lambda}_{i,i+1} + \bar{\mu}_{i,i+1}\delta_i\} \times \frac{1}{\sigma_\lambda \sqrt{2\pi\delta_i}} \exp\left\{-\frac{(\Delta\lambda_i - \alpha\delta_i)^2}{2\delta_i\sigma_\lambda^2}\right\} \times \frac{1}{\sigma_\mu \sqrt{2\pi\delta_i}} \exp\left\{-\frac{(\Delta\mu_i)^2}{2\delta_i\sigma_\mu^2}\right\}$$

269 where \mathcal{M}_{BDD} represents the Birth-Death Diffusion model given by Eq. 7. The first part of the equation is
 270 the probability that no speciation or extinction events happen during δ_i and the second part is the probability
 271 of the speciation and extinction GBM diffusing from λ_i to λ_{i+1} and μ_i to μ_{i+1} , respectively. Given the data
 272 augmented stochastic diffusion for speciation and extinction rates across the tree, Λ and M , respectively, the
 273 likelihood for the full tree under the BDD process can be then straightforwardly approximated by

$$p(\Psi | \Lambda, M, \lambda_r, \mu_r, \alpha, \sigma_\lambda, \sigma_\mu, \mathcal{M}_{\text{BDD}}) = \\ \prod_{b \in \Psi_I} \lambda_{t_b} \prod_{b \in \Psi_{T_\mu}} \mu_{t_b} \prod_b \prod_{i=1}^m p(\text{no event in } [t_i, t_{i+1}] | \lambda_i, \lambda_{i+1}, \mu_i, \mu_{i+1}, \sigma_\lambda, \sigma_\mu, \mathcal{M}_{\text{BDD}}), \quad (8)$$

274 where Ψ_{T_μ} are the set of terminal extinct branches.

275 **Data augmentation and parameter updates** To augment the reconstructed tree and obtain complete
 276 trees, we use forward simulation, as described above. For internal branches, new complications arise given
 277 the underlying GBM in speciation and extinction. Specifically, once the tree is simulated along a branch
 278 and one of the tips selected at random as the one that leads to the observed speciation event, both the
 279 speciation rate instantaneously before speciation $\lambda(t_{pr})'$ and the extinction rate $\mu(t_{pr})'$ do not correspond
 280 to the initial speciation and extinction rate in the daughter lineages, $\lambda_{d1}(t_1)$ and $\lambda_{d2}(t_1)$ and $\mu_{d1}(t_1)$ and
 281 $\mu_{d2}(t_1)$. Thus, we make Brownian bridge proposals on the daughter branches to match the new proposed
 282 rates at speciation, $\lambda(t_{pr})'$ and $\mu(t_{pr})'$. Figure 1 illustrates our forward simulation proposals. Let Ψ, Λ, M
 283 represent the topology, the latent speciation GBM and the latent extinction GBM, respectively, then the
 284 acceptance ratio for this proposal is

$$a = \min \left\{ 1, \frac{\ell(\Psi', \Lambda', M' | \alpha, \sigma_\lambda, \sigma_\mu, \mathcal{M}_{\text{BDD}}) p(\lambda(t_{pr}) | \lambda(t_{d1}), \lambda(t_{d2})) p(\mu(t_{pr}) | \mu(t_{d1}), \mu(t_{d2}))}{\ell(\Psi, \Lambda, M | \alpha, \sigma_\lambda, \sigma_\mu, \mathcal{M}_{\text{BDD}}) p(\lambda(t_{pr})' | \lambda(t_{d1}), \lambda(t_{d2})) p(\mu(t_{pr})' | \mu(t_{d1}), \mu(t_{d2}))} \times \frac{p(\Lambda | \alpha, \sigma_\lambda, \mathcal{M}_{\text{GBM}}) p(M | \sigma_\mu, \mathcal{M}_{\text{GBM}}) \ell(\psi | \alpha, \sigma_\lambda, \sigma_\mu, \mathcal{M}_{\text{BDD}})}{p(\Lambda' | \alpha, \sigma_\lambda, \mathcal{M}_{\text{GBM}}) p(M' | \sigma_\mu, \mathcal{M}_{\text{GBM}}) \ell(\psi' | \alpha, \sigma_\lambda, \sigma_\mu, \mathcal{M}_{\text{BDD}})} \right\},$$

which simplifies to

$$a = \min \left\{ 1, \frac{\lambda(t_{pr})' p(\lambda(t_{pr}) | \lambda(t_{d1}), \lambda(t_{d2})) p(\mu(t_{pr}) | \mu(t_{d1}), \mu(t_{d2}))}{\lambda(t_{pr}) p(\lambda(t_{pr})' | \lambda(t_{d1}), \lambda(t_{d2})) p(\mu(t_{pr})' | \mu(t_{d1}), \mu(t_{d2}))} \times \frac{\ell(\Lambda'_{d1,d2}, M'_{d1,d2} | \Psi_{d1,d2}, \alpha, \sigma_\lambda, \sigma_\mu, \mathcal{M}_{\text{BDD}})}{\ell(\Lambda_{d1,d2}, M_{d1,d2} | \Psi_{d1,d2}, \alpha, \sigma_\lambda, \sigma_\mu, \mathcal{M}_{\text{BDD}})} \right\},$$

285 where

$$p(\lambda(t_{pr}) | \lambda(t_{d1}), \lambda(t_{d2})) = \lambda(t_{pr}) \sim N \left(\frac{t_{d2} \lambda(t_{d1}) + t_{d1} \lambda(t_{d2})}{t_{d1} + t_{d2}}, \frac{t_{d1} t_{d2}}{t_{d1} + t_{d2}} \sigma_\lambda^2 \right),$$

286 and similarly for $p(\mu(t_{pr}) | \mu(t_{d1}), \mu(t_{d2}))$. If the branch is internal we add the factor $n_b(t_b)' / n_b(t_b)$ as in Eq.
 287 2. In practice, to make proposals more efficient, we sample the tip that leads to the observed speciation event
 288 proportional to the probability that it's rates would yield the currently observed rates at the daughters.

289 We use the Eq. 5 to sample α , use Eq. 6 to sample σ_λ , and to sample σ_μ we use

$$p(\sigma_\mu^2 | M, \kappa, \varsigma, \mathcal{M}_{\text{BDD}}) \sim \Gamma^{-1} \left(\kappa + \frac{N}{2}, \varsigma + \sum_b \sum_{i=1}^m \frac{(\Delta \mu_i)^2}{2 \delta_i} \right), \quad (9)$$

290 where M represents the full diffusion of $\mu(t)$ and $N = \sum_b \sum_{i=1}^m 1$. For λ_r and μ_r we use Uniform priors
 291 of $(0, 100)$. We note that other sort of priors, particularly on extinction, can severely affect the posterior
 292 distribution when there is not sufficient information in the tree. For the CED and CTD models, we sample
 293 μ using Gibbs sampling as in the CBD model and ϵ using standard MH updates and specify a Uniform prior
 294 of $(0, 100)$.

295 Model Behavior

296 **Simulations** We use simulations to explore behavior for each of the four different model assumptions:
 297 PBD, CED, CTD, and BDD. To simulate under a diffusion model we take advantage of the following

298 approximation. Let V be a random variable for the time of an event and let λ_i and λ_{i+1} , be the event rate
299 at time t_i and t_{i+1} , respectively, where $t_{i+1} - t_i = \delta \geq 0$, then we have

$$\Pr(t_i \leq V < t_{i+1} \mid t_i \leq V) \approx \bar{\lambda}_{i,i+1} \delta.$$

300 For each model we simulated 100 trees with 50, 100 and 200 tips by sampling from a range of parameter
301 space (see Fig 2). We first made simulations and inference without $\alpha = 0$, and then simulations with a range
302 of values of α around 0 (Fig 2 & 3). To simulate non-biased samples from the model, we started with two
303 lineages and made sure both of them survived until the present and followed Stadler (2011b) to sample a tree
304 with a determined number of extant species. In total we have 100 simulations for each of the 4 models, each
305 with 3 different tree sizes, with and without α , yielding 2400 total simulations. We ran MCMC inference
306 on each simulation for 10^6 iterations, logging every 10^3 iteration, and discarding the first 2×10^5 samples
307 as burn-in. These guaranteed Effective Sample Sizes of at least 300 across all parameters. We used weakly
308 informative priors for most parameters. Specifically, we used a uniform $U(0, 100)$ for $\ln(\lambda_r)$, $\ln(\mu_r)$ and ϵ ,
309 an Inverse Gamma $\Gamma^{-1}(0.5, 0.1)$ for σ_λ , a Gaussian $N(0, 10)$ for α , and a Gamma $\Gamma(1, 1)$ (*i.e.*, $\text{Exp}(1)$) for
310 μ . Initial simulations showed that for trees with less than 100 tips, σ_μ would largely follow the prior, which,
311 given the high variance of our prior, would yield very bad mixing and numerical inaccuracies. Thus, we used
312 the stronger prior for σ_μ of $\Gamma^{-1}(5, 0.5)$, which concentrates more density on parameter values closer to 0.

313 **Statistical coverage and accuracy** Figure 2 & 3 show the simulation results without and with the drift
314 parameter α , respectively. Across all scenarios we have a median coverage of the speciation rates along each
315 tree above *ca* 95% (a coverage of 95% for one simulation indicates that 95% of the true speciation rates
316 across the phylogenetic tree are within the 95% Highest Posterior Density (HPD) intervals of the estimated
317 speciation rates). Furthermore, the median average relative error (*i.e.*, $\ln(\lambda_{true}(t_i)) / \ln(\lambda_{inferred}(t_i)) - 1$)
318 for the diffusion of speciation rates is close to 0 across most simulations, supporting no strong bias in the
319 model's estimates (Fig 2 & 3). Nonetheless, for CTD we find a small downward bias (*i.e.*, median of 0.2
320 relative log units) in speciation rates. Finally, the diffusion coefficient for speciation σ_λ shows adequate
321 accuracy and coverage across all simulations (Fig 2 & 3), taking into account the influence of its prior for
322 smaller trees.

323 As expected, the influence of the prior is proportional to the amount of data (*i.e.*, tree size): Supple-
324 mentary Figure 2 shows the influence of two different Inverse Gamma priors using the same data on the
325 posterior distribution of σ_λ and compares coverage and accuracy across simulations with increasing tree size.
326 Overall, σ_λ is overestimated in values close to 0 of parameter space (*i.e.*, $\sigma_\lambda < 0.1$) when using an Inverse
327 Gamma prior of $\Gamma^{-1}(0.5, 0.1)$. In this range of parameter space, using an Inverse Gamma prior with higher
328 density towards 0 (*e.g.*, $\Gamma^{-1}(0.5, 0.1)$) can increase coverage, yet lead to some underestimation when rates
329 are very high. While, as with all Bayesian analyses, the choice of prior is important and should be carefully
330 considered, the underlying rate estimates do not suffer in statistical coverage nor accuracy, and the prior
331 influence dwindles with increasing tree size (Fig S2).

332 Without drift, extinction rates for constant-extinction ($\mu(t) = \mu$) seem to be underestimated for trees
333 with 50 tips, but gradually improve with trees of 100 and 200 tips; still, the 95% HPD coverage across

334 these simulations is of 95% (Fig 2b). For constant turnover ($\mu(t) = \epsilon\lambda(t)$), ϵ estimates are become more
335 accurate and coverage is better as one increases tree size. For the BDD ($\mu(t)$), extinction rates have very
336 poor coverage and a underestimation bias, suggesting that, for extant-only small trees (at least to 200 tips),
337 extinction diffusion dynamics seem not to be recoverable.

338 With drift, we get good estimates and coverage of the drift, α , and the diffusion, σ_λ , parameters (Fig
339 3) across diffusion models (again taking into account the influence of the prior on the diffusion coefficient
340 for small tree sizes; Fig S2). For CTD, we find good estimates of turnover rates, ϵ , with good accuracy
341 increasing with tree size. For CED, however, accuracy is low and, at least for the range of tree sizes used,
342 does not seem to improve with more data (at least up to trees with 200 tips). HPD coverage of μ remains
343 very high since the posterior distribution has high variance. For the BDD model, median extinction diffusion
344 coverage is very low and shows a strong downward bias, again suggesting that either more data is needed
345 or an impossibility to appropriately recover diffusing extinction rates. Note, however, this does not lead to
346 non-identifiability, but rather to biased estimates in extinction rates. While these results reinforce known
347 difficulties of retrieving extinction rates from extant-only trees, particularly in our model with lineage specific
348 speciation diffusion with drift, speciation rates continue displaying good statistical behavior.

349 **Results summaries** On top of returning posterior samples for all hyperparameters, our model returns a
350 posterior sample of data augmented trees, each with unobserved lineages that went extinct and their latent
351 speciation, and in the case of BDD, extinction, rates. We summarize rate patterns in two ways. First,
352 to provide posterior speciation and extinction (only in BDD) rate distributions across the tree, we remove
353 all unobserved (data augmented) branches from the data augmented trees, and then estimate the posterior
354 distribution for each λ_i in the reconstructed tree. Note that we can only estimate a posterior distribution
355 of rates along the branches of the reconstructed tree, since this is the only part that remains fixed across
356 the whole MCMC run. For clarity, we dub these estimates “posterior reconstructed rates”, which we can
357 summarize at any point along the phylogenetic tree, or take their geometric mean across contemporary
358 lineages through time.

359 Second, we estimate average rates through time taking into account all data augmented lineages. Here, we
360 take each data augmented tree sample and estimate the cross-lineage geometric mean of their rates through
361 time. We call these rates “posterior DA rates”.

362 **Temporal patterns** In simulations of our diffusion models, the drift parameter controls whether the me-
363 dian speciation rates increase or decrease through time. Nonetheless, we wanted to further test the flexibility
364 of our diffusion models in capturing temporal heterogeneity that is neither linear nor constant, making our
365 models more useful when applied to empirical data. Thus, we performed a simulation scenario in which spe-
366 ciation rates were low, suddenly increased and then decreased, while maintaining branch heterogeneity, and
367 with some extinction diffusion, as shown in Figure 4a. We then performed inference on this simulation using
368 PBD (Fig 4b), CED (Fig 4c), CTD (Fig 4b), and BDD (Fig 4e,f). Remarkably, we find that all diffusion
369 models are able to capture this temporal fluctuation in speciation rates, independent of our assumption on
370 extinction rates.

371 Implementation

372 We implemented all these diversification methods in the ‘Tapestry’ package for Julia (Bezanson et al.,
373 2017), available at <https://github.com/ignacioq/Tapestry.jl>. The software documentation details the
374 simulation, inference and data structures implemented in the package as well as its summary and plotting
375 capabilities.

376 Mammal Birth-Death Diffusion

377 We characterize the across-lineage temporal diversification of mammals using our new diffusion models.
378 We used the recent time-calibrated phylogenetic tree integrating phylogenomic data from Álvarez-Carretero
379 et al. (2021), which comprises 4705 genetically represented extant lineages. Because many of these tips are
380 still considered subspecies, we used only recognized species as tips for taxonomic consistency in the pat-
381 terns observed, yielding a tree with 4071 species. We estimated clade-specific sampling fractions for each
382 of 16 subclades we identified and which broadly correspond to those divisions used to infer the full mam-
383 mal phylogenetic tree in Álvarez-Carretero et al. (2021). For this, we matched current mammal diversity
384 following the Mammal Diversity Database (‘MDD’; 2022), which at the time of the analyses (May 2022)
385 recognized 6457 total extant mammal species. We matched the species with the taxonomy in the MDD and
386 estimated how many species were not represented in the tree to estimate species specific sampling fractions
387 to be passed to the diffusion models (subclade specific sampling fractions are shown in Table S1). We ran
388 3 MCMC chains on the Maximum Clade Credibility (MCC) mammal phylogeny for 10^6 iterations, saving
389 every 10^3 and discarding an additional 3×10^5 as burn-in, for each of the diversification models: PBD, CED,
390 CTD and BDD, both with and without the drift parameter. Finally, we added a constant turnover diffusion
391 model fixing turnover to 1 (equal rates of speciation and extinction at any given time) following empirical
392 paleontological evidence (Marshall, 2017).

393 As expected, we find substantial variation across surviving mammals lineages in speciation and extinction
394 rates (Fig 5 & 6), with a median posterior diffusion coefficient for speciation rates of $\sigma_\lambda = 0.117 \text{ spp/My}^{3/2}$
395 and of extinction rates of $\sigma_\mu = 0.067 \text{ spp/My}^{3/2}$. Indeed, median posterior reconstructed rates of speciation
396 range from about 0.01 spp/My and of extinction of up to about 0.025 spp/My. We find that posterior
397 reconstructed and DA speciation rates were stable or even slightly decreasing since the origin of extant
398 mammals throughout the Jurassic and most of the Cretaceous, but that, during the late Cretaceous (*ca.* 80
399 Mya), there was a dramatic increase in speciation rates (Fig 5a,c). Thereafter, speciation rates remained
400 stable until recently where a final surge in speciation rates is evinced, mostly driven by the recent fast
401 diversification of rodents. These patterns of speciation rates were congruent across diffusion models, *i.e.*,
402 PBD (Fig S3), CED (Fig S4), CTD (Fig S5) and CTD with turnover fixed to 1 (Fig S6). Finally, diversity
403 curves show different patterns across the different diffusion models (Fig 5 & Fig S3-S5), yet, regardless of
404 the assumption in extinction, we find an initial slow accumulation of diversity with a sharp increase in the
405 Late Cretaceous mirroring the temporal pattern of speciation rates.

406 Discussion

407 We developed and implemented new flexible birth-death diffusion models, which allow fast estimation of
408 speciation and extinction rates that vary instantaneously and continuously in time and across lineages. Our
409 model provides increased resolution than other approaches assuming a few constant regime shifts across the
410 tree (Rabosky, 2014; Höhna et al., 2019; Barido-Sottani et al., 2020; Vasconcelos et al., 2022), and even
411 ‘ClaDS’, which assumes constancy in diversification rates within a branch (Maliot et al., 2019). Indeed, for
412 any lineage, for any instantaneous point in time, our model returns a posterior distribution of speciation
413 (and extinction) rates. Furthermore, we present four different models with different assumptions about
414 extinction, three of which are analogous to ClaDS (*i.e.*, ‘pure-birth’, ‘constant-extinction’ and ‘constant-
415 turnover’), and a more flexible one that assumes extinction to also vary according to a GBM, the ‘BDD’
416 model. As with the data augmented implementation of ClaDS (Maliot and Morlon, 2021), our inference
417 yields a set of data augmented posterior trees, that is, probable histories given the model and data, which
418 can be used to estimate diversity through time and average rates (*e.g.*, as we did for mammals in Fig 5 &
419 6), avoiding biases for not taking into account the dynamics of unobserved lineages on these estimates. We
420 also generalize the data augmentation scheme for phylogenetic birth-death processes and show that it can be
421 applied to other models of diversification for which the likelihood has no analytical solution or its calculation
422 is computationally costly.

423 Our diffusion models assume that diversification rate variation is heritable, accumulates in a gradual
424 fashion and in proportion to time. This is consistent with the idea that the interdependence of species
425 traits with their environment at any given moment in time will largely govern their diversification rates.
426 Heritable traits such as body size, dispersal and generation time, among others, are posited to influence
427 rates of speciation and extinction, and result from a gradual, heritable evolutionary process (Heard, 1996).
428 Further evidence of the heritability of diversification rates is found in the observed imbalance across empirical
429 phylogenetic trees (Heard, 1996), as expected from speciation and extinction rates that evolve through
430 time and are transferred into daughter species. However, non-heritable variation, such as environmental
431 oscillations, are also thought of influencing diversification rates (Heard, 1996; Barnosky, 2001; Benton, 2009).
432 We show that, while not explicitly accounted for, our diffusion models are able to detect overarching temporal
433 trends while accounting for fine-grained lineage heterogeneity.

434 All diffusion models support the ‘early-rise’ hypothesis for the clades that led to the present-day richness of
435 mammals (Bininda-Emonds et al., 2007; Meredith et al., 2011; Springer et al., 2019; Grossnickle et al., 2019).
436 Specifically, we find that the origin of the major clades leading to extant mammal diversity results from a
437 sharp average increase across contemporary lineages in speciation rates before the K-Pg extinction event (Fig
438 5). This refutes the explosive-rise hypothesis wherein mammals were suppressed in their ecomorphological
439 and taxonomic diversity during the Cretaceous and experienced a release after the K-Pg extinction event
440 (Archibald and Deutschman, 2001). Our results are concordant with findings using previous phylogenetic
441 evidence from (Meredith et al., 2011), and are supported by other lines of recent paleontological evidence
442 showing that a major ecomorphological diversification in mammal lineages following the rise of angiosperms
443 occurred 10-20 My before the K-Pg (Wilson et al., 2012; Grossnickle and Newham, 2016; Halliday and

444 Goswami, 2016; Grossnickle et al., 2019; Upham et al., 2021). Specifically, fossil dental and body size analyses
445 reveal an expansion into herbivorous and carnivorous diets and increased body size disparity concomitant
446 with taxonomic diversification (Grossnickle and Newham, 2016; Wilson, 2014).

447 Before the diversification burst of the Late Cretaceous, our diffusion models that include extinction
448 exhibits low speciation rates coupled with high extinction rates rates, leading to a slow initial accumulation
449 of mammal diversity (Fig 5c-g). This pattern is also mirrored in paleontological evidence from the Cretaceous
450 Terrestrial Revolution that resulted in substantial lineage turnover (Luo, 2007; Grossnickle and Polly, 2013;
451 Benson et al., 2013, 2016; Grossnickle et al., 2019). While we do not find support for a delayed-rise only
452 scenario of mammals, diversification rates remained high after the K-Pg during most of the Paleogene with
453 a sustained increase in diversity (Fig 5c-g). This findings contrast with the results from (Meredith et al.,
454 2011), wherein speciation rates increase at *ca.* 83 Mya but decrease sharply again at *ca.* 78 Mya, and agrees
455 with paleontological evidence of continued ecological and taxonomic diversification (O’Leary et al., 2013;
456 Grossnickle et al., 2019). Our results also show no effect of the K-Pg extinction event on diversification rates
457 of the surviving mammals lineages but instead show a continued acceleration until the Eocene, concomitant
458 with the radiation of many of the crown group members of extant placental orders (Grossnickle et al., 2019).
459 Regardless of the assumption in extinction rates (*i.e.*, no extinction, constant extinction, constant turnover,
460 constant turnover of 1 or extinction following a diffusion process), our results suggest that the K-Pg did
461 not drive the explosive radiation of present-day mammals (Fig 5,S3-S6). Instead, while the accelerated
462 diversification of surviving mammals started before the K-Pg, the aftermath of this extinction event allowed
463 most lineages to maintain comparable (or even higher) levels of high rates of diversification up to the present.

464 Underlying these overarching processes, our model reveals substantial lineage and time heterogeneity of
465 diversification rates across the mammal tree (Fig 6). To illustrate, we find posterior median lineage speciation
466 rates that range from almost 0 spp/My in Monotremes, up to more than 0.2 spp/My in some Rodents.
467 Concomitant to their high extant diversity, mouse-related Rodents (excluding Ctenohystrica and Sciuridae
468 and related clades) exhibit a dramatic surge in their speciation rates from *ca.* the start of the Miocene (Fig 6).
469 Multiple hypotheses, such as developing hypsodonty, colonization of South America, environmental changes
470 and extinction of competitors, have been proposed to explain this evolutionary radiation (Fabre et al., 2012).
471 Speciation rates in Primates also follow a post-Oligocene increase, yet this was preceded by a sharp decrease
472 after the K-Pg up to the Oligocene, coinciding with a terrestrial fauna turnover event from glaciation and
473 cooling (Springer et al., 2012). For their part, Marsupials show an increase of speciation rates during the
474 Paleocene, correlated with biogeographic dispersal from South America to Australia through Antarctica
475 (Nilsson et al., 2004), but then remained stable. These examples of clade specific routes to extant diversity
476 demonstrate the strength and flexibility of our birth-death diffusion models in capturing how lineage-specific
477 diversification rates evolve in continuous time.

478 Our model and results assume that the phylogenetic tree used is the ‘true’ tree of extant mammalian
479 evolution, and relies on its specific topology and timing of speciation events. Consequently, we use the most
480 up-to-date phylogenetic time-tree of mammals, built with a novel Bayesian framework that incorporated 72
481 mammalian genomes and thousands of species genetic data in an integrated framework, with substantial less
482 uncertainty in the timing of past cladogenetic events (Álvarez-Carretero et al., 2021). Our diversification

483 results contrast with previous phylogenetic analyses that used different model assumptions and phylogenetic
484 trees. Bininda-Emonds et al. (2007) developed the first species-level phylogenetic tree and estimated changes
485 in the slope in the Lineage-Through-Time (LTT) plot to show a diversification peak at around 90 Mya, fol-
486 lowed by a decrease until the K-Pg, where rates increased again until the very present. Using the same
487 phylogenetic tree, Stadler (2011a) demonstrated that using the LTT slope as means to estimate temporal
488 changes in diversification rates can lead to biased results, and developed a birth-death model where con-
489 temporary lineages experience the same diversification rates (*i.e.*, ‘lineage-homogeneous’), but are allowed
490 to shift at different epochs. Using this model, a peak in mammalian diversification rates around 33 Mya was
491 inferred, followed by stable high rates until near the present, wherein they declined. Meredith et al. (2011)
492 built a family level tree with 169 mammal species and recovered a diversification peak that lasted only during
493 *ca.* 83-78 Mya. Using the same model, but with an updated tree and dating scheme of placental mammals,
494 Liu et al. (2017) found an upward shift around 90 Mya followed by a downward shift around 52 Mya, and
495 no effect of the K-Pg. Finally, using the tree from Upham et al. (2019), Upham et al. (2021) performed
496 temporal diversification analyses under a model of lineage-heterogeneous rates resulting from discrete shifts
497 (Rabosky, 2014) (but see Moore et al. (2016) for concerns on this method) and found a burst in speciation
498 rates in the Late Cretaceous followed by a steady increase towards the present.

499 A more general caveat persists for diversification analyses that are conducted on fixed phylogenetic trees
500 (the great majority), which, in turn, were usually inferred using a dating method that does entail explicit
501 (typically through constant pure-birth or birth-death priors) or implicit assumptions about the underlying
502 diversification process. Ultimately, however, a joint model where co-estimation of divergence times and
503 diversification rates would be more appropriate, and has recently been developed for some birth-death
504 models with discrete shifts (Kühnert et al., 2016; Höhna et al., 2016; Barido-Sottani et al., 2020). The
505 data augmented approaches developed here could be used as priors over divergence times in an integrative
506 approach and is an exciting avenue for future research.

507 As with most diversification models based on extant-taxa alone, our model and results are susceptible
508 to inferential limits, specially as one moves into the deep past. For instance, Louca and Pennell (2020)
509 showed that for time-varying lineage-homogeneous speciation and extinction rates and in the absence of
510 any constrain on the functional form of the time-varying rate functions, there are an infinite number of
511 such functions that return the same likelihood for any extant-only time-tree. Our model overcomes this
512 particular non-identifiability issue by incorporating lineage-heterogeneous rates informed by topology by
513 means of speciation and extinction rates following a GBM diffusion model. More applicable, however, are
514 concerns in reliably estimating extinction rates from extant-only time-trees (Rabosky, 2010; Beaulieu and
515 O’Meara, 2015; Louca and Pennell, 2021). Note that our inference model is similar, yet more complex, to the
516 simulation model in Beaulieu and O’Meara (2015): they simulated trees with branch-constant rates that are
517 inherited with a change proportional to the product of a diffusion rate and the squared root of elapsed time,
518 rectifying the time proportionality of the diffusion rates used by Rabosky (2010), who used the product of
519 time (rather than squared root of time) with the diffusion rate. Using these simulations, Rabosky (2010)
520 and Beaulieu and O’Meara (2015) found that a constant rate birth-death model will overestimate extinction
521 and lead to spurious results if the variance in speciation rates was high. Our model explicitly takes into

522 account this type of heritability and thus is immune to this specific across-lineage heterogeneity that could
523 lead to inaccurate inferences when assuming a simpler model.

524 This is not to say that other sources of variation in diversification could not potentially bias inference.
525 From the different extinction assumptions in our diffusion models, the best behavior is achieved for constant-
526 turnover (CTD) and constant-extinction (CED), with the extinction diffusion model (BDD) only achieving
527 fair accuracy with sufficient data. Nonetheless, we are cautious on over-interpreting what these extinction
528 estimates represent, particularly when faced with empirical data. We emphasize that our results reflect
529 the diversification dynamics of those mammal lineages under a still simple model of diversification, and,
530 using extant-species alone trees could prove insufficient in recovering very high extinction rates in the past
531 (Marshall, 2017). For instance, without fossil information the probability of a mammal clade that diversified
532 greatly yet went extinct, such as the Cimolodontan multituberculates (Grossnickle et al., 2019), is marginal
533 under birth-death models conditioned on surviving taxa and thus is unlikely to be inferred as the main
534 diversification history. Furthermore, our model does not account for episodic mass extinction events, which
535 could substantially enhance the biological realism of our diversification histories, conditional on having a
536 good estimate of the proportion of lineages that went extinct. Our model and implementation renders
537 the future incorporation of fossils into the model straightforward and opens an exciting avenue for future
538 macroevolutionary research as phylogenetic trees that combine neontological and paleontological data become
539 increasingly available.

540 One of the main properties of the diffusion models is that speciation (and extinction in the BDD) vary
541 continuously and stochastically within any lineage at any moment in time. On one hand, this unlocks a
542 new series of possibilities for answering questions about evolutionary radiations. Specifically, it provides
543 the foundation in which to test drivers of diversification rates that also vary instantaneously through time
544 and across lineages, such as concomitant evolution of species intrinsic traits or abiotic factors, including
545 environmental fluctuations, while accounting for independent, stochastic variation. This would be the con-
546 tinuous counterpart to the hidden states framework that considers residual variation in State dependent
547 Speciation and Extinction methods (Beaulieu and O’Meara, 2016), which have proved pivotal in assessing
548 discrete trait drivers of diversification. ‘QuaSSE’ provides a correlation test between diversification and a
549 continuous trait FitzJohn (2010), but, aside of not accounting for residual variation (increasing Type I error),
550 uses a numerical method that would not scale-up well with the number of Brownian processes. Conversely,
551 we have shown with our data augmented framework that this would not be a major limiting factor. More
552 generally, our framework is analogous to a simple linear regression in having both a deterministic effect (*i.e.*,
553 slope) together with unexplained variance. This analogy illustrates the strength of such an approach for
554 it will be contentious to perform inference using only a linear slope to explain the association between two
555 variables. Future advancements in this direction will help in identifying the underlying processes governing
556 macroevolutionary dynamics.

557 With our new model, there are now three different paradigms for modeling lineage-specific variation in
558 diversification rates: rare discrete anagenetic shifts (*e.g.*, ‘BAMM’, ‘MiSSE’, and related; Rabosky, 2014;
559 Höhna et al., 2019; Vasconcelos et al., 2022), discrete cladogenetic shifts (*i.e.*, ‘ClaDS’; Maliet et al., 2019),
560 and, with our models, continuous anagenetic diffusion. The level at which diversification rates change in a

561 few discrete bouts or gradually is unknown, and probably both types of events have contributed in generating
562 diversification rate variance (Barnosky, 2001; Benton, 2009; Maliet et al., 2019). Arguably, even those novel
563 traits thought of substantially impacting lineage diversification rates (*e.g.*, ‘key’ innovations) do not appear
564 ‘instantaneously’ in time, but take some time to evolve, even if brief (Mayr, 1963; Heard and Hauser, 1995;
565 Hunter, 1998). Similarly, it remains an open and intriguing question whether changes in diversification
566 rates occur mostly at cladogenesis, a likely prediction under punctuated equilibrium with trait dependence
567 of diversification rates (Gould and Eldredge, 1972; Maliet et al., 2019), or evolve throughout a lineage’s
568 lifetime. We are only aware of one phylogenetic study that compared BAMM, a model of few discrete shifts
569 (Rabosky, 2014), with ClaDS, a model with many branch specific shifts (Maliet et al., 2019), and found
570 higher Bayesian support for the latter across some empirical trees, supporting the idea that most variation
571 in diversification rates are accumulated through frequent small-scale variation (Ronquist et al., 2021). Given
572 that all these processes could be present to some degree in empirical datasets, comparing models with
573 exclusive processes may not be particularly meaningful. A promising and exciting perspective of the data
574 augmentation methods introduced here is that they can easily be recruited to develop models that would
575 combine those hitherto disconnected processes into a joint model, where their respective contributions could
576 be directly estimated.

577 Our new model development uncovered a clear signal of an early-rise of modern mammals, refuting the
578 idea that suppression before the K-Pg was limiting the diversification of those mammal lineages that survived
579 to the present, but rather suggest the angiosperm radiation opened substantial ecological opportunities for
580 these lineages to diversify in the Late Cretaceous and up to the present. These results derive from inferring
581 rates of speciation and extinction that can vary at any time for any lineage, as theoretically expected from
582 the interplay of intrinsic and extrinsic factors acting at a given time on species, and opens new modeling
583 and computational possibilities in the study of macroevolutionary dynamics.

584 **Acknowledgments and funding**

585 This project has received funding from the European Union’s Horizon 2020 research and innovation pro-
586 gramme under the Marie Skłodowska-Curie grant agreement No 897225 for IQ.

587 **References**

- 588 Álvarez-Carretero, S., A. U. Tamuri, M. Battini, F. F. Nascimento, E. Carlisle, R. J. Asher, Z. Yang, P. C.
589 Donoghue, and M. Dos Reis. 2021. A species-level timeline of mammal evolution integrating phylogenomic
590 data. *Nature Pages* 1–8.
- 591 Andrieu, C. and G. O. Roberts. 2009. The pseudo-marginal approach for efficient monte carlo computations.
592 *The Annals of Statistics* 37:697–725.
- 593 Archibald, J. D. and D. H. Deutschman. 2001. Quantitative analysis of the timing of the origin and diversi-
594 fication of extant placental orders. *Journal of Mammalian Evolution* 8:107–124.

- 595 Barido-Sottani, J., T. G. Vaughan, and T. Stadler. 2020. A Multitype Birth–Death Model for Bayesian
596 Inference of Lineage-Specific Birth and Death Rates. *Systematic Biology* 69:973–986.
- 597 Barnosky, A. D. 2001. Distinguishing the effects of the red queen and court jester on miocene mammal
598 evolution in the northern rocky mountains. *Journal of Vertebrate Paleontology* 21:172–185.
- 599 Beaulieu, J. M. and B. C. O’Meara. 2015. Extinction can be estimated from moderately sized molecular
600 phylogenies. *Evolution* 69:1036–1043.
- 601 Beaulieu, J. M. and B. C. O’Meara. 2016. Detecting hidden diversification shifts in models of trait-dependent
602 speciation and extinction. *Systematic biology* 65:583–601.
- 603 Benson, R. B., P. D. Mannion, R. J. Butler, P. Upchurch, A. Goswami, and S. E. Evans. 2013. Cretaceous
604 tetrapod fossil record sampling and faunal turnover: Implications for biogeography and the rise of mod-
605 ern clades. *Palaeogeography, Palaeoclimatology, Palaeoecology* 372:88–107 vertebrate palaeobiodiversity
606 patterns and the impact of sampling bias.
- 607 Benson, R. B. J., R. J. Butler, J. Alroy, P. D. Mannion, M. T. Carrano, and G. T. Lloyd. 2016. Near-stasis
608 in the long-term diversification of mesozoic tetrapods. *PLOS Biology* 14:1–27.
- 609 Benton, M. J. 2009. The red queen and the court jester: species diversity and the role of biotic and abiotic
610 factors through time. *Science* 323:728–732.
- 611 Bezanson, J., A. Edelman, S. Karpinski, and V. B. Shah. 2017. Julia: A fresh approach to numerical
612 computing. *SIAM Review* 59:65–98.
- 613 Bininda-Emonds, O. R., M. Cardillo, K. E. Jones, R. D. MacPhee, R. Beck, R. Grenyer, S. A. Price, R. A.
614 Vos, J. L. Gittleman, and A. Purvis. 2007. The delayed rise of present-day mammals. *Nature* 446:507–512.
- 615 Davies, T. J., S. A. Fritz, R. Grenyer, C. D. L. Orme, J. Bielby, O. R. P. Bininda-Emonds, M. Cardillo, K. E.
616 Jones, J. L. Gittleman, G. M. Mace, and A. Purvis. 2008. Phylogenetic trees and the future of mammalian
617 biodiversity. *Proceedings of the National Academy of Sciences* 105:11556–11563.
- 618 Fabre, P.-H., L. Hautier, D. Dimitrov, and E. J. P. Douzery. 2012. A glimpse on the pattern of rodent
619 diversification: a phylogenetic approach. *BMC evolutionary biology* 12:1–19.
- 620 FitzJohn, R. G. 2010. Quantitative Traits and Diversification. *Systematic Biology* 59:619–633.
- 621 Gould, S. J. and N. Eldredge. 1972. Punctuated equilibria: an alternative to phyletic gradualism. *Models in*
622 *paleobiology* 1972:82–115.
- 623 Grossnickle, D. M. and E. Newham. 2016. Therian mammals experience an ecomorphological radiation
624 during the late cretaceous and selective extinction at the k–pg boundary. *Proceedings of the*
625 *Royal Society B: Biological Sciences* 283:20160256.
- 626 Grossnickle, D. M. and P. D. Polly. 2013. Mammal disparity decreases during the cretaceous angiosperm
627 radiation. *Proceedings of the Royal Society B: Biological Sciences* 280:20132110.

- 628 Grossnickle, D. M., S. M. Smith, and G. P. Wilson. 2019. Untangling the multiple ecological radiations of
629 early mammals. *Trends in Ecology & Evolution* 34:936–949.
- 630 Halliday, T. J. D. and A. Goswami. 2016. Eutherian morphological disparity across the end-Cretaceous mass
631 extinction. *Biological Journal of the Linnean Society* 118:152–168.
- 632 Heard, S. B. 1996. Patterns in phylogenetic tree balance with variable and evolving speciation rates. *Evolution*
633 50:2141–2148.
- 634 Heard, S. B. and D. L. Hauser. 1995. Key evolutionary innovations and their ecological mechanisms. *Historical*
635 *biology* 10:151–173.
- 636 Höhna, S., W. A. Freyman, Z. Nolen, J. P. Huelsenbeck, M. R. May, and B. R. Moore. 2019. A bayesian
637 approach for estimating branch-specific speciation and extinction rates. *bioRxiv* .
- 638 Höhna, S., M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot, B. R. Moore, J. P. Huelsenbeck, and
639 F. Ronquist. 2016. Revbayes: Bayesian phylogenetic inference using graphical models and an interactive
640 model-specification language. *Systematic biology* 65:726–736.
- 641 Horvillour, B. and N. Lartillot. 2014. Monte carlo algorithms for brownian phylogenetic models. *Bioinform-*
642 *atics* 30:3020–3028.
- 643 Hunter, J. P. 1998. Key innovations and the ecology of macroevolution. *Trends in Ecology & Evolution*
644 13:31–36.
- 645 Jablonski, D. 2008. Species selection: theory and data. *Annual Review of Ecology, Evolution, and Systematics*
646 39:501–524.
- 647 Kühnert, D., T. Stadler, T. G. Vaughan, and A. J. Drummond. 2016. Phylodynamics with Migration: A
648 Computational Framework to Quantify Population Structure from Genomic Data. *Molecular Biology and*
649 *Evolution* 33:2102–2116.
- 650 Liu, L., J. Zhang, F. E. Rheindt, F. Lei, Y. Qu, Y. Wang, Y. Zhang, C. Sullivan, W. Nie, J. Wang,
651 F. Yang, J. Chen, S. V. Edwards, J. Meng, and S. Wu. 2017. Genomic evidence reveals a radiation of
652 placental mammals uninterrupted by the kpg boundary. *Proceedings of the National Academy of Sciences*
653 114:E7282–E7290.
- 654 Louca, S. and M. W. Pennell. 2020. Extant timetrees are consistent with a myriad of diversification histories.
655 *Nature* 580:502–505.
- 656 Louca, S. and M. W. Pennell. 2021. Why extinction estimates from extant phylogenies are so often zero.
657 *Current Biology* 31:3168–3173.
- 658 Luo, Z.-X. 2007. Transformation and diversification in early mammal evolution. *Nature* 450:1011–1019.
- 659 Maliet, O., F. Hartig, and H. Morlon. 2019. A model with many small shifts for estimating species-specific
660 diversification rates. *Nature ecology & evolution* 3:1086–1092.

- 661 Maliet, O. and H. Morlon. 2021. Fast and accurate estimation of species-specific diversification rates using
662 data augmentation. *Systematic biology* .
- 663 Mammal Diversity Database. 2022. Mammal diversity database (version 1.8).
- 664 Marshall, C. R. 2017. Five palaeobiological laws needed to understand the evolution of the living biota.
665 *Nature Ecology & Evolution* 1:1–6.
- 666 Mayr, E. 1963. *Animal Species and Evolution*. Harvard University Press.
- 667 Meredith, R. W., J. E. Janečka, J. Gatesy, O. A. Ryder, C. A. Fisher, E. C. Teeling, A. Goodbla, E. Eizirik,
668 T. L. L. Simão, T. Stadler, D. L. Rabosky, R. L. Honeycutt, J. J. Flynn, C. M. Ingram, C. Steiner,
669 T. L. Williams, T. J. Robinson, A. Burk-Herrick, M. Westerman, N. A. Ayoub, M. S. Springer, and
670 W. J. Murphy. 2011. Impacts of the cretaceous terrestrial revolution and kpg extinction on mammal
671 diversification. *Science* 334:521–524.
- 672 Moore, B. R., S. Höhna, M. R. May, B. Rannala, and J. P. Huelsenbeck. 2016. Critically evaluating the
673 theory and performance of bayesian analysis of macroevolutionary mixtures. *Proceedings of the National*
674 *Academy of Sciences* 113:9569–9574.
- 675 Nee, S., R. M. May, and P. H. Harvey. 1994. The reconstructed evolutionary process. *Philosophical Trans-*
676 *actions of the Royal Society of London. Series B: Biological Sciences* 344:305–311.
- 677 Nilsson, M. A., U. Arnason, P. B. Spencer, and A. Janke. 2004. Marsupial relationships and a timeline for
678 marsupial radiation in south gondwana. *Gene* 340:189–196.
- 679 O’Leary, M. A., J. I. Bloch, J. J. Flynn, T. J. Gaudin, A. Giallombardo, N. P. Giannini, S. L. Goldberg,
680 B. P. Kraatz, Z.-X. Luo, J. Meng, et al. 2013. The placental mammal ancestor and the post-k-pg radiation
681 of placentals. *science* 339:662–667.
- 682 Quintero, I. and M. J. Landis. 2019. Interdependent Phenotypic and Biogeographic Evolution Driven by
683 Biotic Interactions. *Systematic Biology* 69:739–755.
- 684 R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical
685 Computing Vienna, Austria.
- 686 Rabosky, D. L. 2010. Extinction rates should not be estimated from molecular phylogenies. *Evolution:*
687 *International Journal of Organic Evolution* 64:1816–1824.
- 688 Rabosky, D. L. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on
689 phylogenetic trees. *PloS one* 9:e89543.
- 690 Ronquist, F., J. Kudlicka, V. Senderov, J. Borgström, N. Lartillot, D. Lundén, L. Murray, T. B. Schön, and
691 D. Broman. 2021. Universal probabilistic programming offers a powerful approach to statistical phyloge-
692 netics. *Communications biology* 4:1–10.

- 693 Springer, M. S., N. M. Foley, P. L. Brady, J. Gatesy, and W. J. Murphy. 2019. Evolutionary models for the
694 diversification of placental mammals across the kpg boundary. *Frontiers in Genetics* 10.
- 695 Springer, M. S., R. W. Meredith, J. Gatesy, C. A. Emerling, J. Park, D. L. Rabosky, T. Stadler, C. Steiner,
696 O. A. Ryder, J. E. Janečka, C. A. Fisher, and W. J. Murphy. 2012. Macroevolutionary dynamics and
697 historical biogeography of primate diversification inferred from a species supermatrix. *PLOS ONE* 7:1–
698 23.
- 699 Stadler, T. 2011a. Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National*
700 *Academy of Sciences* 108:6187–6192.
- 701 Stadler, T. 2011b. Simulating trees with a fixed number of extant species. *Systematic biology* 60:676–684.
- 702 Upham, N. S., J. A. Esselstyn, and W. Jetz. 2019. Inferring the mammal tree: Species-level sets of phylogenies
703 for questions in ecology, evolution, and conservation. *PLOS Biology* 17:1–44.
- 704 Upham, N. S., J. A. Esselstyn, and W. Jetz. 2021. Molecules and fossils tell distinct yet complementary
705 stories of mammal diversification. *Current Biology* 31:4195–4206.e3.
- 706 Van Valen, L. 1973. A new evolutionary law. *Evolutionary Theory* 1:1–30.
- 707 Vasconcelos, T., B. C. O’Meara, and J. M. Beaulieu. 2022. A flexible method for estimating tip diversification
708 rates across a range of speciation and extinction scenarios. *bioRxiv* .
- 709 Wilson, G. P. 2014. Mammalian extinction, survival, and recovery dynamics across the Cretaceous-Paleogene
710 boundary in northeastern Montana, USA. *in* *Through the End of the Cretaceous in the Type Locality of*
711 *the Hell Creek Formation in Montana and Adjacent Areas*. Geological Society of America.
- 712 Wilson, G. P., A. R. Evans, I. J. Corfe, P. D. Smits, M. Fortelius, and J. Jernvall. 2012. Adaptive radiation
713 of multituberculate mammals before the extinction of dinosaurs. *Nature* 483:457–460.

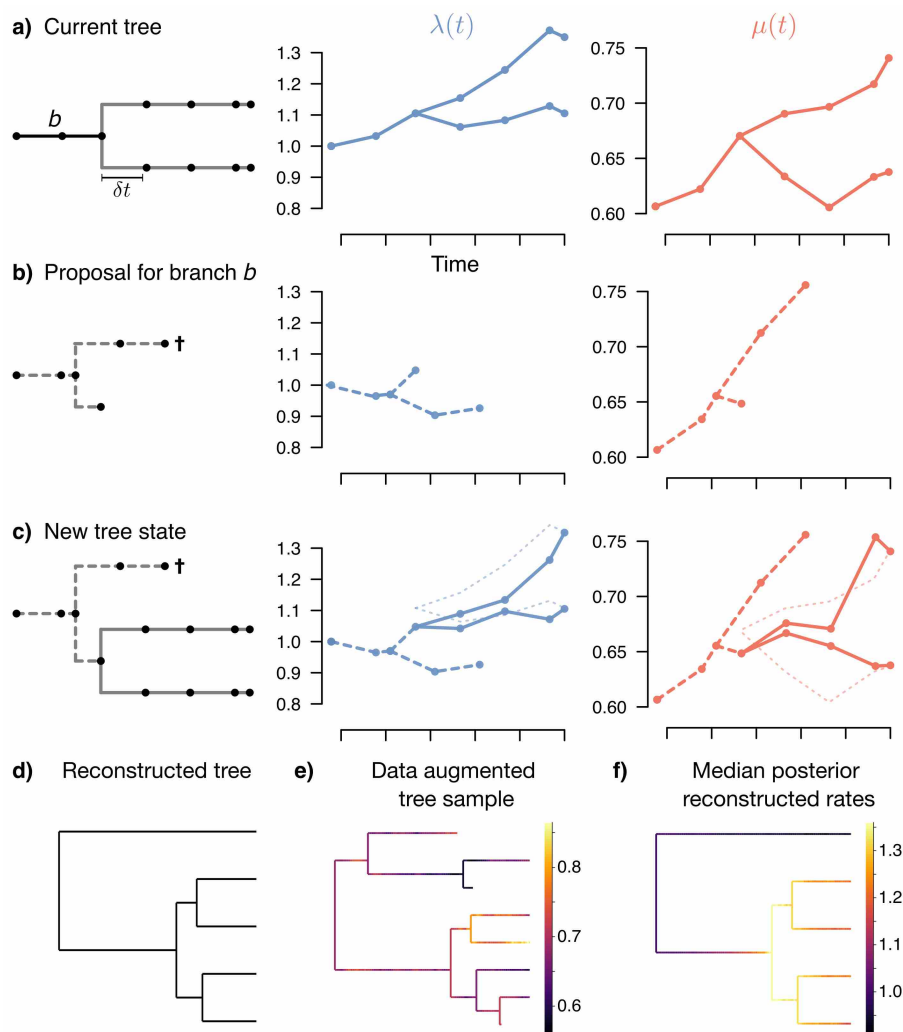


Figure 1: Tree BDD data augmentation. **a)** A phylogenetic tree is divided by a discrete time grid every δ unit times (or less) with underlying speciation $\lambda(t)$ (blue) and extinction $\mu(t)$ rates (red). **b)** To make a new subtree proposal for branch b , we simulate the process with the current configuration of parameters for the duration of the branch. Here, two tips were simulated for this duration, so we picked one at random to be the one leading to the observed speciation event and we continue the simulation for the rest (here the other tip went extinct). **c)** Since the new proposal in branch b has different speciation and extinction rates at the tip than the current subtree of branch b ($\lambda(t)$ and $\mu(t)$ thin dashed lines), we also propose rates in the daughters that match the proposal rates configuration ($\lambda(t)$ and $\mu(t)$ solid lines), and accept or reject according to the MH acceptance ratio. **d)** An example reconstructed tree. **e)** A given data augmented MCMC sample when running the BDD method on the tree from d). **f)** Median posterior reconstructed $\lambda(t)$ from the BDD model across the tree. Color gradient represents lineage-specific variation in speciation rates along the tree, with warmer colors specifying higher rates.

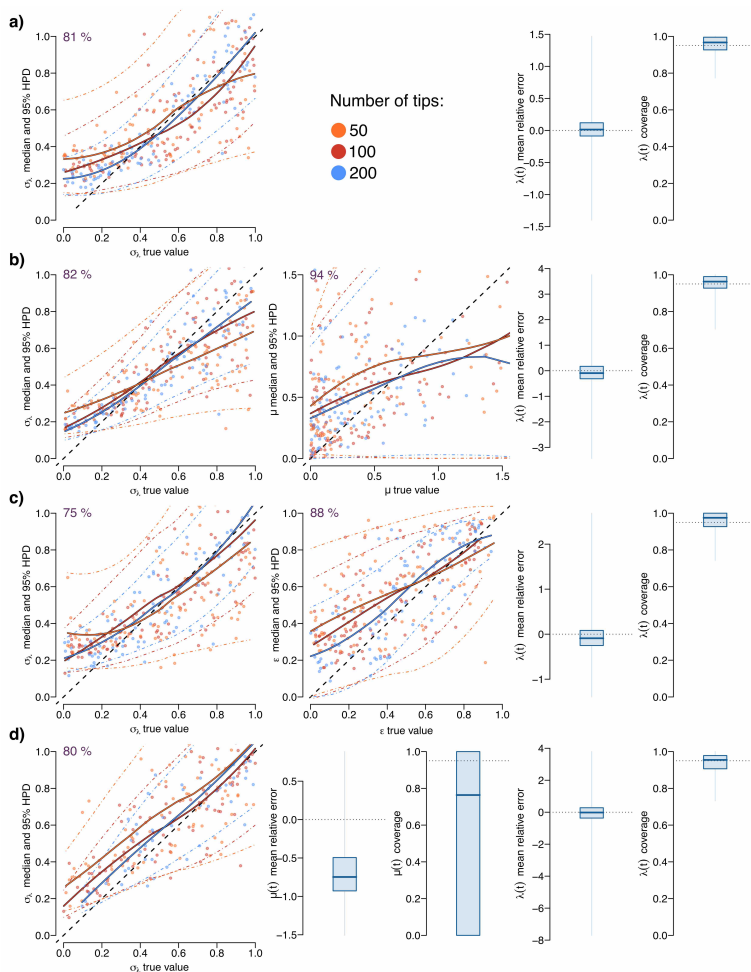


Figure 2: Validation of diffusion model without drift. Comparison between simulated and estimated values for **a)** pure-birth diffusion (‘PBD’: $\mu(t) = 0$), **b)** constant-extinction diffusion (‘CED’: $\mu(t) = \mu$), **c)** constant-turnover diffusion (‘CTD’: $\mu(t) = \epsilon\lambda(t)$), and **d)** extinction-diffusion (‘BDD’: $\mu(t)$), for 50, 100 and 200 tip simulations. For σ_λ , μ , and ϵ , we show the running means across simulations for the median estimates and the 95% HPD for each of the tree sizes, respectively, and the 1:1 line (black dashed). In the upper left corner we show the coverage for those parameters. For $\lambda(t)$ (and, when applicable, $\mu(t)$) we estimate the percentage average relative error across a series of fixed time points along the tree. Similarly, we calculate the coverage by counting the number of true values that fall within the 95% HPD in the estimates across those same fixed time points.

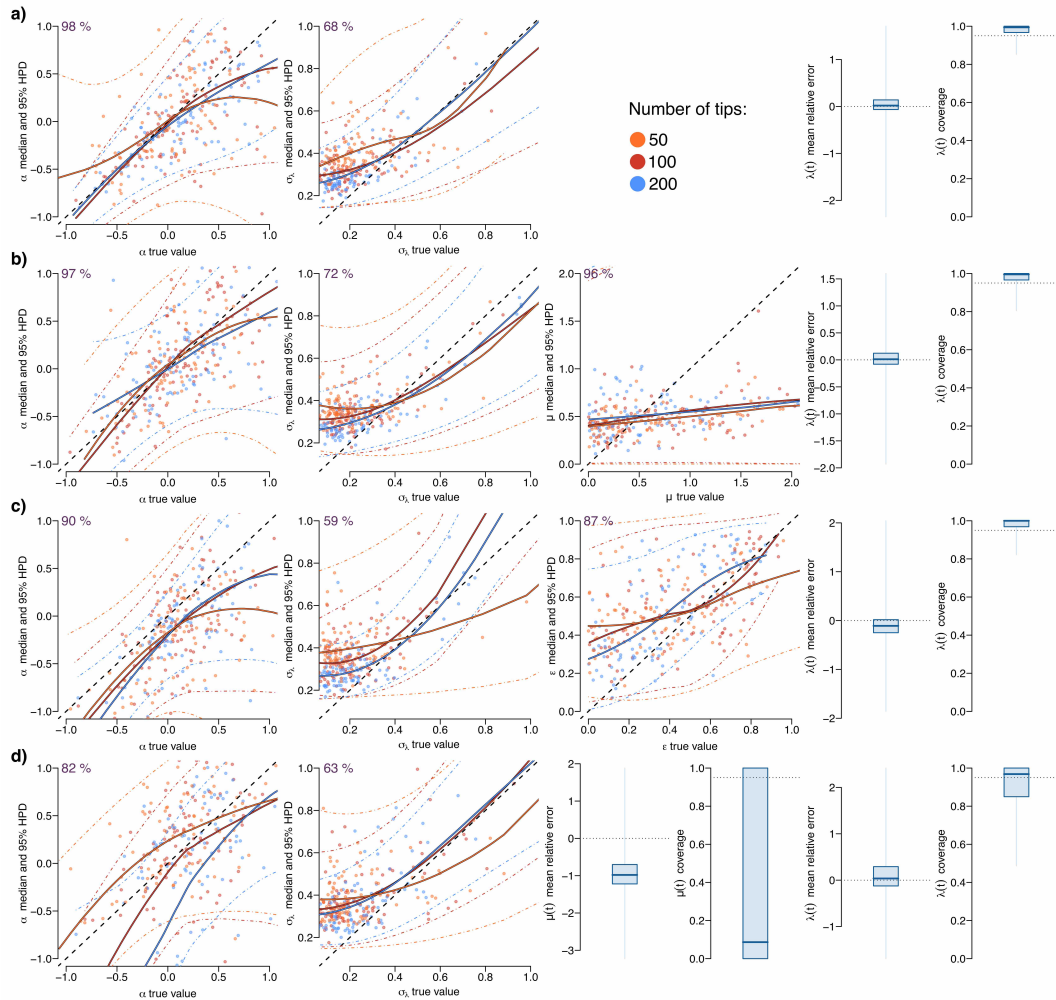


Figure 3: Validation of diffusion model with drift. Comparison between simulated and estimated values when including the drift parameter α for a) PBD, b) CED, c) CTD and d) BDD, for 50, 100 and 200 tip simulations. Plot details as in Figure 2.

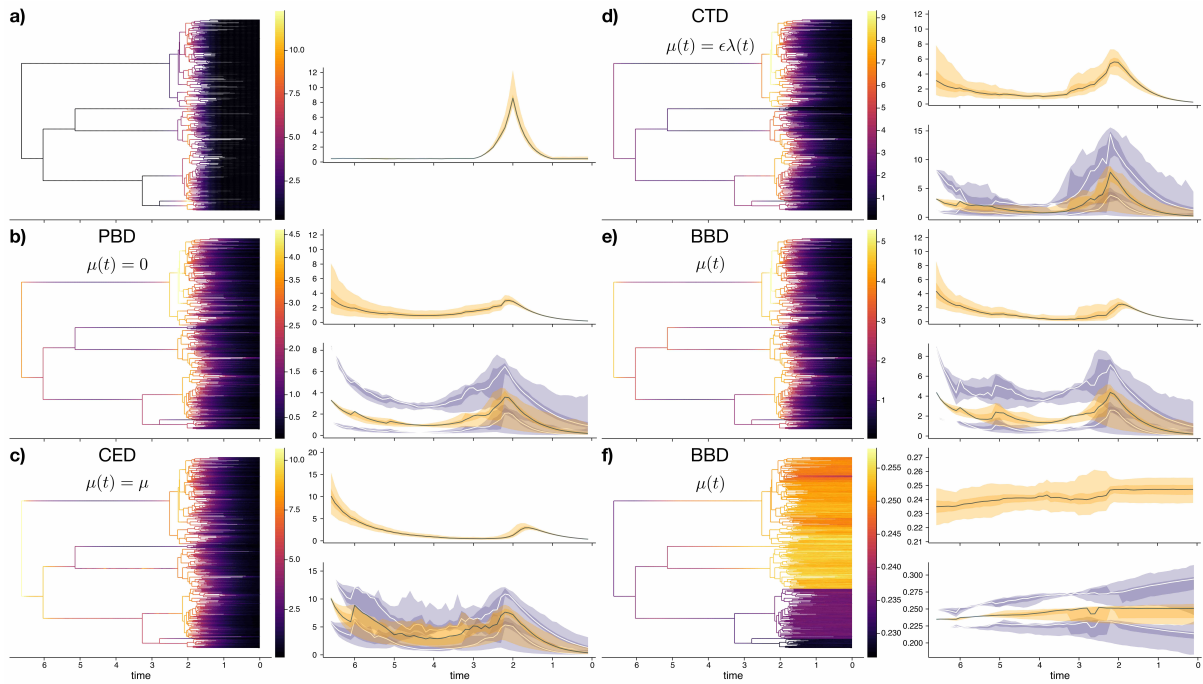


Figure 4: Common temporal patterns through time. **a)** Simulated tree (*left*) with dramatic temporal variation in speciation rates affecting all lineages simultaneously (*right*: geometric average $\lambda(t)$ across lineages and 95% quantiles), small inter-lineage variance $\sigma_\lambda = 0.1$ and a constant extinction rate of 0.2. **b-f)** Show inferred patterns given the simulated tree. For each, *Left*: Median posterior reconstructed rates across the tree (speciation for ‘b-e’ and extinction for ‘f’); *Upper right*: posterior DA rates and 95% Credible Intervals (CI) across all data augmented trees; *Bottom right*: posterior median and 95% CI reconstructed rates. Inferred models are: **b)** PBD (*i.e.*, $\mu(t) = 0$), **c)** CED (*i.e.*, $\mu(t) = \mu$), **d)** CTD (*i.e.*, $\mu(t) = \epsilon\lambda(t)$), and BBD (*i.e.*, $\mu(t)$) **e)** speciation and **f)** extinction rates.

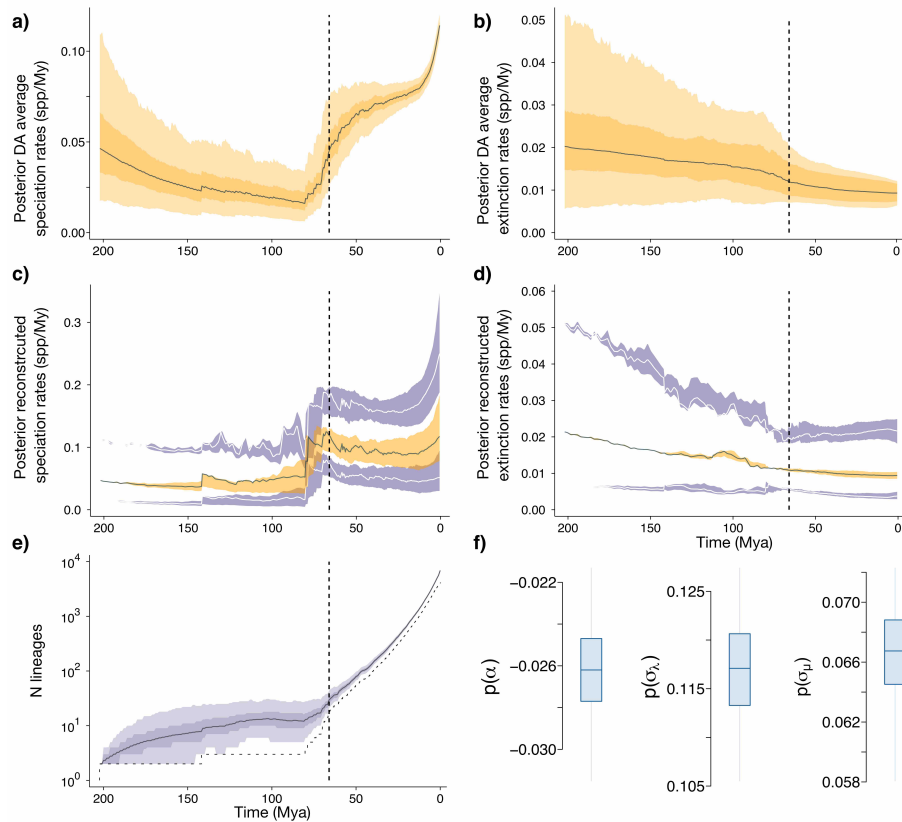


Figure 5: Birth-Death Diffusion for surviving mammals: overarching patterns a) Posterior DA average speciation rates and b) extinction rates across all lineages across augmented tree samples. Solid line shows the median, darker orange the 50% CI, lighter orange the 95% CI across all data augmented trees. c) Posterior reconstructed speciation rates and d) extinction rates for median posterior tree (orange shading showing median and 50% quantiles across the lineages' median posterior rates), and for 2.5% and 97.5% CI trees (purple shading showing median and 50% quantiles across the lineages' 2.5% and 97.5% posterior rates, respectively). e) Distribution through time of the number of mammals lineages in the data augmented trees (solid line shows the median, darker purple the 50% CI, lighter purple the 95% CI). The lower dotted line shows the lineages in the reconstructed tree only. f) Posterior distributions (95 %, 50 % and median HPD) for BDD hyperparameters: drift α , diffusion in speciation rates σ_λ and in extinction rates σ_μ .

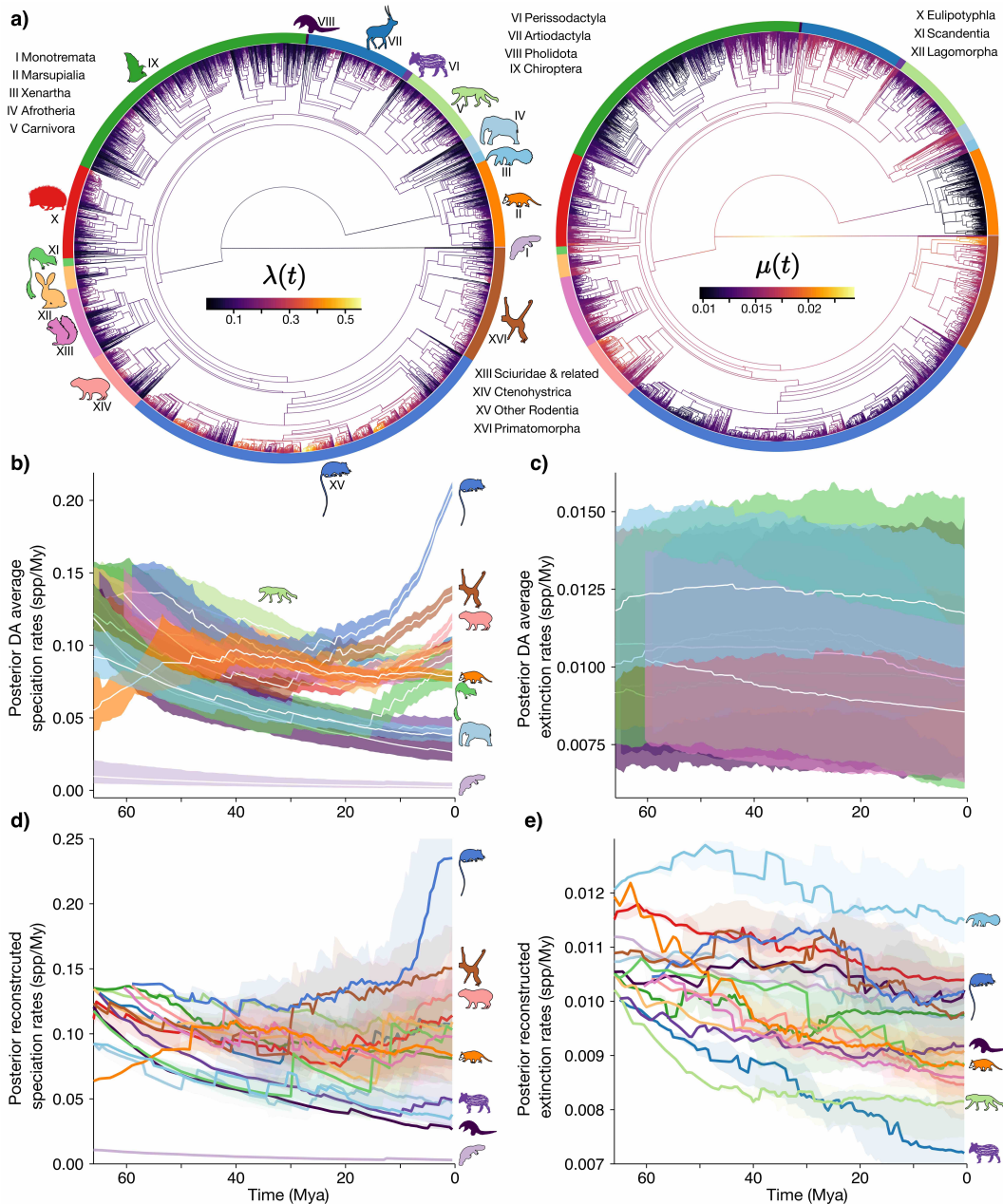


Figure 6: Birth-Death Diffusion for surviving mammals: lineage heterogeneity a) Median posterior reconstructed speciation rates $\lambda(t)$ and extinction rates $\mu(t)$ across the reconstructed mammals tree according to the Birth-Death Diffusion model (note that this are the same model results of Fig 5). Surrounding colors identify 16 mammal clades with embedded species silhouette and roman numerals for identification. b) Posterior DA average speciation rates and c) extinction rates the 16 major clades highlighted in 'a'. Solid white line shows the median, while the darker shade the 50% CI across all data augmented trees. Silhouettes identify some focal clade patterns. d) Posterior reconstructed speciation rates and e) extinction rates for median posterior tree (solid lines shows the median and light shading shows the 50% quantiles across the lineages' median posterior rates).