

# Mammalian phylogeny reveals recent diversification rate shifts

Tanja Stadler<sup>1</sup>

Institut für Integrative Biologie, Eidgenössische Technische Hochschule Zurich, 8092 Zurich, Switzerland

Edited by David M. Hillis, University of Texas at Austin, Austin, TX, and approved March 2, 2011 (received for review November 9, 2010)

**Phylogenetic trees of present-day species allow investigation of the rate of evolution that led to the present-day diversity. A recent analysis of the mammalian phylogeny challenged the view of explosive mammalian evolution after the Cretaceous–Tertiary (K/T) boundary (65 Mya). However, due to lack of appropriate methods, the diversification (speciation minus extinction) rates in the more recent past of mammalian evolution could not be determined. In this paper, I provide a method that reveals that the tempo of mammalian evolution did not change until ~33 Mya. This constant period was followed by a peak of diversification rates between 33 and 30 Mya. Thereafter, diversification rates remained high and constant until 8.55 Mya. Diversification rates declined significantly at 8.55 and 3.35 Mya. Investigation of mammalian subgroups (marsupials, placentals, and the six largest placental subgroups) reveals that the diversification rate peak at 33–30 Mya is mainly driven by rodents, cetartiodactyla, and marsupials. The recent diversification rate decrease is significant for all analyzed subgroups but eulipotyphla, cetartiodactyla, and primates. My likelihood approach is not limited to mammalian evolution. It provides a robust framework to infer diversification rate changes and mass extinction events in phylogenies, reconstructed from, e. g., present-day species or virus data. In particular, the method is very robust toward noise and uncertainty in the phylogeny and can account for incomplete taxon sampling.**

macroevolution | maximum-likelihood inference | speciation rates

It has been a long-standing question whether the rise of present-day mammals began following the mass extinction event at the K/T boundary (1, 2). The hypothesis of a significant rise in mammals following the K/T boundary was challenged when the mammalian phylogeny, including 4,510 present-day species, became available (3). This phylogeny is ~83% complete at the species level (4). It allowed detection of diversification rate (speciation rate minus extinction rate) shifts throughout the evolutionary past of almost all present-day mammals. No shift was detected at the K/T boundary; however, a peak in diversification rates at ~93 Mya was detected (3).

Estimating the time and amount of diversification rate changes is a key toward understanding evolutionary patterns (5, 6). For example, reliable diversification rate estimates can have the power to decide if species richness is typically due to intrinsic causes (the ancestor evolved a new feature) or extrinsic causes (the environment changed) or a complex combination of both. Already in 1996, Sanderson and Donoghue (ref. 5, p. 1) wrote “Few issues in evolutionary biology have received as much attention over the years [ . . . ] as those involving evolutionary rate.” However, until now there has been no general framework available to estimate diversification rates and their changes through time.

Detecting diversification rate changes is typically done by detecting changes in the slope of the lineages-through-time (LTT) plot of the present-day species (3, 7–9, 10). The LTT plot of the mammals and the main mammalian subgroups are displayed in Fig. 1. For constant speciation rate  $\lambda$  and constant extinction rate  $\mu$ , the slope of the semilog LTT plot is  $\lambda - \mu$  in the distant past and  $\lambda$  in the recent past (11), with the change in slope occurring  $\sim 1/(\lambda - \mu)$  time units ago. This change in slope is also called the

“pull of the present” (7). A departure from a straight line in the distant past indicates a diversification rate change.

I argue that a slope change before the pull of the present slope change is not necessarily caused by a diversification rate change. Suppose a very high speciation rate  $\lambda_1$  and a very high extinction rate  $\mu_1$  in the distant past, until say 25 Mya. The LTT plot of the species extant at 25 Mya has first a constant slope  $\lambda_1 - \mu_1$ , followed by slope  $\lambda_1$  when approaching 25 Mya (pull of the present, where the present is 25 Mya). Suppose more recent than 25 Mya, there was a constant speciation rate  $\lambda_0$  and no extinction, meaning that all species alive at 25 Mya have present-day species descending. This observation yields an LTT plot of the present-day species with slope  $\lambda_1 - \mu_1$  in the distant past,  $\lambda_1$  close to 25 Mya, and  $\lambda_0$  more recent than 25 Mya. The diversification rate analysis based on slope changes in LTT plots for the time period before 25 Mya (as done for the mammalian phylogeny) (3) would have detected a period with diversification rate  $\lambda_1 - \mu_1$  followed by a period with diversification rate  $\lambda_1$  before 25 Mya, whereas there was actually only one period with diversification rate  $\lambda_1 - \mu_1$  before 25 Mya.

An approach that estimates simultaneously all rates throughout the evolution of a considered phylogeny, instead of looking for local slope changes, is required to avoid such misleading results.

## A New Likelihood Approach

I formulate an evolutionary model, the *birth–death–shift* process, where the speciation and extinction rates can change through time (*Methods*). When no rate change occurs, my model simplifies to the well-studied constant rate birth–death processes (12–16). By deriving the likelihood function (*Methods*, Theorem 2.7) of a phylogenetic tree under the birth–death–shift model, maximum-likelihood time intervals together with each interval’s diversification rate (speciation rate minus extinction rate) and turnover (extinction rate divided by speciation rate) can be estimated. Likelihood-ratio tests are used to decide how many shifts are best supported. Using simulations (17), I show in *SI Text* that my analytical likelihood approach is powerful in detecting rate shifts accurately and is robust toward noise in data, such as uncertainty in divergence dates and unresolved polytomies. My likelihood parameter estimation method is available as an R package on CRAN (18).

## Tempo of Mammalian Diversification

Using this unique likelihood approach, I estimated the maximum-likelihood speciation and extinction rates together with the shift times for the mammalian phylogeny (3) (*SI Text SI*). I detect four significant diversification rate shifts throughout mam-

Author contributions: T.S. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

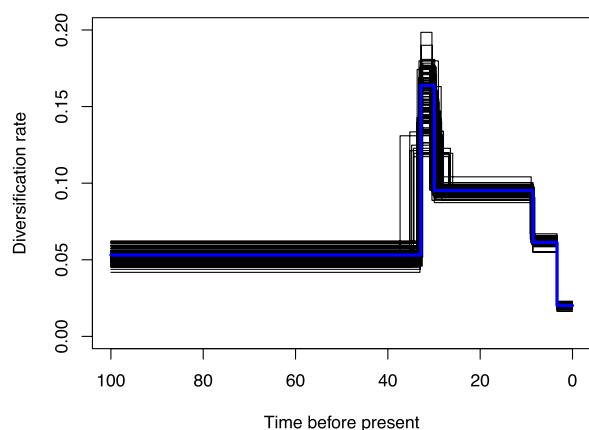
The author declares no conflict of interest.

This article is a PNAS Direct Submission.

<sup>1</sup>E-mail: tanja.stadler@env.ethz.ch.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1016876108/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1016876108/-DCSupplemental).





**Fig. 4.** Maximum-likelihood diversification rate estimates (per million years) under the four-shift model for 100 simulated trees (black) and the mammalian phylogeny (blue). The maximum-likelihood parameters obtained from the mammalian phylogeny were used for simulations.

Up to now, the recent past (the last 25 Myr) could not be analyzed on the basis of the mammalian phylogeny, as the available methods could not distinguish between the pull of the present and a real rate change (11, 3). Two studies (22, 23) based on fossil data investigated the diversification rates of mammals between 30 and 5 Mya. Both studies found a major increase in species richness during 17–14 Mya. When I allow for at least six shifts, a diversification rate increase  $\sim 15.85$  Mya (Fig. S1) is detected. Note, however, that the additional shifts are non-significant ( $P$  value = 0.28).

I detect a decline in diversification rates for all mammals and most mammalian subgroups ~3.35 Mya. To my knowledge, no existing study investigated the diversification rates of mammals during this time.

My study provides maximum-likelihood diversification rate estimates on the basis of the whole mammalian phylogeny. However, it remains controversial why the rates were shifting.

A study on the response of mammals to global warming (ref. 24, p. 1) states “Faunal turnover nears 100% and species diversity may increase when warm temperatures last hundreds of thousands to millions of years, because speciation takes place and faunal changes initiated by a variety of shorter-term processes accumulate.” The two studies discussed above (22, 23) find accelerated diversification during only a single warm period, 17–14 Mya. My study does not suggest higher diversification rates during long warm periods. In particular, during the Early-Eocene climatic optimum at 53–51 Mya, when temperatures reached a long-term

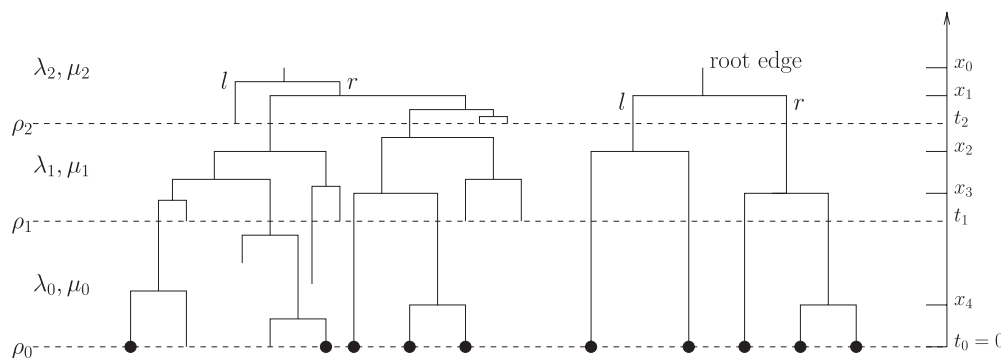
maximum, I detect no diversification rate increase. Further, I do not detect a significant rate increase during the warmest part of the Mid-Miocene climatic optimum at 17–14 Mya.

On the other hand, I find an increase in diversification rate and turnover during cooling. Around 34 Mya, the earth climate changed from warm to cool (25) and remained cool until today. Antarctica began to be glaciated, grasslands expanded, presumably many herbivores increased considerably in population sizes, and, as I show, mammals underwent rapid diversification at 33–30 Mya, in particular rodents, cetartiodactyla, and marsupials. To disentangle whether increased mammalian diversification rates were caused by climate cooling or expanding grassland or neither, I suggest analyzing nonmammalian herbivore phylogenies with my maximum-likelihood method. Note that the increased diversification rate for cetartiodactyla indicates that not only grassland expansion caused the rate increase in mammals.

The statistically most significant rate shift observed in my study, at 3.35 Mya, occurred in an era of high tectonic activity: At 3.5 Mya, the great American biotic interchange took place as the new land bridge between North and South America was formed. However, I find a decrease in diversification rate, as opposed to an increase in rate suggested during times of high tectonic activity (23).

My results are based on the previously published phylogeny of mammals (3). This phylogeny was inferred using a supertree approach that combines reconstructed mammalian subtrees into one phylogeny of all mammals. The accuracy of such supertree approaches is limited (26), and the analyzed mammalian tree will certainly be refined in the future, possibly having an effect on some of the conclusions I draw. In particular, the mammalian supertree is poorly resolved in the recent past, during which I detect two significant rate shifts (3.35 and 8.55 Mya). However, because I show, using simulations, that my method is robust toward unbiased uncertainties in speciation time estimates and unresolved polytomies, I argue that my observed general rate pattern (including the recent shifts) actually reflects the pattern of mammalian evolution. As my method is available in the R package TreePar (18), my hypotheses can be easily checked and corrected with new mammalian phylogenies being inferred. In particular, new mammalian phylogenies will reveal if my assumption of the uncertainties in Bindinda's mammalian phylogeny being unbiased is appropriate.

My diversification rate estimates do not provide any support for the hypotheses of accelerated diversification during warming or during an era of high tectonic activity. However, with the increasing number of reconstructed phylogenies, my likelihood method can be used to obtain a more complete picture of the evolutionary past of present-day species. Such analyses can reveal common diversification patterns across



**Fig. 5.** Tree notation. (Left) An example of a tree that evolved under the birth–death shift process. The sampled species are denoted with a solid circle. (Right) The corresponding sampled tree with root edge. The labels  $l$  and  $r$  are suppressed on most branches for clarity.



different groups of species and, together with paleontological and paleoclimatic data, can help link environmental patterns and diversification patterns.

**Advantages, Extensions, and Limitations of the Presented Likelihood Method.** My likelihood method for detecting rate shifts opens the possibility for analyzing all kinds of phylogenetic trees more accurately. The method is not tailored specifically to species phylogenies but can also be used for, e.g., virus phylogenies. The method can detect in addition to rate shifts also mass extinction events (*Methods*).

My approach, which uses one general model throughout evolutionary time and infers the maximum-likelihood rate estimates, has several advantages over methods using only parts of the phylogeny by looking at local slope changes in the LTT plot:

- i) As discussed in the Introduction, slope changes do not necessarily correspond to rate changes; my method accounts for such slope changes without predicting a false rate change.
- ii) Rates in the recent and distant past can be inferred, for both complete phylogenies and incomplete phylogenies assuming random sampling.
- iii) Simulations reveal that the likelihood method is robust to inaccuracies in speciation time estimates. In particular, unresolved nodes (polytomies) do not pose a problem, whereas the available regression methods require well-dated edges (3).
- iv) When half of the lineages speciate in a short time interval, a rate shift is suggested by standard regression methods. However, if the number of lineages is small, the speciation pattern can easily be explained by stochastic effects. Likelihood methods using the whole tree account for such stochastic effects.
- v) Regression methods for estimating diversification rates require setting some parameters a priori (3) (e.g., basis dimension, gamma parameter, and time steps) and it is unclear how different settings influence the results. A likelihood framework requires no input besides the dated tree.

Previous studies derived likelihood approaches to infer diversification rates under special cases of the birth–death–shift process. Under a model without rate changes, i.e. constant rates throughout the whole evolutionary time, a likelihood approach is well established (14). The maximum-likelihood parameter estimation for the constant rate birth–death process was generalized to changing birth and death rates (14). However, only integral equations were available, which have to be evaluated numerically. Numerical integration has been done, e.g., on an *Enallagma* phylogeny (27). However, their 90% complete trees contained only 20 species and they allowed for only one rate shift. For big phylogenetic trees where several rate shifts can be expected, direct analytical methods become crucial to analyze the data.

A previous study (6) assumes the birth–death–shift model with one rate shift (and constraints on the shift time and rates). The parameters are estimated assuming that the tree before the rate shift time is independent of the tree after the rate shift time. However, this assumption is not valid, because extinction in a later interval affects species that appeared in an earlier interval and survived until the later interval.

Further, a mass extinction event produces an LTT plot with a constant slope interrupted by a plateau (11, 28). However, results were purely based on simulations as no likelihood function was available. The likelihood approach developed in this study accounts for mass extinction events.

My simulations revealed that diversification rates can be estimated with high confidence. However, the turnover cannot be estimated accurately, even in the case of no rate shift. For obtaining very accurate extinction rates, fossil data are necessary (29).

Models have been proposed where speciation and extinction rates are more complex than under the birth–death–shift model. For example, extinction rates might be heritable (30) (e.g., if intrinsic factors determine speciation rate and these intrinsic factors are heritable), or speciation rates might be density dependent (31). Up to now, such models can be investigated only with simulations. Maximum-likelihood methods to infer rates under these more complex models will require new analytic techniques. Once such methods become available, statistical tests can be applied to address the validity of models with changing rates at discrete points in time (presented in this paper) or models with continuously changing rates through time.

The presented likelihood equations can of course also be used in a Markov chain Monte Carlo (MCMC) approach to estimate the posterior parameter distribution for a given phylogeny, if one is willing to assume prior distributions for the parameters.

Also, the birth–death–shift likelihood can be used as an a priori distribution for reconstructing phylogenies in an MCMC framework. This method will be in particular interesting for reconstructing virus phylogenies. When using the birth–death–shift model for viruses, the birth rate corresponds to the transmission rate of a virus. As the transmission rate undergoes major declines when new treatment strategies become available (as treated patients usually cannot transmit), a model with rate shifts is very appropriate for viruses.

## Methods

**1) Birth–Death–Shift Model.** To detect shifts of diversification rates in a phylogeny, we need to define an evolutionary model accounting for such shifts. I define the model very generally, allowing for rate shifts as well as mass extinction events. In this way, the method is applicable to a wide range of datasets.

For modeling the shift of rates and mass extinction events, we define a *birth–death–shift process* as follows. The vector  $t = (t_0, \dots, t_m)$ , with  $0 = t_0 < t_1 < \dots < t_m$ , determines the times (before the present) of rate shift and mass extinction events, where  $t_0 = 0$  denotes the present. We set  $t_{m+1} = \infty$ . The vector  $\lambda = (\lambda_0, \dots, \lambda_m)$ , where  $\lambda_i > 0$  ( $i = 0, 1, \dots, m$ ) defines the speciation rates.  $\lambda_i$  is the speciation rate in time interval  $(t_i, t_{i+1}]$ . The vector  $\mu = (\mu_0, \dots, \mu_m)$ , where  $\lambda_i > \mu_i \geq 0$  ( $i = 0, 1, \dots, m$ ) defines the extinction rates.  $\mu_i$  is the extinction rate in time interval  $(t_i, t_{i+1}]$ . The vector  $\rho = (\rho_0, \dots, \rho_m)$ , where  $1 \geq \rho_i > 0$  ( $i = 0, 1, \dots, m$ ) defines the survival probabilities:  $\rho_i$  for  $i > 0$  is the probability of surviving a mass extinction event at time  $t_i$ . Note that  $\rho_i = 1$  corresponds to a rate shift from  $\lambda_i, \mu_i$  to  $\lambda_{i-1}, \mu_{i-1}$ , but no extinction at time  $t_i$ . The probability  $\rho_0$  is the probability of sampling an extant species.  $\rho_0 = 1$  corresponds to complete sampling of extant species.

The birth–death–shift process starts with one species at time  $x_0$  in the past and is stopped at time  $t_0 = 0$ . At time point  $t$ , where  $t_{i+1} > t > t_i$ , each species gives birth to a new species with rate  $\lambda_i$ ; the two subtending lineages are denoted  $l$  and  $r$  such that we can distinguish between them. Each species dies with rate  $\mu_i$ . At time  $t_i$ , each species survives the mass extinction with probability  $\rho_i$ . The process is stopped at time 0, the present. Each present-day species is sampled with probability  $\rho_0$ .

Such a process induces a tree (Fig. 5, *Left*). Suppressing all extinct and all nonsampled lineages yields the *sampled tree with root edge* (Fig. 5, *Right*). The *sampled tree* is obtained by deleting the edge subtending the origin. Note that the sampled tree is a binary tree with each leaf having the same distance from the root. Each internal vertex has a lineage  $l$  and a lineage  $r$  descending. In the sampled tree, we subdivide the edge from  $t_s$  to  $t_e$  with time  $t_s > t_i > t_e$  at time  $t_i$ . So at time  $t_i$ , we have a degree-two vertex. The  $n - 1$  speciation times in a sampled tree with  $n$  present-day species are  $x_1 > x_2 > \dots > x_{n-1}$ .

Note that a phylogeny inferred from data does not have the orientation labels  $l$  and  $r$  but each leaf has a unique label. We considered oriented trees and calculate their probability density because of convenient notation. The probability density of a labeled tree (each leaf is assigned a unique label, and the orientations  $l$  and  $r$  are dropped) follows from the equations for an oriented tree by multiplying with  $2^{n-1}/n!$ , as each labeling and each orientation is equally likely (32). Note that in maximum-likelihood parameter estimations on a fixed tree, the factor  $2^{n-1}/n!$  is constant, and therefore it can be discarded. Therefore, throughout the rest of the paper, we consider only oriented trees.

**2) Deriving the Likelihood of a Tree.** The aim of this section is to calculate the probability density of a sampled tree  $\mathcal{T}$  under the birth–death–shift process. We do so by first calculating the probability of a species having 0 descendants [2.1) *Probability of 0 extant descendants*]. Using this probability, we calculate the density of a tree conditioned on the time of origin [2.2) *Probability density of a tree conditioned on  $x_0$* ] and the time of the most recent common ancestor [Section 2.3) *Probability density of a tree conditioned on the time of the most recent common ancestor*].

Having the general expression for the probability density of any sampled tree allows us to calculate the probability density of a data phylogeny (which is a particular sampled tree). This probability density is maximized over the parameter space to obtain the maximum-likelihood parameter estimates for the phylogeny.

### 2.1) Probability of 0 extant descendants.

**Theorem 2.1.** *The probability that a species alive at time  $t$  before today has no sampled extant descendants with  $t_i \leq t \leq t_{i+1}$  ( $i = 0, \dots, m$  with  $t_{m+1} = \infty$ ) is*

$$p_0^i(t) = \frac{\mu_i(c_i - 1)e^{-(\lambda_i - \mu_i)t} + (\mu_i - c_i\lambda_i)e^{-(\lambda_i - \mu_i)t}}{\lambda_i(c_i - 1)e^{-(\lambda_i - \mu_i)t} + (\mu_i - c_i\lambda_i)e^{-(\lambda_i - \mu_i)t}} \quad [1]$$

with  $c_i = (1 - \rho_i) + \rho_i p_0^{i-1}(t_i)$ , and  $c_0 = 1 - \rho_0$ .

*Proof.* The master equation for the probability  $p_0^i(t)$  is, for  $t_i \leq t \leq t_{i+1}$ ,

$$\begin{aligned} \frac{d}{dt} p_0^i(t) &= \mu_i - (\lambda_i + \mu_i)p_0^i(t) + \lambda_i p_0^i(t)^2, \\ p_0^i(t_i) &= (1 - \rho_i) + \rho_i p_0^{i-1}(t_i), p_0^0(0) = 1 - \rho_0. \end{aligned}$$

By plugging Eq. 1 into the master equation, one can verify that it is a solution of the master equation. For showing the uniqueness, define

$$\dot{p}_0^i(t) = f(p_0^i, t).$$

As  $f, \partial f / \partial p_0^i$  are continuous, uniqueness follows by the uniqueness theorem for ordinary differential equations (e.g., ref. 33, p. 211).  $\square$

**Remark 2.2.** Note that for  $i = 0$ , we have

$$\begin{aligned} p_0^0(t) &= \frac{\mu_0 \rho_0 - (\mu_0 - (1 - \rho_0)\lambda_0)e^{-(\lambda_0 - \mu_0)t}}{\lambda_0 \rho_0 - (\mu_0 - (1 - \rho_0)\lambda_0)e^{-(\lambda_0 - \mu_0)t}} \\ &= 1 - \frac{\rho_0(\lambda_0 - \mu_0)}{\lambda_0 \rho_0 + ((1 - \rho_0)\lambda_0 - \mu_0)e^{-(\lambda_0 - \mu_0)t}}, \end{aligned}$$

which was already established (14).

In the next section, we need Eq. 1 to calculate the density function of a sampled tree  $\mathcal{T}$ .

**2.2) Probability density of a tree conditioned on  $x_0$ .** First note that the time of the start of the process,  $x_0$ , is a parameter of the birth–death–shift process, and therefore we cannot state a tree probability density independent of  $x_0$  unless we assume a prior distribution for  $x_0$ . In the following, we simply condition on the time  $x_0$  being fixed. In the subtending section, we then calculate the probability density of a tree conditioned on the time of the most recent common ancestor of the extant and sampled species: This tree can be considered as two trees with time of origin at  $x_0$ . The time of the most recent common ancestor of a clade is the age of the clade. This value is known for well-studied clades and therefore conditioning on this value is reasonable and generally done when estimating rates (19).

For  $t_i \leq t \leq t_{i+1}$ , let  $g_{i,e}(t)$  be the probability density that the species corresponding to edge  $e$  at time  $t$  evolved between  $t$  and the present as observed in  $\mathcal{T}$ .

**Theorem 2.3.** *The unique solution of the ordinary differential equation for  $g_{i,e}(t)$  is*

$$g_{i,e}(t) = d_{i,e} \frac{\hat{g}_i(t)}{\hat{g}_i(t_e)} \quad [2]$$

with

$$\hat{g}_i(t) = \frac{e^{-(\lambda_i - \mu_i)(t_i + t)}}{(\lambda_i(c_i - 1)e^{-(\lambda_i - \mu_i)t_i} + (\mu_i - c_i\lambda_i)e^{-(\lambda_i - \mu_i)t})^2}. \quad [3]$$

Note that  $\hat{g}_i(t_i) = (\lambda_i - \mu_i)^{-2}$  for all  $i \geq 0$ .

*Proof.* The master equation for  $g_{i,e}(t)$  along an edge with starting time  $t_s$  and ending time  $t_e$  is, with  $t_s \geq t \geq t_e$ ,

$$\frac{d}{dt} g_{i,e}(t) = -(\lambda_i + \mu_i)g_{i,e}(t) + 2\lambda_i p_0^i(t)g_{i,e}(t) \quad [4]$$

with the initial values at  $t = t_e$ ,

$$g_{i,e}(t_e) := d_{i,e} = \begin{cases} \lambda_i g_{i,e_1}(t_e) g_{i,e_2}(t_e) & \text{if } e \text{ has two descendant edges } e_1, e_2, \\ \rho_i g_{i-1,e_1}(t_i) & \text{if } e \text{ has one descendant edge } e_1, \\ \rho_0 & \text{if } e \text{ is a leaf.} \end{cases}$$

By plugging Eq. 2 into the master equation, one can verify that it is a solution of the master equation. For showing the uniqueness, define

$$\dot{g}_{i,e}(t) = f(g_{i,e}, t).$$

As  $f, \partial f / \partial g_{i,e}$  are continuous, uniqueness follows by the uniqueness theorem for ordinary differential equations (e.g., ref. 33, p. 211).  $\square$

**Corollary 2.4.** *Let  $p_i^j(t)$  be the probability that a species alive at time  $t$  before today has one sampled extant descendant with  $t_i \leq t \leq t_{i+1}$  ( $i = 1, \dots, m$  with  $t_{m+1} = \infty$ ). Then*

$$p_i^j(t) = \rho_i(\lambda_i - \mu_i)^2 \hat{g}_i(t) \prod_{j=0}^{i-1} (\rho_j(\lambda_j - \mu_j)^2 \hat{g}_j(t_{j+1})) = \rho_i(\lambda_i - \mu_i)^2 \hat{g}_i(t) p_{i-1}^{j-1}(t_i).$$

*Proof.* The proof follows from Theorem 2.3 by noting that  $p_i^j(t)$  is the probability of observing a tree with one sampled individual.  $\square$

**Remark 2.5.** For  $i = 0$ , we have

$$p_1^j(t) = \frac{\rho_0(\lambda_0 - \mu_0)^2 e^{-(\lambda_0 - \mu_0)t}}{(\lambda_0 \rho_0 + ((1 - \rho_0)\lambda_0 - \mu_0)e^{-(\lambda_0 - \mu_0)t})^2},$$

which was established in Remark 3.2 in ref. 16.

The density of a sampled tree with root edge,  $\mathcal{T}_{or}$ , given the time of origin,  $f[\mathcal{T}_{or} | t_{or} = x_0]$ , can be calculated by recursively using the formula for  $g_{e,i}(t)$ , starting with  $t = x_0$ .

For obtaining a closed-form solution, note that each  $x_i$  with  $i > 0$  is twice a starting point and once an ending point of an edge.  $x_0$  is only a starting point. The leaves are only ending points.

Further, for  $t_{i+1} > t \geq t_i$ , define  $l(t) := i$ . Further, let  $n_i$  be the number of edges in  $\mathcal{T}_{or}$  that go through time  $t_i$ . Using this notation, we obtain

$$\begin{aligned} f[\mathcal{T}_{or} | t_{or} = x_0] &= \frac{1}{\lambda_l(x_0)} \rho_0^n (\lambda_0 - \mu_0)^{2n} \prod_{i=0}^{n-1} \lambda_{l(x_i)} \hat{g}_{l(x_i)}(x_i) \prod_{i=1}^m \left( \frac{\rho_i \hat{g}_{i-1}(t_i)}{\hat{g}_i(t_i)} \right)^{n_i} \\ &= \frac{1}{\lambda_l(x_0)} \rho_0^n (\lambda_0 - \mu_0)^{2n} \prod_{i=0}^{n-1} \lambda_{l(x_i)} \hat{g}_{l(x_i)}(x_i) \prod_{i=1}^m (\rho_i(\lambda_i - \mu_i)^2 \hat{g}_{i-1}(t_i))^{n_i}. \end{aligned}$$

Therefore, we established the following theorem.

**Theorem 2.6.** *The density of a sampled tree with root edge,  $\mathcal{T}_{or}$ , conditioned on the time of origin being  $x_0$ , is*

$$f[\mathcal{T}_{or} | t_{or} = x_0] = \frac{1}{\lambda_l(x_0)} \rho_0^n (\lambda_0 - \mu_0)^{2n} \prod_{i=0}^{n-1} \lambda_{l(x_i)} \hat{g}_{l(x_i)}(x_i) \prod_{i=1}^m (\rho_i(\lambda_i - \mu_i)^2 \hat{g}_{i-1}(t_i))^{n_i}.$$

**2.3) Probability density of a tree conditioned on the time of the most recent common ancestor.** We next provide the density for an oriented tree conditioned on the time of the most recent common ancestors of the extant sampled species. This formula is used for the maximum-likelihood parameter estimation in phylogenies.

**Theorem 2.7.** *The density of a sampled tree  $\mathcal{T}$ , conditioned on the time of the most recent common ancestor being  $x_1$ , is*

$$\begin{aligned} f[\mathcal{T} | t_{mrca} = x_1] &= \rho_0^n (\lambda_0 - \mu_0)^{2n} \frac{\hat{g}_{l(x_1)}(x_1)^2}{(1 - \rho_0^{l(x_1)}(x_1))^2} \prod_{i=2}^{n-1} \lambda_{l(x_i)} \hat{g}_{l(x_i)}(x_i) \prod_{i=1}^m (\rho_i(\lambda_i - \mu_i)^2 \hat{g}_{i-1}(t_i))^{n_i}. \end{aligned}$$

*Proof.* When conditioning on the time of the most recent common ancestor of the sampled species, we implicitly condition that the two descendants at the time of the most recent common ancestor have sampled descendants, and we call  $S$  the event that a species has at least one sampled extant descendant. Let  $\mathcal{T}_{or}^L$  be the left subtree of  $\mathcal{T}$  and  $\mathcal{T}_{or}^R$  be the right subtree of  $\mathcal{T}$  descending from the most recent common ancestor of  $\mathcal{T}$ . So

$$\begin{aligned}
f[T|t_{mrca} = x_1] &= f[T_{or}|t_{or} = x_1, S] f[T_{or}^L|t_{or} = x_1, S] \\
&= \frac{f[T_{or}^L|t_{or} = x_1] f[T_{or}^L|S|t_{or} = x_1]}{(f[S|t_{or} = x_1])^2} \\
&= \frac{f[T_{or}^L|t_{or} = x_1] f[T_{or}^L|S|t_{or} = x_1]}{(1 - p_0^{I(x_1)}(x_1))^2}.
\end{aligned}$$

This establishes the theorem.  $\square$

**3) Implementation of the Maximum-Likelihood Method.** The probability density in Theorem 2.7 is used for maximum-likelihood rate estimation in phylogenetic trees inferred from (species) data. A model with  $m$  rate shifts (or mass extinction events) can be tested against a model with  $m + 1$  rate shifts (or mass extinction events) using the likelihood-ratio test. The method is implemented in the R package TreePar (18). For the maximization of the likelihood function, I use the R package subplex (34) available on CRAN. The likelihood estimations were performed on the ETH Brutus cluster.

The maximization performs well for fixed rate shift times, but terminates in local optima instead of the global optimum when estimating simultaneously the rate shift times and the speciation and extinction rates. I therefore use a fine grid on shift times and optimize the speciation and extinction rates for the fixed rate shift times on the grid.

When analyzing the mammalian phylogeny, I first allow for one shift and estimate rates in 0.1-Myr steps between 2.05 and 100.05 Mya. Then I allow for two shifts by fixing the maximum-likelihood time shift from the one-shift estimation and using for the second shift again a grid of 0.1-Myr steps be-

tween 2.05 and 100.05 Mya. I proceed in this way to add more shifts. Note that estimating all time shifts simultaneously would be favorable over the greedy approach, but it is computationally not feasible. I show through simulations that the greedy approach recovers the rate shifts reliably (SI Text S2).

I showed in a previous paper (17) that it is not possible to distinguish a mass extinction event from constant rates interrupted by a period of stasis (i.e., basically no speciation and extinction), as both events leave a similar footprint on the phylogeny. Distinguishing between mass extinction and rate shift events requires additional paleontological data. If only the phylogenetic tree is available, then for each time interval, we need to know one of the three parameters  $\lambda_{it}$ ,  $\mu_{it}$  and  $\rho_i$  or we have to assume dependencies, like  $\lambda_i = \lambda_{i-1}$ . Note that requiring one of the three parameters  $\lambda_{it}$ ,  $\mu_{it}$  and  $\rho_i$  to be known is already necessary if no shift and no mass extinction occur (35).

In the mammal phylogeny, I investigate whether and when rate shift events occurred; i.e., I set  $\rho_i = 1$  for all  $i$ . I investigate the accuracy of the maximum-likelihood rate estimates through simulating trees with  $n$  species and then estimating their rates. For the simulations, I use the R package TreeSim (36).

Trees with polytomies can be analyzed with my method. Polytomies are considered by the method as a random binary resolution with edges of length 0 (note that each binary resolution of the polytomy has the same likelihood).

**ACKNOWLEDGMENTS.** I thank Alexandre Antonelli, Marcelo Sanchez, Helen Alexander, Jan Engelstaedter, Roland Regoes, Olin Silander, the editor, and the two anonymous reviewers for very helpful comments. Funding for this work came from Eidgenössische Technische Hochschule, Zurich.

- Asher RJ, et al. (2005) Stem Lagomorpha and the antiquity of Glires. *Science* 307: 1091–1094.
- Wible JR, Rougier GW, Novacek MJ, Asher RJ (2007) Cretaceous eutherians and Laurasian origin for placental mammals near the K/T boundary. *Nature* 447: 1003–1006.
- Bininda-Emonds OR, et al. (2007) The delayed rise of present-day mammals. *Nature* 446:507–512.
- Wilson D, Reeder D (2005) *Mammal Species of the World: A Taxonomic and Geographic Reference* (Johns Hopkins Univ Press, Baltimore).
- Sanderson MJ, Donoghue MJ (1996) Reconstructing shifts in diversification rates on phylogenetic trees. *Trends Ecol Evol* 11:15–20.
- Rabosky DL (2006) Likelihood methods for detecting temporal shifts in diversification rates. *Evolution* 60:1152–1164.
- Nee S, Holmes EC, May RM, Harvey PH (1994) Extinction rates can be estimated from molecular phylogenies. *Philos Trans R Soc Lond B Biol Sci* 344:77–82.
- Paradis E (1997) Assessing temporal variations in diversification rates from phylogenies: Estimation and hypothesis testing. *Proc Biol Sci* 264:1141.
- Pybus OG, Harvey PH (2000) Testing macro-evolutionary models using incomplete molecular phylogenies. *Proc Biol Sci* 267:2267–2272.
- Ricklefs RE (2007) Estimating diversification rates from phylogenetic information. *Trends Ecol Evol* 22:601–610.
- Harvey PH, May RM, Nee S (1994) Phylogenies without fossils. *Evolution* 48:523–529.
- Kendall DG (1948) On the generalized “birth-and-death” process. *Ann Math Stat* 19: 1–15.
- Bailey N (1964) *The Elements of Stochastic Processes with Applications to the Natural Sciences* (Wiley, New York).
- Nee SC, May RM, Harvey PH (1994) The reconstructed evolutionary process. *Philos Trans R Soc Lond B Biol Sci* 344:305–311.
- Gernhard T (2008) The conditioned reconstructed process. *J Theor Biol* 253:769–778.
- Stadler T (2010) Sampling-through-time in birth-death trees. *J Theor Biol* 267: 396–404.
- Stadler T (2011) Simulating trees on a fixed number of extant species. *Syst Biol*, in press.
- Stadler T (2011) TreePar in R - Estimating diversification rates in phylogenies. Available at <http://cran.r-project.org/web/packages/TreePar/index.html>. Accessed January 25, 2011.
- Nee SC (2001) Inferring speciation rates from phylogenies. *Evolution* 55:661–668.
- Rabosky DL (2010) Extinction rates should not be estimated from molecular phylogenies. *Evolution* 64:1816–1824.
- Quental TB, Marshall CR (2010) Diversity dynamics: Molecular phylogenies need the fossil record. *Trends Ecol Evol* 25:434–441.
- Barnosky A, Carrasco M (2002) Effects of Oligo-Miocene global climate changes on mammalian species richness in the northwestern quarter of the USA. *Evol Ecol Res* 4: 811–841.
- Finarelli J, Badgley C (2010) Diversity dynamics of Miocene mammals in relation to the history of tectonism and climate. *Proc R Soc B* 277:2721–2726.
- Barnosky A, Hadly E, Bell C (2003) Mammalian response to global warming on varied temporal scales. *J Mammal* 84:354–368.
- Zachos JC, Dickens GR, Zeebe RE (2008) An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* 451:279–283.
- Swenson MS, Barbançon F, Warnow T, Linder CR (2010) A simulation study comparing supertree and combined analysis methods using SMIDGen. *Algorithms Mol Biol* 5:8.
- Turgeon J, Stoks R, Thum RA, Brown JM, McPeck MA (2005) Simultaneous Quaternary radiations of three damselfly clades across the Holarctic. *Am Nat* 165:E78–E107.
- Crisp MD, Cook LG (2009) Explosive radiation or cryptic mass extinction? Interpreting signatures in molecular phylogenies. *Evolution* 63:2257–2265.
- Alroy J (2000) New methods for quantifying macroevolutionary patterns and processes. *Paleobiology* 26:707–733.
- Rabosky DL (2009) Heritability of extinction rates links diversification patterns in molecular phylogenies and fossils. *Syst Biol* 58:629–640.
- Rabosky DL, Lovette IJ (2008) Density-dependent diversification in North American wood warblers. *Proc Biol Sci* 275:2363–2371.
- Ford D, Matsen FA, Stadler T (2009) A method for investigating relative timing information on phylogenetic trees. *Syst Biol* 58:167–183.
- King A, Billingham J, Otto S (2003) *Differential Equations: Linear, Nonlinear, Ordinary, Partial* (Cambridge Univ Press, Cambridge, UK).
- King A (2008) The subplex algorithm for unconstrained optimization. Available at <http://cran.r-project.org/web/packages/subplex/index.html>. Accessed July 28, 2008.
- Stadler T (2009) On incomplete sampling under birth-death models and connections to the sampling-based coalescent. *J Theor Biol* 261:58–66.
- Stadler T (2010) TreeSim in R - Simulating trees under the birth-death model. Available at <http://cran.r-project.org/web/packages/TreeSim/index.html>. Accessed February 23, 2010.