



HAL
open science

Time-dependent speciation and extinction from phylogenies: a least squares approach

Emmanuel Paradis

► **To cite this version:**

Emmanuel Paradis. Time-dependent speciation and extinction from phylogenies: a least squares approach. *Evolution - International Journal of Organic Evolution*, 2011, 65 (3), pp.661 - 672. 10.1111/j.1558-5646.2010.01179.x . hal-01822128

HAL Id: hal-01822128

<https://hal.science/hal-01822128>

Submitted on 23 Jun 2018

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 destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Running head: TEMPORAL DYNAMICS OF DIVERSIFICATION

TIME-DEPENDENT SPECIATION AND EXTINCTION FROM PHYLOGENIES: A LEAST SQUARES APPROACH

Emmanuel Paradis

*Institut de Recherche pour le Développement, UR175 CAVIAR, BP 5095, 361 rue
Jean-François Breton, F-34196 Montpellier Cédex 5, France*

E-mail: Emmanuel.Paradis@ird.fr

tel: +33 (0)4 67 16 64 47 fax: +33 (0)4 67 16 64 40

Molecular phylogenies contribute to the study of the patterns and processes of macroevolution even though past events (fossils) are not recorded in these data. In this paper, I consider the general time-dependent birth–death model in order to fit any model of temporal variation in speciation and extinction to phylogenies. I establish formulae to compute the expected cumulative distribution function of branching times for any model, and, building on previous published works, I derive maximum likelihood estimators. Some limitations of the likelihood approach are described, and a fitting procedure based on least squares is developed that alleviates the shortcomings of maximum likelihood in the present context. Parametric and nonparametric bootstrap procedures are developed to assess uncertainty in the parameter estimates, the latter version giving narrower confidence intervals and being faster to compute. I also present several general algorithms of tree simulation in continuous time. I illustrate the application of this approach with the analysis of simulated data sets, and two published phylogenies of primates (Catarrhinae) and lizards (Agamidae).

KEY WORDS: Birth–death, diversification, extinction, phylogeny, speciation.

The statistical analysis of diversification using phylogenetic trees has progressed considerably during the last decade. New insights have been given into the estimation of speciation and extinction rates (Nee 2001; Ricklefs 2007), quantifying the temporal decline in diversification following an adaptive radiation (Rabosky and Lovette 2008), linking traits to increased or decreased diversification (Paradis 2005; Ree 2005; Maddison et al. 2007; Moore and Donoghue 2009), as well as more general issues like community assembly (McPeck 2008). A notable feature of this progress is the move from purely correlative methods (such as sister-group comparisons) towards approaches based on explicit models of diversification and likelihood-based inference (for a comparison, see Rabosky 2006). In spite of these advances, two major issues remain unsolved.

First, how accurately can temporal variation in diversification be quantified? Different evolutionary mechanisms imply different temporal patterns of changes in speciation and extinction rates. For instance, an increasing speciation rate may be the result of some ecological niches being vacant or from population fragmentation, whereas an increasing extinction rate may be a consequence of high interspecific competition, habitat destruction, or decreased genetic variability. In real situations, these different mechanisms will combine to create evolutionary scenarios. For instance, an adaptive radiation may be characterized by a very high speciation rate at or near the origin of a clade that decreases subsequently (Schluter 2000; Gavrillets and Losos 2009). On the other hand, a phylogenetic fuse process is characterized by an initial low speciation rate during some time that increases later (Cooper and Fortey 1998). A clade expands, on average, if speciation rate is greater than extinction rate, whereas it declines if this is the opposite (Fig. 1). Moreover, for a fixed difference in these rates, species compositions will show a high turn-over if both rates are high. Since the contrast between speciation and extinction rates is likely to change through time, a clade will move through different phases of expansion, decline, or high/low turn-over. These different evolutionary

mechanisms may act separately or in combination, resulting in potentially complex evolutionary diversification scenarios. It appears clearly that a statistical modeling approach must take a central place in studies of diversification. Only the confrontation of alternative models will be able to test the adequacy of complex scenarios of diversification to phylogenetic data.

Second, how accurately can in diversification be disentangled in its components of speciation and extinction? This question has remained elusive since some studies emphasized the difficulties in estimating accurately extinction rate (Kubo and Iwasa 1995; Paradis 2004) while others emphasized the importance of considering extinction explicitly in the estimation procedure (Nee et al. 1994a; Nee 2001). Recently, Quental and Marshall (2010) demonstrated that the estimates of extinction rates in cetaceans inferred from a molecular phylogeny are greatly underestimated compared to the same estimates inferred from the fossil record. Their study is of great importance because most groups of animals and plants do not have a sufficiently documented fossil record to be able to make such a comparison between molecular and fossil data.

In one of the first contributions to the statistical analysis of diversification, Nee et al. (1994b) described how to do likelihood-based inference of temporal variation in speciation and extinction rates, but their approach does not seem to have been implemented. In a previous paper, I developed a method to test for temporal variation in diversification rate, but it can consider only a continuous increase (or decrease) through time, and speciation and extinction rates cannot be distinguished (Paradis 1997). Pybus and Harvey (2000) introduced a simple statistic, the γ -statistic, to test whether diversification has increased or decreased through time. This test has been widely used in the literature. Finally, Rabosky and Lovette (2008) implemented Nee et al.'s (1994b) likelihood approach and applied it to study decreasing diversification with several models; in particular, they considered a decline in speciation rate, an increase in extinction rate, and a combination of both. These authors suggested that this likelihood

approach can be generalized to any time-dependent model, though they did not show other examples of such models. Therefore, whether we can fit more complex models of diversification is still an open question.

In this paper, I introduce several new analytical tools for the analysis of temporal variation in speciation and extinction. I derive formulae to compute the cumulative distribution function of branching times for any time-dependent birth–death model. I introduce a new method of model fitting based on least squares, and compare it with maximum likelihood. I also introduce several algorithms to simulate phylogenies under such models. This approach is illustrated with simulated and real data.

Time-Dependent Birth–Death Models

The data under consideration are a dated phylogenetic tree with N species. We assume that this tree is rooted and fully dichotomous; consequently, it has $N - 1$ ($= m$) internal nodes. By convention, the root is the origin of the time scale, and the node times, measured from the root, are denoted t_i ($i = 1, \dots, m$) with $t_1 = 0$, and the present is T .

Birth–death models are a class of continuous-time branching models where each species (or individual) is continuously exposed to speciation (birth) and extinction (death). The time-dependent birth–death model (or generalized birth–death model in the terminology used in Kendall (1948)) assumes that the probability of a speciation event varies with time and follows a function $\lambda(t)$, and similarly for the probability of extinction $\mu(t)$. This formulation implies no lineage-specific variation: at a give time, all species have the same probabilities of speciation and extinction.

Under such a model, we can write the probability that a lineage, originating from a single species at time t , has exactly one descendant at time T :

$$\Pr(t, n_T = 1) = \frac{e^{-\rho(t,T)}}{W(t)^2},$$

with

$$\rho(t, T) = \int_t^T [\mu(u) - \lambda(u)] du,$$

and

$$W(t) = e^{-\rho(t, T)} \left[1 + \int_t^T e^{\rho(t, u)} \mu(u) du \right].$$

Similarly, the probability of a lineage being extinct during the same time interval is

$$\Pr(t, n_T = 0) = 1 - \frac{e^{-\rho(t, T)}}{W(t)},$$

from which we can write the probability that a lineage is not extinct:

$$\Pr(t, n_T \geq 1) = 1 - \Pr(t, n_T = 0) = \frac{e^{-\rho(t, T)}}{W(t)}. \quad (1)$$

For simplification, let $W(t) = e^{-\rho(t, T)} [1 + I(t)]$ with:

$$I(t) = \int_t^T e^{\rho(t, u)} \mu(u) du.$$

It appears then that all the above equations depend only on two functions: $\rho(t, T)$ and $I(t)$. Therefore once we have specified $\lambda(t)$ and $\mu(t)$, we may derive these two functions. This is straightforward for $\rho(t, T)$, especially if primitives are available for $\lambda(t)$ and $\mu(t)$; otherwise this integral can be calculated numerically. For $I(t)$, it appears that the latter method is needed in all cases. Appendix 1 shows how these functions simplify when speciation and extinction probabilities are constant through time.

In order to fit a model of time-dependent diversification, one has to specify a parametric function for $\lambda(t)$ and $\mu(t)$. It is common to assume monotonically increasing or decreasing functions, in which case a logistic model is convenient, for instance for

speciation:

$$\lambda(t) = \frac{1}{1 + e^{-(\beta_\lambda t + \alpha_\lambda)}}, \quad (2)$$

where α_λ and β_λ are two parameters. Equation 2 is an increasing function with respect to time if $\beta_\lambda > 0$, or decreasing if $\beta_\lambda < 0$; if $\beta_\lambda = 0$ it reduces to a constant probability $\lambda = 1/(1 + e^{-\alpha_\lambda})$. The function $\mu(t)$ can be written similarly with parameters α_μ and β_μ . Another advantage of the logistic model is that the primitives of equation 2 can be used to accelerate the computations. However, the present approach is not limited to monotonically changing functions. Step functions or linearly changing functions with multiple break-points are also convenient because their primitives are easily calculated with simple geometry. Furthermore these functions may lead to biologically interesting scenarios. A step model with a single break-point for speciation probability can be defined with:

$$\lambda(t) = \begin{cases} \lambda_1 & t \leq \tau_\lambda \\ \lambda_2 & t > \tau_\lambda. \end{cases}$$

This model has three parameters (λ_1 , λ_2 , and τ_λ). A step model with two break-points would have five parameters (see application below).

Model Fitting to the Distribution of Branching Times

Since the probabilities of observing the different elements of a phylogeny have been written down, it is possible to develop a likelihood function following the lines of Nee et al. (1994b):

$$\ln L = 2\rho(t_1, T) - 4\ln[1 + I(t_1)] + (m - 1)\rho(0, T) + (m - 1)\ln 2 + \sum_{i=2}^m \{-2\rho(0, t_i) + \ln \lambda(t_i) - 3\ln[1 + I(t_i)]\}.$$

The details of this derivation are provided in the Supplementary Information.

It appears from this likelihood function that the function of extinction μ appears only in the integral ρ and in the product with the exponential of ρ . On the other hand, the function of speciation λ appears as a single term. I made several attempts to implement this likelihood but this always led to meaningless estimates, suggesting it is degenerate in some way or another. This observation does not prove that it is generally impossible to fit a time-dependent model by maximum likelihood, but Appendix 2 presents a theorem suggesting that it is impossible to prove that, in general, time-dependent models can be fitted by maximum likelihood.

An alternative to a likelihood approach is to consider the distribution of branching times from the phylogeny. Since at a given time t , speciation and extinction probabilities are homogeneous (they have the same values for all species), branching times summarize all the information on diversification of the clade. The probability of observing a branching event in a phylogeny, denoted by $\pi(t)$, is given by the probability that two lineages originating at time t survive until present multiplied by the probability of a speciation event at t . The latter is given by the product of the speciation probability at t with the number of species living at this time. Obviously, we do not know the latter but we can substitute it by its expectation: $E(n_t) = 2e^{-\rho(0,t)}$. Consequently, we have:

$$\pi(t) = \Pr^2(t, n_T \geq 1)E(n_t)\lambda(t).$$

We rescale these quantities with respect to t on the interval $[0, T]$ to obtain a cumulative density function (CDF) of the branching times:

$$\mathcal{F}(t) = \int_0^t \pi(x)dx / \int_0^T \pi(x)dx. \quad (3)$$

It is then possible to compare this theoretical CDF with the ECDF (empirical CDF) from the tree and find the parameter values that minimize the deviation of the former on

the latter using a least squares criterion to take all information along the tree into account. The ECDF of branching times is similar to a standardized lineage-through-time (LTT) plot where both x - and y -scales are scaled on $[0, 1]$. Considering that the points of the ECDF are equally likely to be above or below the curve of the theoretical CDF, a least squares criterion is justified and makes possible hypothesis testing by comparing different models with an F -test as a ratio of sums of squares (SS) with the most parsimonious model on the numerator. The degrees of freedom of this test are given by the number of additional parameters, and the residual degrees of freedom ($m - \text{total number of parameters}$). The simulation results below show that this test does not follow exactly an F distribution so that the type I error rate is inflated. So a correction is applied by dividing the ratio of sums of squares by the number of additional parameters. This has the effect of reducing statistical power, but the type I error rate is kept at a reasonable level (see below). For a given model, uncertainty of the parameter estimates can be assessed with a bootstrap, either parametric or nonparametric (Efron and Tibshirani 1993). Nonparametric bootstrapping is done by sampling with replacement the observed branching times and estimating the parameters of the model for each bootstrap sample. In the parametric version, random samples of size N are simulated from the estimated CDF. In both cases, the standard-errors of the parameters are calculated with the standard deviation of the bootstrap estimates. The proposed SS-based approach is compared to maximum likelihood below using a simulated data set.

In most studies, it is possible to consider a priori different hypotheses about the diversification of a given clade. In practice, this implies selecting different functions $\lambda(t)$ and $\mu(t)$ that quantify these different hypotheses, choosing some parameter values, calculating $\mathcal{F}(t)$ for these different hypotheses, and comparing them. Figure 2A shows four hypothetical scenarios of change in diversification with different patterns for speciation and extinction: some of them would result in CDF of branching times that are very close (Fig. 2B). Figure 3 shows a more striking example with two very different

scenarios that would result in very similar CDFs of branching times. These results are further discussed below.

The present approach has been implemented in R (?) and the package *ape* (Paradis et al. 2004). Additional code has been developed and is available upon request.

Simulation of Phylogenies in Continuous Time

The formulae derived above allow us to write general algorithms for simulating phylogenetic trees under any time-dependent birth–death model. Three algorithms are described in Appendix C of the Supplementary Information. The general idea behind these algorithms is the following: start the simulation from the root, and consider successively whether each lineage survives until T using equation 1. If it survives, consider whether there has been a branching event on this branch before T using appropriate equations; if yes, then apply the same procedure recursively. The difference between algorithms 1 and 2 is that the latter ignores lineages that do not survive until T , thus producing a “pruned” ultrametric phylogeny (algorithm 1 produces a “complete” phylogeny). Algorithm 3 generates a random sample of branching times (i.e., without topology) from a theoretical CDF using equation 3 and the reverse method.

These algorithms are implemented in R so that the user can easily specify functions for the temporal variation in speciation and extinction rates (some examples are provided in *ape*).

Simulation Study

Figure 4A shows a tree simulated with constant probabilities $\lambda = 0.13$ and $\mu = 0.05$. The model fitted assuming constant probabilities gave $\hat{\lambda} = 0.106$ and $\hat{\mu} = 0.004$ (SS = 13.322, Fig. 4B). A time-dependent model with both probabilities following equation 2 gave SS = 12.448 (Fig. 4C). This second model fits better to the data but this is expected because it has two additional parameters. The test comparing them is $F_{2,129} = 0.535$ ($P = 0.587$),

leading to the acceptance of the constant-probability model. The nonparametric bootstrap standard-errors for the first model based on 1000 bootstrap samples were $SE(\hat{\lambda}) = 0.014$ and $SE(\hat{\mu}) = 0.013$. However, when looking at the distributions of the individual bootstrap estimates it appeared that they were highly skewed, so uncertainty in the estimation is better represented by computing the 0.025 and 0.975 quantiles of the bootstrap estimates in order to derive 95% confidence intervals. These intervals were [0.010, 0.130] and [0, 0.034] for $\hat{\lambda}$ and $\hat{\mu}$, respectively. The parametric bootstrap gave, for the same parameters, [0.031, 0.140] and [0, 0.051]. Henceforth I report only the confidence intervals using the nonparametric version which are easier to compute.

In order to assess the bias of the estimators and the type I error rate of the F -test, I simulated 1000 phylogenies with the same parameters ($\lambda = 0.13$, $\mu = 0.05$) using algorithm 2. The number of species in these trees ranged between 9 and 879 (median = 153, mean = 181.3). Both models were fitted to each phylogeny and the F -test was computed. The observed type I error rate (proportion of tests where the null model, which is correct here, was rejected at the nominal level 0.05) was 0.083. The results are detailed in the Supplementary Information. There was no relationship between sample size (number of species) and the rate of rejection of the null hypothesis (Fig. S3). The estimator of λ showed a slight negative bias: the mean of the 1000 estimates was 0.1148 (min = 0.0417, max = 0.1937, median = 0.1157). This bias was not related to sample size; however, the variance of the estimates decreased with increasing sample size (Fig. S4). On the other hand, the estimator of μ showed systematic and considerable variance (Fig. S5). This variance decreased with increasing number of species, although the bias remained constant with a median value of $\hat{\mu}$ close to zero. This appears as a discouraging result because the trees with many species did not yield more accurate estimates of μ . This is particularly contrasting with the fact λ was more accurately estimated with the large trees. The two estimates $\hat{\lambda}$ and $\hat{\mu}$ showed strong covariation (Fig. S6). Interestingly, this line of covariation seems to converge to the point of the real

parameters.

Figure 5A shows a tree simulated with time-dependent probabilities according to a logistic model as shown on Fig. 5B. The model fitted with constant rates gave $\hat{\lambda} = 0.268$ and $\hat{\mu} = 0.198$, and the SS was 246.654. The model with time-dependent rates gave an SS value of 15.973. The test comparing these two models is thus $F_{2,127} = 16.14$ ($P < 0.001$). Figure 5B also displays the estimated curve of speciation and extinction probabilities according to the time-dependent rates model. For this particular simulated tree, I plotted the likelihood function and the SS curve around the real values of the extinction parameters ($\beta_{\mu} = 0.02$, $\alpha_{\mu} = -2$). The likelihood function displayed no local or global maximum, and decreased continuously while the parameter values tend towards $-\infty$. On the other hand, the SS curve displayed a global minimum with the parameter estimates $\hat{\beta}_{\mu} = 0.088$ and $\hat{\alpha}_{\mu} = -5.465$ (Fig. 6). The profile of the SS curve is much shallower with respect to the extinction parameters than to the speciation ones. This is reflected by wider confidence intervals for the former than for the latter (see legend of Figure 5 and details in SI). The confidence interval of $\hat{\alpha}_{\mu}$ is remarkably large, and the estimated value is close to the upper bound of this interval. This surprising result is due to the fact that the estimation of this parameter can be greatly influenced by a few observations, so it varies accordingly among bootstrap samples. This is confirmed by the simulation results below.

As in the constant-probability case, I simulated 1000 phylogenies with the same four-parameter model and followed the protocol of model fitting and testing as above. The number of species (N) in these simulated trees ranged between 5 and 690 (median = 115.5, mean = 138.1). The null hypothesis was rejected in 477 cases, so the power of the test was estimated to be 0.477. Surprisingly, there was no relationship between sample size and P -value: the rate of rejection of the null hypothesis was similar with small or large phylogenies (Fig. S7). The estimates of the slope of the logistic model for speciation (β_{λ}) showed a positive bias for the smallest trees, but this bias converged to

zero with increasing N and was nearly null for $N > 200$ (Fig. S8). The variance of these estimates also decreased with increasing sample size. The estimates of β_μ were also positively biased but with higher variance and it converged only partially to zero with increasing N ; however, the variance decreased considerably for $N > 150$ (Fig. S9). The estimates of α_μ varied greatly and their distribution was highly skewed: 996 estimated values were less than the true value (median = -7.78). This is consistent with the bootstrap results in the previous paragraph. This suggests care in the interpretation of the estimation of this parameter. Encouragingly, the bias and dispersion of these estimates seem to decrease with increasing N (Fig. S7).

Applications

Two data sets were considered: the phylogeny of Catarrhini (Old World monkeys) taken from Purvis (1995), and the phylogeny of Agamidae taken from ?. These trees have 97 and 69 species, respectively. They have been chosen in order to provide relatively homogeneous clades since large phylogenies (e.g., supertrees) are likely to include several clades with different parameters of diversification (Purvis et al. 1995; Paradis 1998, 2005). These two clades of primates and reptiles have attracted attention from evolutionists because they are considered as adaptive radiations (Purvis et al. 1995; ?).

As emphasized above, the choice of models is crucial when testing hypotheses. In the simulation study, I only considered two models because the objective was to assess the properties of the estimators and the tests. With the present applications, I fitted several models in order to quantify the temporal variations in speciation and extinction. Simple models were fitted first, then more complicated models depending on the results. An attempt was made to model temporal variation in extinction even though it is anticipated that this will be difficult.

Six models were fitted to the Catarrhine data set (Table 1). All tests comparing the different models with the constant-parameter model were not significant, leading to the

conclusion that diversification occurred at a constant rate for these primates.

Interestingly, the model assuming that extinction probability follows a step model (4) fits better than the null model (1), though not significantly; in that case the extinction probability before 3.02 Ma was estimated to be 0.109. All models suggest a slight increase in diversification through time, though not statistically significant (Table 2). For the sake of comparison, I calculated the γ -statistic (which follows a standard normal distribution under the null hypothesis of constant diversification; Pybus and Harvey 2000) for this data set. The result was not statistically significant ($\gamma = -0.94$, $P = 0.344$) further indicating that the Catarrhini diversified at a constant rate.

Five models were fitted to the Agamid data set (Table 3). The two models with time-varying speciation probability fitted significantly better than the null model ($P < 0.001$ in both cases). Among these two models with a logistic model (2) or a step model (3), the former fitted better though it has one parameter less than the latter. None of the time-varying extinction probability models fitted better than the null model. These results provide evidence for a continuous decrease in speciation rate for the agamid lineage (Table 4). The γ -statistic also indicated that diversification rate in this lineage was not constant ($\gamma = -4.49$, $P < 0.001$).

Discussion

Molecular phylogenies (actually all phylogenies estimated with recent species) are valuable sources of information in the study of macroevolutionary processes. However, the fact that extinctions are not observed, and so must be inferred from the distribution of branching times is a difficulty that has remained essentially unsolved because the uncertainty caused by extinctions cannot be quantified easily (Harvey et al. 1994; Kubo and Iwasa 1995; Paradis 2003, 2004). It is intuitive that the absence of past events implies some biases in the inference of diversification rates. Different scenarios of temporal variation in speciation and extinction will result in different expected

distributions of branching times; however, some of these scenarios will be very difficult to distinguish because these distributions will be very close. One explanation for this phenomenon is that the mean values of speciation and extinction probabilities over time are relevant instead of the exact temporal variation because of the integration over time. Nevertheless, it is not completely clear whether these different branching times distributions may be completely undistinguishable, or whether they are actually unique and could be distinguished with sufficiently large data sets. This question needs further work to be clarified.

The simulation study presented here shows some interesting results. In the case of a constant-probability model, there is a moderate bias in the estimator $\hat{\lambda}$ though this bias does not seem to be affected by the size of the tree. It is not clear what is the cause of this bias. The bias in the estimator $\hat{\mu}$ is large and this estimator has a large variance as well. The type I error rate (probability of wrongly rejecting the null hypothesis) is low ($\approx 8\%$). On the other hand, the same test performed on phylogenies simulated with a time-dependent model had a moderate power ($\approx 50\%$) even though the distribution of the F -test is not completely determined. Further work on this point is clearly needed to improve the performance of the test (increase power while keeping the type I error rate at a low level). In the case of time-dependent speciation and extinction probabilities, the estimators of the slopes $\hat{\beta}_\lambda$ and $\hat{\beta}_\mu$ were positively biased, but this bias converged towards zero with increasing tree size. As in the constant-probability model, the variance of the estimates related to extinction ($\hat{\beta}_\mu$) were greater than for speciation. It is clear that the range of scenarios considered in this paper is restricted, and it will be necessary to assess these statistical properties for a wider range of situations.

Two previous studies showed that when speciation and extinction probabilities are both constant, the extinction probability estimator may be very imprecise (Kubo and Iwasa 1995; Paradis 2004). Particularly, unbiased estimation of a constant extinction probability can be achieved only when this parameter is close to zero and the age of the

root of the reconstructed tree is close to the age of the actual ancestor of the lineage, in other words when the tree without dead lineages is close to the tree with them (Paradis 2004). Recently, Rabosky (2009a, 2010) showed that when extinction rate is heterogeneous among species or through time some biases in the estimation of extinction rate are also observed. It must be emphasized that even when extinction rate is constant, its estimation is biased in most cases.

Rabosky (2009b) developed a model of constant diversity and heritable extinction rates (further discussed below) which can produce phylogenies with estimated extinction rates equal to zero. However, this may be a special case of a more general explanation: if some species have higher extinction risk (i.e., if extinction rate is heterogeneous among species), they are less likely to be present in phylogenies of living species. Consequently, these phylogenies provide a biased sample including a majority of the species with the lowest extinction risk. This more general explanation does not require constant diversity or heritable extinction risk. Various other explanations have been proposed to explain the failure to estimate extinction rates (e.g., lack of appropriate models or of fitting methods), but it could be simply that since extinctions are not observed in phylogenies of recent species, estimating the rate at which these events occur is not generally possible.

Temporal variation in diversification parameters may create patterns of branching times similar to a constant-rate model. It is crucial to appreciate this point because many relevant hypotheses about phylogenetic diversification consider temporal variation in speciation and extinction probabilities (Nee et al. 1992; Cooper and Fortey 1998; Bininda-Emonds et al. 2007; Rabosky and Lovette 2008). The striking example illustrated on Fig. 3 can be understood as both scenarios have the same mean speciation probabilities over time; consequently, the integrals computed above will be very close. Therefore, it may be expected that complex scenarios of temporal change in diversification rates will be difficult to test. Future studies on phylogenetic diversification will need to consider this point.

There are a large number of methods to deal with model uncertainty in the statistical literature (e.g., Claeskens and Hjort 2008) but it does not seem easy to apply them to the present problem. Usually, techniques such as model averaging or multimodel inference are useful when the subject of interest is not the models themselves but another quantity we wish to estimate or predict. In the present context, the models are central to the question of interest since we want to know whether speciation or extinction changed through time. If we change our perspective and focus on other quantities which are products of speciation and extinction (e.g., diversity, morphological disparity, distribution, etc) then a multimodel approach might be particularly relevant because of this model indeterminacy. This raises many challenges though, such as how to tackle a potentially extremely vast model and parameter space. Some Bayesian methods may offer a way to attack this problem such as Occam's window (Madigan and Raftery 1994) where the model space is explored in a hierarchical way.

Although maximum likelihood seems the ideal method for fitting diversification models, alternative methods have been proposed like Bokma's (2006) neural networks. Rabosky (2009b) used approximate Bayesian computation (ABC) to fit a model of heritable extinction with pulsed turnover (HEPT). In this model, diversity is constant through time, so extinctions are balanced by speciations and the former are assumed to have correlated probabilities of occurrence among species. Consequently, the HEPT model does not belong to the class of birth–death models and its likelihood function cannot be derived easily (Rabosky 2009b). However, the HEPT model is particularly interesting to model diversification because it relaxes several assumptions of birth–death models. Since the ECDF of branching times of the HEPT model can be obtained, it is possible to fit this model by the least squares method presented here and compare it to alternative models. This is a potential interesting area of future research because alternative models of diversification have already been considered in the past, but it is difficult to compare them with standard birth–death models (see, e.g., Losos and Adler

1995).

Incomplete taxon sampling has been of interest in the literature on diversification (e.g., FitzJohn et al. 2009), particularly for extremely-rich clades where data will not be available for all species in the near future (e.g., Pie and Tschá 2009). It is possible to generalize the above formulae to this situation by realizing that such subsampling of species within a clade is similar to the sampling of individuals in a population where their genealogy is modelled by the coalescent (Kingman 1982). Therefore, the ECDF of branching times in the case of incomplete sampling may be obtained by superposing two processes: birth–death, forward in time, and coalescent, backward in time (Stadler 2009). Since the number of sampled species within a clade may not be negligible compared to the total number of species, exact formulae for the coalescent should be used in this situation (see Wakeley 2009).

The analysis of the simulated and real phylogenies showed examples with constant, declining, or increasing diversification. The results from the analysis of the Agamid data indicate that the decline in speciation probability (also found by Rabosky and Lovette 2008, with a maximum likelihood approach) was more likely to have been continuous than discontinuous as suggested by the better fit of the logistic model compared to the step model. It will be interesting to test if the same pattern can be observed in other groups since a continuous decline is more likely to be intrinsic (e.g., loss of genetic variation) whereas a discontinuous decline may be environmentally-driven.

Acknowledgments

I am grateful to Dan Rabosky for his comments on a previous version of this paper, and to two anonymous referees for their constructive comments. Financial support was provided by grant ANR-2006-BDIV002. This is publication IRD-DIVA-ISEM 2010-116.

LITERATURE CITED

- Bininda-Emonds, O. R. P., M. Cardillo, K. E. Jones, R. D. E. MacPhee, R. M. D. Beck, R. Grenyer, S. A. Price, R. A. Vos, J. L. Gittleman, and A. Purvis. 2007. The delayed rise of present-day mammals. *Nature* 446:507–512.
- Bokma, F. 2006. Artificial neural networks can learn to estimate extinction rates from molecular phylogenies. *J. Theor. Biol.* 243:449–454.
- Claeskens, G. and N. L. Hjort. 2008. *Model Selection and Model Averaging*. Cambridge University Press, Cambridge.
- Cooper, A. and R. Fortey. 1998. Evolutionary explosions and the phylogenetic fuse. *Trends Ecol. Evol.* 13:151–156.
- Efron, B. and R. J. Tibshirani. 1993. *An Introduction to the Bootstrap*. Chapman & Hall/CRC, Boca Raton, FL.
- FitzJohn, R. G., W. P. Maddison, and S. P. Otto. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Syst. Biol.* 58:595–611.
- Gavrilets, S. and J. B. Losos. 2009. Adaptive radiation: contrasting theory with data. *Science* 323:732–737.
- Harvey, P. H., R. M. May, and S. Nee. 1994. Phylogenies without fossils. *Evolution* 48:523–529.
- Kendall, D. G. 1948. On the generalized “birth-and-death” process. *Ann. Math. Stat.* 19:1–15.
- Kingman, J. F. C. 1982. On the genealogy of large populations. *J. Appl. Prob.* 19A:27–43.
- Kubo, T. and Y. Iwasa. 1995. Inferring the rates of branching and extinction from molecular phylogenies. *Evolution* 49:694–704.
- Losos, J. B. and F. R. Adler. 1995. Stumped by trees? A generalized null model for patterns of organismal diversity. *Am. Nat.* 145:329–342.

- Maddison, W. P., P. E. Midford, and S. P. Otto. 2007. Estimating a binary character's effect on speciation and extinction. *Syst. Biol.* 56:701–710.
- Madigan, D. and A. E. Raftery. 1994. Model selection and accounting for model uncertainty in graphical models using Occam's window. *J. Am. Statist. Assoc.* 89:1535–1546.
- McPeck, M. A. 2008. The ecological dynamics of clade diversification and community assembly. *Am. Nat.* 172:E270–E284.
- Moore, B. R. and M. J. Donoghue. 2009. A Bayesian approach for evaluating the impact of historical events on rates of diversification. *Proc. Natl. Acad. Sci. USA* 106:4307–4312.
- Nee, S. 2001. Inferring speciation rates from phylogenies. *Evolution* 55:661–668.
- Nee, S., E. C. Holmes, R. M. May, and P. H. Harvey. 1994a. Extinction rates can be estimated from molecular phylogenies. *Phil. Trans. R. Soc. Lond. B* 344:77–82.
- Nee, S., R. M. May, and P. H. Harvey. 1994b. The reconstructed evolutionary process. *Phil. Trans. R. Soc. Lond. B* 344:305–311.
- Nee, S., A. Ø. Mooers, and P. H. Harvey. 1992. Tempo and mode of evolution revealed from molecular phylogenies. *Proc. Natl. Acad. Sci. USA* 89:8322–8326.
- Paradis, E. 1997. Assessing temporal variations in diversification rates from phylogenies: estimation and hypothesis testing. *Proc. R. Soc. Lond. B* 264:1141–1147.
- Paradis, E. 1998. Detecting shifts in diversification rates without fossils. *Am. Nat.* 152:176–187.
- Paradis, E. 2003. Analysis of diversification: combining phylogenetic and taxonomic data. *Proc. R. Soc. Lond. B* 270:2499–2505.
- Paradis, E. 2004. Can extinction rates be estimated without fossils? *J. Theor. Biol.* 229:19–30.
- Paradis, E. 2005. Statistical analysis of diversification with species traits. *Evolution*

59:1–12.

- Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
- Pie, M. and M. K. Tschá. 2009. The macroevolutionary dynamics of ant diversification. *Evolution* 63:3023–3030.
- Purvis, A. 1995. A composite estimate of primate phylogeny. *Phil. Trans. R. Soc. Lond. B* 348:405–421.
- Purvis, A., S. Nee, and P. H. Harvey. 1995. Macroevolutionary inferences from primate phylogeny. *Proc. R. Soc. Lond. B* 260:329–333.
- Pybus, O. G. and P. H. Harvey. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. *Proc. R. Soc. Lond. B* 267:2267–2272.
- Quental, T. B. and C. R. Marshall. 2010. Diversity dynamics: molecular phylogenies need the fossil record. *Trends Ecol. Evol.* 25:434–441.
- Rabosky, D. L. 2006. Likelihood methods for detecting temporal shifts in diversification rates. *Evolution* 60:1152–1164.
- Rabosky, D. L. 2009a. Ecological limits on clade diversification in higher taxa. *Am. Nat.* 173:662–674.
- Rabosky, D. L. 2009b. Heritability of extinction rates links diversification patterns in molecular phylogenies and fossils. *Syst. Biol.* 58:629–640.
- Rabosky, D. L. 2010. Extinction rates should not be estimated from molecular phylogenies. *Evolution* 64:1816–1824.
- Rabosky, D. L. and I. J. Lovette. 2008. Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? *Evolution* 62:1866–1875.
- Ree, R. H. 2005. Detecting the historical signature of key innovations using stochastic models of character evolution and cladogenesis. *Evolution* 59:257–265.
- Ricklefs, R. E. 2007. Estimating diversification rates from phylogenetic information. *Trends Ecol. Evol.* 22:601–610.

Schluter, D. 2000. *The Ecology of Adaptive Radiation*. Oxford University Press, Oxford.

Stadler, T. 2009. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. *J. Theor. Biol.* 261:58–66.

Wakeley, J. 2009. *Coalescent theory: an introduction*. Roberts & Company Publishers, Greenwood Village, CO.

Appendix 1

If λ and μ are constant through time we have:

$$\rho(t, T) = (\mu - \lambda)(T - t),$$

and

$$\begin{aligned} I(t) &= \int_t^T e^{(\mu-\lambda)(u-t)} \mu du \\ &= \mu \int_t^T e^{(\mu-\lambda)(u-t)} du \\ &= \mu \left[\frac{e^{(\mu-\lambda)(T-t)}}{\mu-\lambda} - \frac{1}{\mu-\lambda} \right] \\ &= \mu \frac{e^{(\mu-\lambda)(T-t)} - 1}{\mu-\lambda}. \end{aligned}$$

From this we can derive the probability in eq. 1 when both rates are constant:

$$\begin{aligned} \Pr(t, n_T \geq 1) &= \frac{1}{1 + \mu \frac{e^{(\mu-\lambda)(T-t)} - 1}{\mu-\lambda}} \\ &= \frac{\mu - \lambda}{\mu - \lambda + \mu [e^{(\mu-\lambda)(T-t)}]} \\ &= \frac{\mu - \lambda}{\mu e^{(\mu-\lambda)(T-t)} - \lambda}. \end{aligned}$$

Appendix 2

Theorem. *The integral $I(t)$ has no analytical solution for any non-constant function $\mu(t)$ independent of $\lambda(t)$.*

Proof. To simplify the notation, functions of u are denoted with their names (i.e., $f = f(u)$) and all derivatives are with respect to u (i.e., $\rho' = \partial\rho(t, u)/\partial u$). Suppose that $I(t)$ has an analytical solution, therefore $e^{\rho} \cdot \mu$ has an antiderivative of the form $f \cdot e^{\rho}$ because, from the derivative of the product of two functions, $(f \cdot e^{\rho})' = (f' + f \cdot \rho')e^{\rho}$. So we should be able to find a function f such that $f' + f \cdot \rho' = \mu$. Since $\rho' = \mu - \lambda$, we have:

$$f' + f \cdot (\mu - \lambda) = \mu,$$

which can be rearranged as:

$$f' + (f + 1) \cdot \mu = f \cdot \lambda.$$

Except in the case where λ and μ are constant, there will be a relationship between $\lambda(u)$ and $\mu(u)$ implied by the existence of f . For instance, suppose f is a constant, so $f' = 0$ and the solution is:

$$f = \frac{\mu}{\mu - \lambda} = \frac{1}{1 - \lambda/\mu}$$

This will be true only if the ratio $\lambda(u)/\mu(u)$ is constant over u . □

This proof is only partial, but it is highly suggestive that $I(t)$ cannot be solved analytically. Furthermore, attempts to find this integral with algebra computer programs (e.g., www.quickmath.com) failed even with some simple models such as the logistic model.

An important corollary of the above theorem is that it is impossible to prove that ML

fitting of time-dependent birth–death models can generally be done.

Table 1. Results for the Catarrhinae phylogeny. np: number of parameters; SS: sums of squares.

The *F*-test compares the model with the constant-parameter model (1).

Model	np	SS	<i>F</i>	df	<i>P</i>
1 λ and μ constant	2	48.867	–		
2 λ follows a logistic model, μ constant	3	46.079	1.06	1, 93	0.31
3 λ follows a step model, μ constant	4	40.808	0.60	2, 92	0.55
4 λ constant, μ follows a step model	4	42.676	0.57	2, 92	0.57
5 λ follows a logistic model, μ follows a step model	5	42.640	0.38	3, 91	0.77
6 λ follows a step model with 3 rates, μ constant	6	36.711	0.33	4, 90	0.85

Table 2. Parameter estimates for Catarrhinae. τ is the time (measured from present in Myr) when the parameter changes. The 95% confidence intervals of model 1 based on 500 bootstrap samples are [0.324, 0.447] and [0, 0].

Model	Speciation	Extinction
1	$\hat{\lambda} = 0.357$	$\hat{\mu} = 0$
2	$\hat{\beta}_{\lambda} = 0.022, \hat{\alpha}_{\lambda} = -0.86$	$\hat{\mu} = 0$
3	$\hat{\lambda}_1 = 0.283, \hat{\lambda}_2 = 0.363, \hat{\tau}_{\lambda} = 5.77$	$\hat{\mu} = 0$
4	$\hat{\lambda} = 0.412$	$\hat{\mu}_1 = 0.109, \hat{\mu}_2 = 0, \hat{\tau}_{\mu} = 3.02$
5	$\hat{\beta}_{\lambda} = -0.014, \hat{\alpha}_{\lambda} = 0.23$	$\hat{\mu}_1 = 0.23, \hat{\mu}_2 = 0, \hat{\tau}_{\mu} = 4.02$
6	$\hat{\lambda}_1 = 0.268, \hat{\lambda}_2 = 0.4, \hat{\lambda}_3 = 0.358,$ $\hat{\tau}_{\lambda} = 6.41, \hat{\tau}'_{\lambda} = 4.34$	$\hat{\mu} = 0$

Table 3. Results for the Agamidae phylogeny. np: number of parameters; SS: sums of squares.

The F -test compares the model with the constant-parameter model (1).

Model	np	SS	F	df	P
1 λ and μ constant	2	993.034	–		
2 λ follows a logistic model, μ constant	3	6.911	143.69	1, 67	0.0001
3 λ follows a step model, μ constant	4	86.757	5.72	2, 66	0.005
4 λ constant, μ follows a logistic model	3	993.000	1.00	1, 67	0.37
5 λ constant, μ follows a step model	4	993.000	0.50	2, 66	0.61

Table 4. Parameter estimates for Agamidae. τ is the time (measured from present in Myr) when the parameter changes. The 95% confidence intervals of model 2 based on 1000 bootstrap samples are $[-0.210, 0.016]$, $[-2.169, 0.362]$, and $[0, 0.012]$.

Model	Speciation	Extinction
1	$\hat{\lambda} = 0.12$	$\hat{\mu} = 0$
2	$\hat{\beta}_{\lambda} = -0.093, \hat{\alpha}_{\lambda} = -0.646$	$\hat{\mu} = 0$
3	$\hat{\lambda}_1 = 0.461, \hat{\lambda}_2 = 0.06, \hat{\tau}_{\lambda} = 27.04$	$\hat{\mu} = 0$
4	$\hat{\lambda} = 0.12$	$\hat{\beta}_{\mu} = 0.19, \hat{\alpha}_{\mu} = -3.25$
5	$\hat{\lambda} = 0.12$	$\hat{\mu}_1 = 0, \hat{\mu}_2 = 0, \hat{\tau}_{\mu} = 27.43$

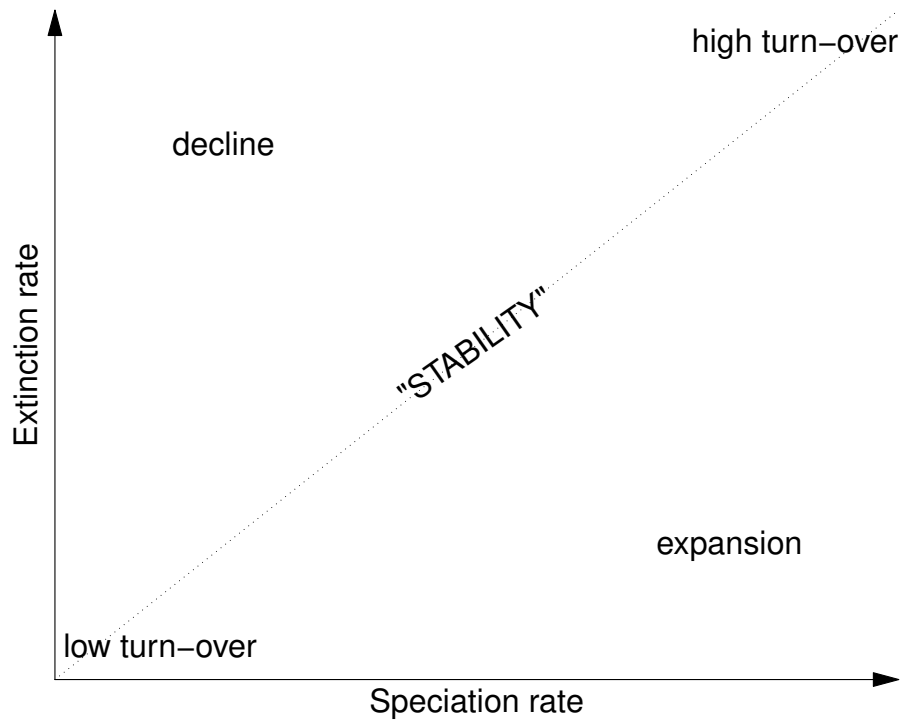


Figure 1. Relationships between differences in speciation and extinction rates and phylogenetic diversification patterns.

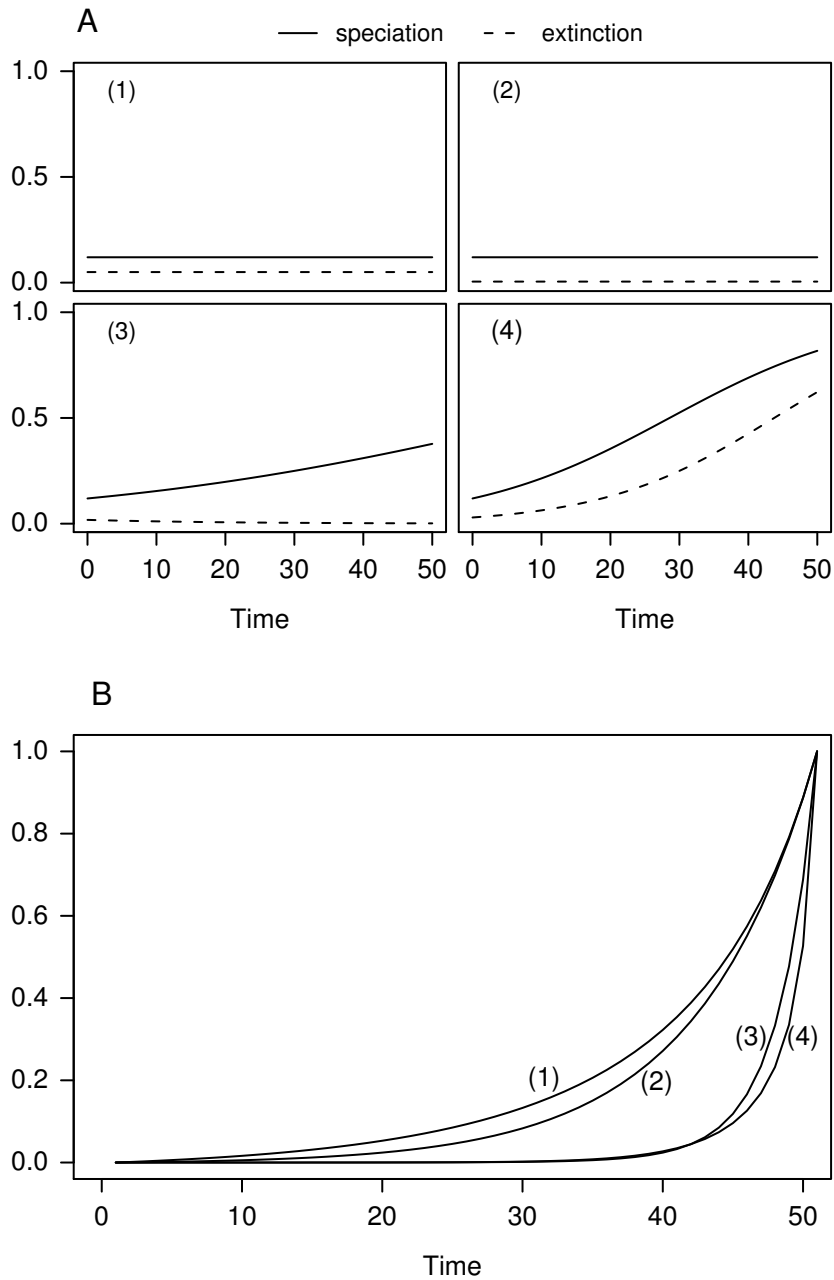


Figure 2. (A) Four different diversification scenarios with constant (1 and 2) or time-dependent (3 and 4) speciation and extinction probabilities, and (B) their respective predicted cumulative distribution of branching times calculated with eq. 3.

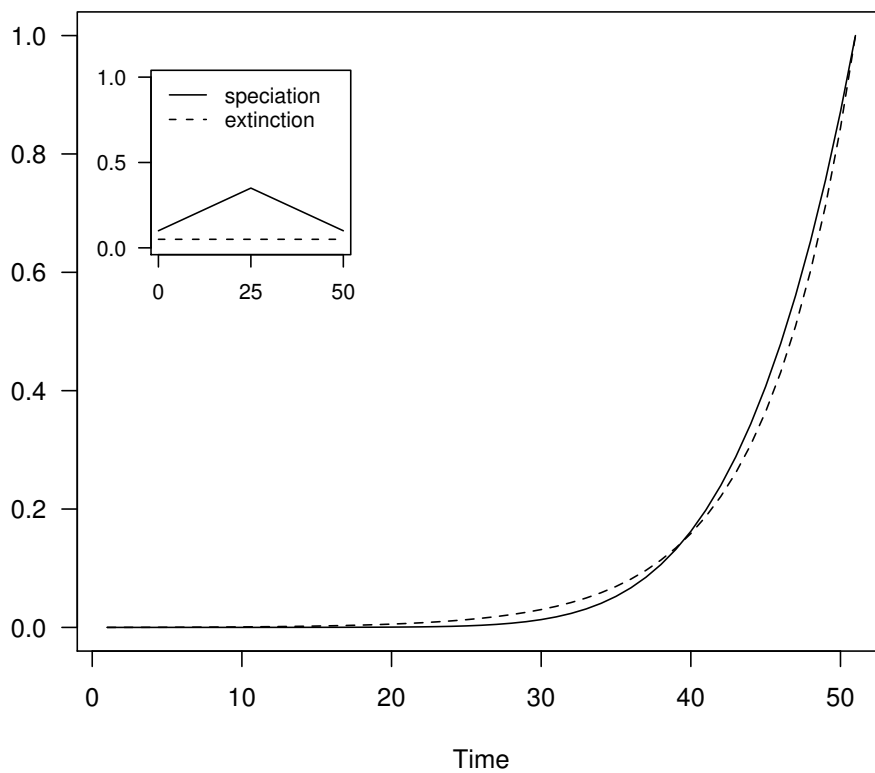


Figure 3. Two different diversification scenarios that would result in very similar cumulative distributions of branching times. The straight curve is the CDF under the scenario given in the inset (time-dependent speciation and constant extinction); the dashed line is the CDF under a scenario of constant diversification with $\lambda = 0.17$ and $\mu = 0.005$.

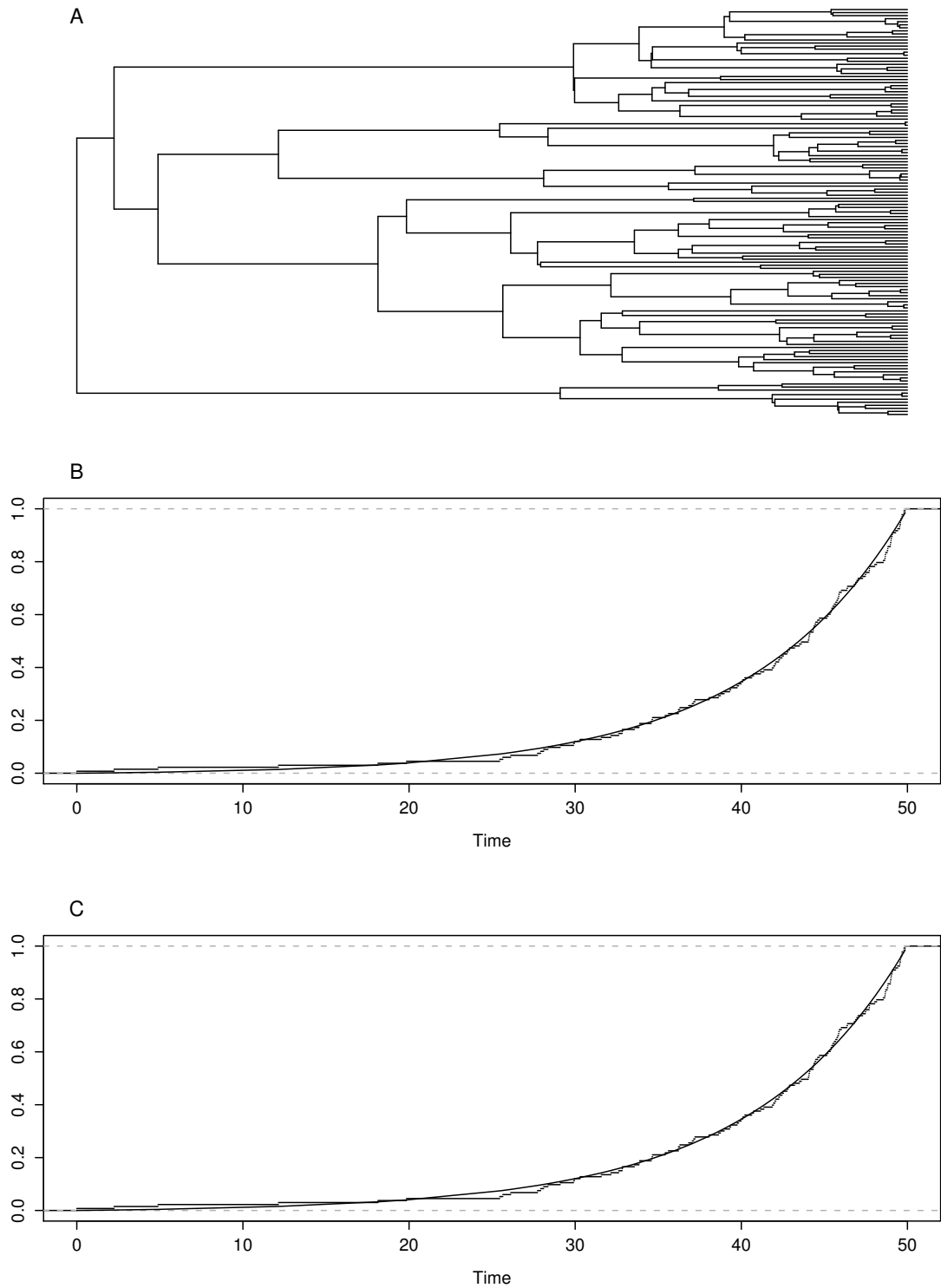


Figure 4. (A) A tree simulated with constant parameters $\lambda = 0.13$ and $\mu = 0.05$ ($N = 134$). (B) The empirical cumulative distribution function of branching times (step curve) together with fitted curve assuming constant speciation and extinction probabilities. (C) Same as in (B) but the fitted curve assumes time-dependent probabilities following a logistic model. The F -test comparing both models showed no significant difference leading to acceptance of the constant-parameter model.

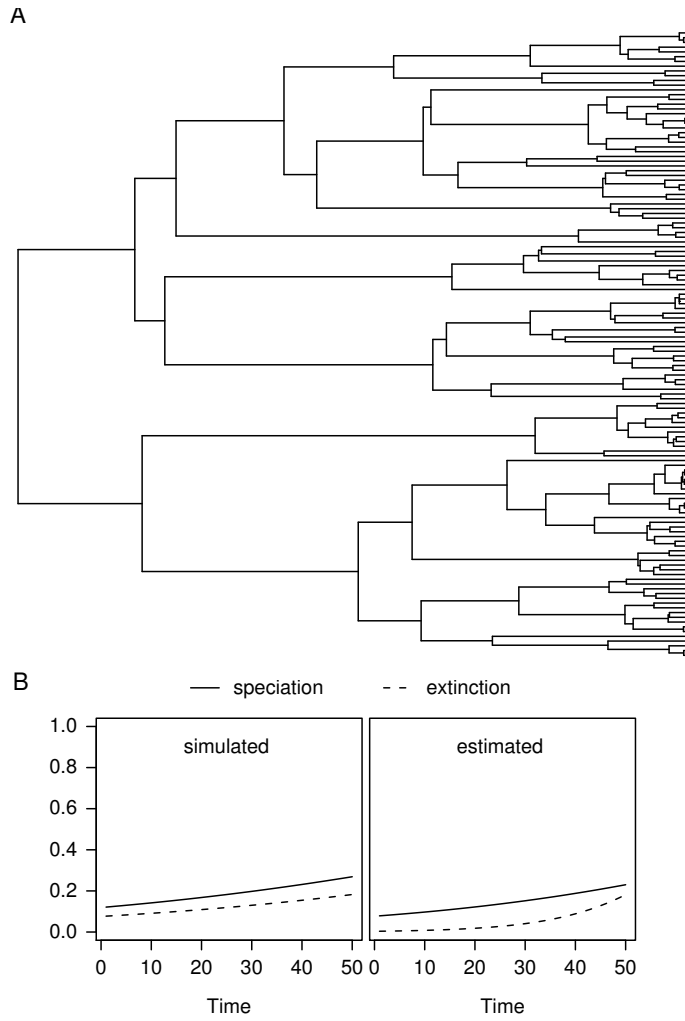


Figure 5. (A) A tree simulated with increasing speciation and extinction probabilities following logistic models ($N = 132$). (B) The simulated variation of speciation and extinction probabilities with parameters $\beta_\lambda = 0.02$, $\alpha_\lambda = -2$, $\beta_\mu = 0.02$, $\alpha_\mu = -2.5$ (left), and the inferred one from the estimated parameters of the logistic model (right). The estimates are [95% confidence interval]: $\hat{\beta}_\lambda = 0.026$ $[-0.014, 0.044]$, $\hat{\alpha}_\lambda = -2.485$ $[-3.353, -1.842]$, $\hat{\beta}_\mu = 0.088$ $[0.020, 2.224]$, $\hat{\alpha}_\mu = -5.465$ $[-109.980, -5.528]$. The confidence intervals were computed with 100 nonparametric bootstrap samples.

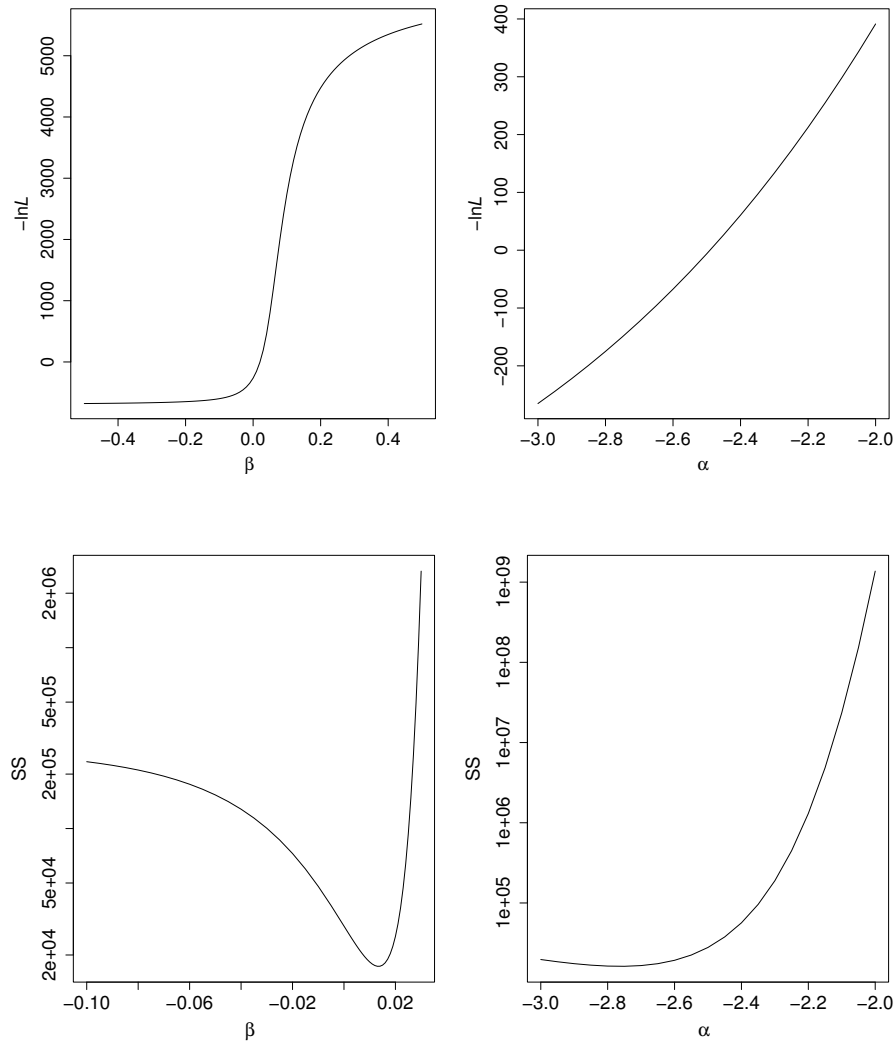


Figure 6. Variation in minus log-likelihood ($-\ln L$) and sums of squares (SS) with respect to the extinction parameters for the simulated phylogeny shown on Fig. 5A. The log-likelihood function shows no local or global minimum resulting in a failure of maximum likelihood estimation with these data.