Random phylogenies and the distribution of branching times

Emmanuel Paradis

Random phylogenies and the distribution of branching times

Emmanuel Paradis

Institut des Sciences de l'Évolution, Université Montpellier – CNRS – IRD – EPHE, CC 064, Place Eugène Bataillon, F-34095 Montpellier cédex 05, France

Abstract

Phylogenetic trees reconstructed without fossils have become an important source of information to study evolutionary processes. A widely used method to describe patterns of phylogenetic diversification is known as the lineages-through-time (LTT) plot. Recently, it has been shown that it is possible to predict the distribution of the branching times of a phylogeny, thus making possible to derive a theoretical LTT curve conditioned on diversification parameters. Here, I review some aspects related to this prediction showing how to derive it for any time-dependent model of diversification, as well as calculating a prediction interval around a theoretical LTT curve. The accuracy of the prediction interval was assessed with simulations using fixed or random tree sizes under constant-rate models as well as two models of time-dependent diversification. The prediction intervals were relatively narrower and more accurate for larger trees. The features of this approach are discussed as well as its potential applications.

Keywords: diversity, extinction, LTT plot, speciation

Tel: +33 (0) 4 67 14 46 85
Fax: +33 (0) 4 67 14 36 14
E-mail: Emmanuel.Paradis@ird.fr

1. Introduction

Biological diversity on Earth is the result of several physical, chemical, and biological forces which affect the rates of species origination and extinction. The investigation of evolutionary diversification with phylogenetic trees has received a lot of attention during the last two decades (see recent reviews in Pennell and Harmon, 2013; Morlon, 2014). In particular, a wide range of methods have been developed for the statistical analysis of phylogenies in order to estimate speciation and extinction parameters or to test hypotheses about their variation through time or among different lineages. One popular approach is a graphical method known as the lineage-through time (LTT) plot (Nee et al., 1992). This method requires a dated phylogenetic tree (also known as chronogram) from which the dates of divergence are computed. A plot of the increase in the number of lineages (on the y-axis) with respect to time (on the x-axis) is then done to realise the LTT plot. The y-axis can be log-transformed if, for instance, the increase in the number of lineages is very strong close to present, so that the structure of the curve back in the past may be more visible. The initial motivation behind this method was to find a relationship between the shape of the curve displayed by the LTT plot and the temporal variation (or lack of) in the speciation and extinction rates of the lineage under study. Two approaches to this end can be found in the literature. The first one is based on deriving explicit equations of the distribution of the branching times of a tree with respect to the diversification parameters (e.g., Stadler, 2008; Paradis, 2011; Hallinan, 2012). The second one, which has enjoyed some success in the recent literature, is to simulate many trees under a given model of diversification and then derive a null distribution of the LTT plot (e.g., Crisp and Cook, 2009; Pie and Tschá, 2009; Fordyce, 2010; Lorén et al., 2014).

Recently, Lambert and Stadler (2013) investigated in details general macroevolutionary models of speciation and extinction. An important result from this study is that under a homogeneous model of diversification, the distribution of branching times depends only on the rates of speciation and extinction. Therefore, the LTT plot of a phylogenetic tree completely summarises the information required to estimate these rates under the assumption of homogeneity (i.e., speciation and extinction rates are the same for all species at a given time

though these parameters may vary through time) and the assumption that both rates are not correlated. On the other hand, remarkable progress has been accomplished on the simulation of phylogenetic trees under various diversification models with specified or random numbers of species (Paradis, 2011; Stadler, 2011; Höhna, 2013). An issue that remains open concerns the links between these different methods and approaches, particularly whether explicit equations and simulations of phylogenetic trees lead to similar prediction of branching times distribution. Another question is about the effect of simulating a phylogeny with a specified number of species on the distribution of branching times. Finally, the nature of the variability around the predicted distribution of branching times has not yet been assessed.

In the present paper, I review the question of predicting the distribution of branching times of a phylogeny given a model of temporal variation in speciation and extinction rates. Further elaborating on the work presented in Paradis (2011), I show how the calculations can be improved and applied to any model of time-dependent diversification. I then discuss how to derive a prediction interval around the theoretical LTT curve. I also review the methods for simulating phylogenetic trees from speciation–extinction models, and discuss the relationships among the different approaches. Using stochastic simulations, it is shown that the prediction interval of the LTT plot can be accurate in different settings of diversification parameters and tree sizes.

2. Methods

2.1. Speciation–extinction models and the distribution of branching times

Throughout this paper, we consider a dated (i.e., ultrametric) phylogenetic tree with N species. We assume that this tree is rooted and fully dichotomous; consequently, it has $N-1$ ($=m$) internal nodes. The branching times, measured from the root, are denoted $t_1,\ldots,t_m$. By convention in this paper, the root is the origin of the time scale, so that $t_1 = 0$, and the present is denoted T.

Birth–death models have been widely used to analyse or simulate the diversification of clades. The basic assumption of these models is that each species is continuously exposed to speciation and extinction. The rates at which these events happen may vary through time. In such

a time-dependent birth–death model, the probability that a species undergo speciation follows a function denoted $\lambda(t)$, and similarly for the probability of extinction $\mu(t)$. This formulation implies no lineage-specific variation: at a given time, all species have the same probabilities of speciation and extinction (i.e., the assumption of homogeneity). In order to derive the predicted LTT curve under a time-dependent birth–death model, we use the probability that a lineage, originating from a single species at time t, is not extinct at time T (Kendall, 1948):

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

where n_T is the number of living species in the lineage at time T, and with the function ρ defined as:

$$\rho(t,T) = \int_t^T \mu(u) - \lambda(u)\mathrm{d}u,$$

and:

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

We note that in eq. 1, t and T are given fixed values; only n_T is a random variable.

We can now derive the distribution of branching times from the phylogeny. The probability of observing a branching event in a phylogeny, denoted as $\pi(t)$, 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 probability is given by the product of the speciation probability at t, $\lambda(t)$, with the number of species living at this time. Obviously, we do not know the latter but we can substitute it by its expectation: $\mathbb{E}(n_t) = \mathrm{e}^{-\rho(0,t)}$. Consequently, we have:

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

This density gives the relative probabilities of a branching time on the interval $0 < t < T$, but this is not a probability density because this function does not integrate (since t is continuous) to one.

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

$$\mathcal{F}(t) = \frac{\int_0^t \pi(u)\mathrm{d}u}{\int_0^T \pi(u)\mathrm{d}u}. \tag{3}$$

This equation is used to derive the predicted curve of the LTT plot with given functions $\lambda(t)$ and $\mu(t)$ (Paradis, 2011). $\mathcal{F}(t)$ is an increasing function with increasing t, and we have $\mathcal{F}(0) = 0$ and $\mathcal{F}(T) = 1$. Höhna (2013) derived an equation which is close but different from eq. 2 above. However, it can be shown that both forms are equivalent (see Appendix A).

As one can see from the above formulae, the calculations involved in computing $\mathcal{F}(t)$ are quite heavy with several integrations which, in general, cannot be solved analytically (a trivial exception is the case with λ and μ constant). In practice, standard numerical methods for integration can be used but they are usually slow. Hallinan (2012) found an interesting solution to this problem: if one assumes that $\lambda(t)$ and $\mu(t)$ are constant across some time intervals (i.e., they vary following some step functions), then eq. 1 can be calculated easily. This is an interesting solution because the assumption that speciation and extinction rates are constant over some short time periods (e.g., a few thousand years) may be reasonable on a macroevolutionary scale (several million years).

Here I develop some improvements on the computation of the above equations that make the calculation of $\mathcal{F}(t)$ faster for any function $\lambda(t)$ and $\mu(t)$. First, we note that eq. 1 simplifies to:

$$\Pr(t, n_T \geq 1) = \left[1 + \int_t^T \mathrm{e}^{\rho(t,u)}\mu(u)\mathrm{d}u\right]^{-1}.$$

We then log-transform eq. 2:

$$\ln\pi(t) = -2\ln\left[1 + \int_t^T \mathrm{e}^{\rho(t,u)}\mu(u)\mathrm{d}u\right] - \rho(0,t) + \ln\lambda(t).$$

Working on a logarithmic scale is less likely to inflate computing errors since we work here with sums instead of products. Note that if the antiderivates of $\lambda(t)$ and of $\mu(t)$ can be found, say $\Lambda(t)$

and $M(t)$, then $\rho(t,u)$ can be computed directly:

$$\rho(t,u) = M(u) - \Lambda(u) - M(t) + \Lambda(t).$$

For instance, if both follow some step functions, as assumed in Hallinan (2012), then these antiderivates are calculated simply with rectangle areas. The two remaining difficulties are the integral of $e^{\rho(t,u)}\mu(u)$ and the one of $\pi(t)$. To compute them efficiently, the following strategy has been adopted. The function to be integrated is first evaluated on a sufficiently large number of points which are defined so that two neighbouring points are close enough so the area under the curve can be approximated with a rectangle. For instance (where δ is a small value):

$$\int_x^{x+\delta} \pi(u)\mathrm{d}u \approx \pi(x+\delta/2) \times \delta.$$

This is a standard step in computing an integral with an unknown antiderivative and is called the midpoint method. The integral between 0 and t is then calculated with the sum of all the rectangle areas between these two values. We have to be careful that the error generated by the midpoint method, which may be negligible for a single small interval, will tend to accumulate when summing over many such intervals. A solution is to use Simpson's rule which is defined by:

$$\int_x^{x+\delta} \pi(u)\mathrm{d}u \approx \frac{\pi(x) + 4\pi(x+\delta/2) + \pi(x+\delta)}{6} \times \delta.$$

Over a wider interval made of many elementary intervals, the values of all left-, mid-, and right-points of each elementary interval are summed using the same coefficients as above (1, 4, and 1). In the framework of the present application, it was found that defining 500 elementary intervals was enough, so that it is needed to calculate 999 values of $\pi(t)$. Both integrals in eq. 3 are then calculated with Simpson's rule (the denominator being calculated over all elementary intervals).

2.2. The lineage-through time plot and its prediction interval

Simulations show clearly that the distribution of branching times is highly variable even with the same parameter values. So it is desirable to predict the variability around the expected LTT plot in the same way than prediction bands are derived in statistical practice (Liu et al., 2008). We use here an approach introduced by Hallinan (2012) whose rationale is as follows: consider a point in time, say τ, and let us write $p = \mathcal{F}(\tau)$. Consider now a branching time t taken at random in a tree generated with the same parameters used to compute $\mathcal{F}$. Then, by the definition of a CDF, the event $t \leq \tau$ follows a Bernoulli distribution with probability p. We can thus derive upper and lower bounds for the CDF of branching times by calculating the quantiles of the number of observed branching times for each value of t. Suppose we observe n branching times, then the number of those smaller than or equal to τ follows a binomial distribution with parameters n and p. It is therefore possible to calculate a 95% prediction interval of this distribution using the quantiles of the binomial distribution at the probabilities 0.025 and 0.975. In practice, these bounds are calculated at a number of points in the interval $[0, T]$ (every 0.1 time unit in this paper).

2.3. Simulating speciation–extinction trees

Simulating evolutionary data has become an important endeavour in data analysis with the constant development of Monte Carlo or Bayesian approaches. There is a long history of the simulation of trees with different purposes (see Paradis, 2014, for a brief overview). The simulation of speciation–extinction (or birth–death) trees is only a subset of these methods. Two broad categories of methods to simulate speciation–extinction trees can be defined: the homogeneous methods which are continuous in time, and the heterogeneous methods which can model any kind of variation in speciation and extinction rates (in relation to species biological traits, environment, climate, ...) I consider only the former category here, and distinguish three approaches to this:

1. Starting from a single species, evolution is simulated forward in time. The times of speciation and extinction events are simulated with time-to-event (or waiting time)

formulae (Paradis, 2011).

2. Starting from present and a given number of species, evolution is simulated backward in time. The times of speciation and extinction events are simulated with a coalescent approach (Stadler, 2011).
3. Branching times are sampled randomly from the expected distribution described in the previous section. The topology of the tree is built by random successive agglomeration of species.

There can be many variants around these broad lines. For instance, the simulation can start from a (root) node instead of a single species. Another variation is to record or not the extinction events: if one is only interested in ultrametric trees (i.e., without fossils) then the extinct lineages may be simply discarded to speed-up calculations (Paradis, 2011).

These three approaches share in common that they simulate branch lengths in continuous time making comparisons easier because the outputs do not depend on the definition of arbitrary time steps. The first approach makes possible to simulate a tree with a fixed value of T (N is always random), while the second one can simulate trees with fixed N (T is random), and the third one can do simulations with both T and N fixed. The first approach has been used in earlier works to simulate trees with N fixed; however, Hartmann et al. (2010) demonstrated that this approach results in biased simulated trees and presented a method to remedy this problem.

Various implementations of these approaches are available in different computer programs. The first approach is implemented in ape (Paradis et al., 2004) which provides two algorithms (depending on whether extinct lineages are recorded or not) which simulate trees with any time-dependent variation in λ and μ (these parameters may be constant). The second approach is implemented in TreeSim (Stadler, 2014) which can simulate trees with constant λ and μ and the possibility to include random sampling of species as well as episodic (mass) extinctions. In order to complete this, I implemented in ape the possibility to simulate a tree with N fixed and any kind of time-dependent variation in both λ and μ (which still can be constant). This uses a slightly modified version of the algorithm given in Stadler (2011). The difference is in the way waiting times are generated (step 2 of the original algorithm), specifically, the formula used to generate a

random time x starting from time t is here:

$$n_t\left[\lambda(t)+\mu(t)\right]\exp\left[-n_t\int_x^t \lambda(u)+\mu(u)\mathrm{d}u\right].$$

We keep in mind that in the framework used in this paper the time scale goes from past to present, so $x<t$ since the algorithm goes backward in time. If λ and μ are constant, the above formula simplifies to $n_t(\lambda+\mu)\mathrm{e}^{n_t(\lambda+\mu)(t-x)}$ which is the probability density function of an exponential distribution with rate $n_t(\lambda+\mu)$, thus making possible to use efficient random number generators from common statistical software.

Finally, the third approach is implemented in TESS which can simulate trees with any time-dependent model (Höhna, 2013).

2.4. Simulation study

A simulation study was conducted in order to assess the effect of tree size and diversification parameters on the predictions of the distribution of branching times. Two of the above simulation algorithms were used to generate phylogenetic trees. The first one is described in Paradis (2011) and simulates a tree in continuous time where λ and μ are allowed to vary with time under any user-specified model. This is a time-forward algorithm where T is fixed and N is variable (first approach of the previous section). The second algorithm is due to Stadler (2011) and simulates trees in continuous time using a time-backward method with N fixed (second approach of the previous section). Both algorithms require as input the values of the parameters λ and μ and are implemented in the R (R Core Team, 2014) packages ape and TreeSim (Stadler, 2014), respectively. The third approach described in the previous section was not considered here because it samples branching times directly from the expected LTT curve, so a good agreement between the simulated trees and the predictions should be no surprise.

The predicted LTT plot and its 95% prediction interval were computed for a given set of parameter values. Ten thousand trees were simulated for each set of parameter values with both algorithms. For each tree, the LTT plot was calculated, and the agreement with the predictions were assessed in two ways. First, the proportion of points falling within the 95% prediction

interval was counted. Second, the proportion of points falling above or below the predicted curve was also counted. Two sets of simulations were done. In the first set, λ and μ were kept constant with values given in Table 1. In the second set, two models of temporal variation were considered: the first one is an early burst model of speciation with decreasing λ through time and constant μ (Fig. 2a); the second one is a model of increasing extinction in recent time and constant λ (Fig. 3a).

3. Results

The proportion of points of the LTT plots falling within the prediction intervals depended mainly on the size of the tree (Table 1). For trees simulated with $N = 10$, about 69% of the points fell within the prediction intervals for all values of λ and μ. This rose to about 90% for $N = 50$, and 94% for $N = 200$. For $N = 1000$, this percentage was slightly larger when $\lambda = 0.1$ (94%) than when $\lambda = 0.2$ (93%). For trees generated with a random value of N, the percentage of points outside the prediction interval was related to the value of the difference $\lambda - \mu$: when this difference was large, the percentage was above 93% whereas it was below 70% for the smallest values.

A similar pattern emerged from the proportion of points above the predicted curve (Table 2). Except for trees simulated with $N \leq 20$, the points were as likely to be above than below the predicted LTT curve. The same result was observed with random N. Figure 1 shows the predicted LTT plots under two sets of parameters: $\lambda = 0.1, \mu = 0$, and $\lambda = 0.3, \mu = 0.2$. The expected number of species is the same in both cases since the difference $\lambda - \mu$ is the same. As expected, the 95% prediction intervals depended on the size of the tree (shown for $N = 20$ and $N = 200$).

Figure 2a shows the modelled rates λ and μ under the scenario of early burst together with the predicted LTT curve. Figure 2b shows the LTT plot from a tree simulated with this model. By contrast to the cases with constant speciation and extinction rates, the proportion of points of the LTT plots falling within the prediction intervals varied little with respect to N and was around 90% (Table 3). The proportion of points below the predicted LTT curve was around 35% and did not vary much with tree size except when N was random where it was above 50% (Table 4).

Figure 3a shows the modelled rates λ and μ under the scenario of increasing extinction together with the predicted LTT curve. Figure 3b shows the LTT plot from a tree simulated with this model. With this scenario, the proportion of points of the LTT plots within the prediction intervals decreased slightly with increasing tree size ranging from around 92% to around 82% (Table 3). Similarly to the previous scenario, most of these points were below the predicted curve except when N was random (Table 4).

4. Discussion

The LTT plot is one of the most widely used methods to assess diversification from dated phylogenies. Its simplicity to run and its ability to display temporal phylogenetic patterns make it an attractive tool for evolutionists studying many groups. This method has some limitations, particularly some are intrinsic to the analysis of reconstructed phylogenies which ignore fossils (see review in Morlon, 2014). In spite of these limitations, which are now well-known, the LTT plot has proven to be a valuable tool for investigating evolutionary diversification from molecular phylogenies (e.g., Moreau et al., 2006; Bininda-Emonds et al., 2007; Lorén et al., 2014).

This paper aims to review some recent developments on the analysis of branching times from phylogenetic trees with a focus on predicting their distribution. It is possible to compute a predicted LTT curve as well as a prediction interval around the predicted curve. Some improvements to compute this predicted LTT curve are presented above. The prediction interval is computed with the assumption that the branching times of the phylogeny follow a binomial distribution at any given time.

One issue for computing a predicted LTT curve is the justification of the parameter values. In the examples presented in this paper, these values were taken as fixed and given a priori, because the data were simulated so the real diversification parameter values of the phylogenies were known. With real data, the situation is different and the parameters must be estimated from the data. Two applications are possible in this context. First, it is possible to calculate an "empirical" confidence interval around the observed LTT plot with the same approach as described above: the values of p would thus be calculated from the observed cumulated proportions of number of

lineages. Such a confidence interval is actually "model free" (i.e., they do not require to estimate the diversification parameters). The second application is to infer a predicted LTT curve from a fitted model and draw it together with the observed LTT plot. For instance, suppose a constant-parameter birth–death model is fitted to a tree, then the estimated values of λ and μ can be used to draw a predicted LTT curve and compare it with the actual distribution of branching times. Such an approach can help to assess the goodness of the fit of a particular model of diversification (see also Wollenberg et al., 1996; Paradis, 1998).

The match between the prediction intervals and the simulated trees showed various degrees of agreement. For the constant-parameter models, the degree of success of the prediction intervals was mainly related to the tree size (N). With $N = 1000$, almost 95% of the points from the LTT plots of the simulated trees were within the 95% prediction intervals. When trees were simulated with random N, the precision of the predictions were positively related to the net diversification rate ($\lambda - \mu$), thus confirming the importance of tree size since trees simulated with a higher diversification rate are expected to have larger sizes (since the time of evolution of the simulated trees, T, was kept constant). An interesting result was that, for almost all simulations with constant parameters, the points of the LTT plots have the same probability to fall above or below the predicted LTT curve, with only a slightly higher proportion (*ca.* 0.55) of points below this curve for the smaller trees (see Table 2).

The success of the predictions was more contrasted when trees were simulated with the time-dependent models of diversification. Overall, the proportion of points of the LTT plots within the prediction inverval was similar to the contant-rate case with similar tree sizes. However, this proportion did not change much with tree size and did not reach 95% like for the contant-rate case. On the other hand, the points tend to be on one side of the predicted LTT curve and slightly preferentially above it.

Paradis (2011) and Etienne and Rosindell (2012) proposed to fit models of diversification and estimate the parameters by minimising the deviation between the observed LTT curve and the one predicted from the model using a least squares criterion. It has been shown that this fitting procedure performs better than fitting the same model by maximum likelihood at least in some

situations (Paradis, 2011). Hallinan (2012) questioned the use of LTT plots as a tool to assess the fit of a diversification model to a phylogenetic tree. The difficulty appears to be the autocorrelation of the individual points of an LTT plot: successive points tend to be above or below the predicted line as shown by the examples in this paper. However, the simulation results reported here suggest that, when speciation and extinction rates are constant through time, the points do not statistically tend to be either above or below the predicted line over the complete range of the plot, even within a single LTT plot. This seems to validate the use of a least squares approach to model fitting. On the other hand, when these rates varied through time, the points tended to be above the predicted LTT curve which may result in biased parameter estimation. This will require further study.

To conlude, the distribution of branching times can be well predicted by the above formula (eq. 3) when the speciation and extinction rates are constant through time. The predictions are still accurate but less consistent for the time-dependent models considered here, which is a point requiring further study. A prediction interval around the predicted distribution (i.e., the LTT curve) can be derived and is accurate except for the smallest tree sizes. In spite of being more than two-decade old, the LTT plot is a flexible and very useful graphical tool to display the distribution of branching times, either from phylogenetic trees, or from predictions derived from speciation–extinction models.

Acknowledgements

I am grateful to Tanja Stadler for clarification about some issues on tree simulation, and to two anonymous reviewers for their comments. This is publication ISEM 201x-xxx.

Appendix A

Höhna (2013, eq. 8) derived the probability of speciation events in a phylogenetic tree as (using the notation of the present paper):

$$\pi(t) \propto \Pr(t, n_T = 1)\lambda(t).$$

The probability that a lineage originating at time t has a single species at time T used by Höhna is (Kendall, 1948, eq. 16):

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

We now develop eq. 2 of the present paper:

$$\begin{aligned}\pi(t) &= \Pr(t, n_T \geq 1)^2 \mathbb{E}(n_t) \lambda(t) \\ &= \left[\frac{e^{-\rho(t,T)}}{W(t)}\right]^2 e^{-\rho(0,t)} \lambda(t) \\ &= \frac{e^{-\rho(t,T)}}{W(t)^2} e^{-\rho(t,T)} e^{-\rho(0,t)} \lambda(t).\end{aligned}$$

We note that:

$$e^{-\rho(t,T)} e^{-\rho(0,t)} = e^{-\rho(0,T)}.$$

We also note that the fraction is equal to $\Pr(t, n_T = 1)$, and reordering the terms in the product this gives:

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

which is identical to Höhna's equation and shows what is the proportionality factor. We note that the latter does not depend on t, so its value will cancel out in the ratio of integrals in eq. 3 (also in Höhna, 2013, eq. 9).

References

Bininda-Emonds, O. R. P., Cardillo, M., Jones, K. E., MacPhee, R. D. E., Beck, R. M. D., Grenyer, R., Price, S. A., Vos, R. A., Gittleman, J. L., Purvis, A., 2007. The delayed rise of present-day mammals. Nature 446, 507–512.

Crisp, M. D., Cook, L. G., 2009. Explosive radiation or cryptic mass extinction? Interpreting signatures in molecular phylogenies. Evolution 63 (9), 2257–2265.

Etienne, R. S., Rosindell, J., 2012. Prolonging the past counteracts the pull of the present: protracted speciation can explain observed slowdowns in diversification. Syst. Biol. 61 (2), 204–213.

Fordyce, J. A., 2010. Interpreting the γ statistic in phylogenetic diversification rate studies: a rate decrease does not necessarily indicate an early burst. PLoS ONE 5 (7), e11781.

Hallinan, N., 2012. The generalized time variable reconstructed birth–death process. J. Theor. Biol. 300, 265–276.

Hartmann, K., Wong, D., Stadler, T., 2010. Sampling trees from evolutionary models. Syst. Biol. 59 (4), 465–476.

Höhna, S., 2013. Fast simulation of reconstructed phylogenies under global time-dependent birth–death processes. Bioinformatics 29 (11), 1367–1374.

Kendall, D. G., 1948. On the generalized "birth-and-death" process. Ann. Math. Stat. 19, 1–15.

Lambert, A., Stadler, T., 2013. Birth–death models and coalescent point processes: the shape and probability of reconstructed phylogenies. Theor. Pop. Biol. 90, 113–128.

Liu, W., Lin, S., Piegorsch, W. W., 2008. Construction of exact simultaneous confidence bands for a simple linear regression model. Int. Statist. Rev. 76 (1), 39–57.

Lorén, J. G., Farfán, M., Fusté, M. C., 2014. Molecular phylogenetics and temporal diversification in the genus *Aeromonas* based on the sequences of five housekeeping genes. PLoS ONE 9 (2), e88805.

Moreau, C. S., Bell, C. D., Vila, R., Archibald, S. B., Pierce, N. E., 2006. Phylogeny of the ants: diversification in the age of angiosperms. Science 312, 101–104.

Morlon, H., 2014. Phylogenetic approaches for studying diversification. Ecol. Lett. 17 (4), 508–525.

Nee, S., Mooers, A. Ø., Harvey, P. H., 1992. Tempo and mode of evolution revealed from molecular phylogenies. Proc. Natl. Acad. Sci. USA 89 (17), 8322–8326.

Paradis, E., 1998. Testing for constant diversification rates using molecular phylogenies: a

general approach based on statistical tests for goodness of fit. Mol. Biol. Evol. 15 (4), 476–479.

Paradis, E., 2011. Time-dependent speciation and extinction from phylogenies: a least squares approach. Evolution 65 (3), 661–672.

Paradis, E., 2014. Simulation of phylogenetic data. In: Garamszegi, L. Z. (Ed.), Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology. Springer-Verlag, Berlin Heidelberg, pp. 335–350.

Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20 (2), 289–290.

Pennell, M. W., Harmon, L. J., 2013. An integrative view of phylogenetic comparative methods: connections to population genetics, community ecology, and paleobiology. Ann. N.Y. Acad. Sci. 1289, 90–105.

Pie, M., Tschá, M. K., 2009. The macroevolutionary dynamics of ant diversification. Evolution 63 (11), 3023–3030.

R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL http://www.R-project.org

Stadler, T., 2008. Lineages-through-time plots of neutral models for speciation. Math. Biosci. 216 (2), 163–171.

Stadler, T., 2011. Simulating trees with a fixed number of extant species. Syst. Biol. 60 (5), 676–684.

Stadler, T., 2014. TreeSim: Simulating trees under the birth-death model. R package version 2.1.
URL http://CRAN.R-project.org/package=TreeSim

Wollenberg, K., Arnold, J., Avise, J. C., 1996. Recognizing the forest for the trees: testing temporal patterns of cladogenesis using a null model of stochastic diversification. Mol. Biol. Evol. 13 (6), 833–849.

Fig. 1. Lineage-through time plots (grey lines) for ten simulated trees with different values of speciation (λ) and extinction (μ) rates and tree size (N). The thick line shows the predicted LTT distribution, and the dashed lines delimit the 95% prediction intervals. The y-axes are scaled to one. F: cumulative proportions of lineages.

Fig. 2. (a) The early burst model of speciation showing the parameters (λ and μ) and the predicted LTT curve. (b) The LTT plot from a phylogeny simulated with this model (black line) and the number of living lineages (grey line) from the same simulation.

Fig. 3. (a) A model of increasing extinction in recent time showing the parameters (λ and μ) and the predicted LTT curve. (b) The LTT plot from a phylogeny simulated with this model (black line) and the number of living lineages (grey line) from the same simulation.

Table 1: Proportion of points from the LTT plots falling within the 95% prediction interval. The simulations were replicated 10,000 times. All trees were simulated with $T = 50$ and $n_0 = 1$. N: number of species in the tree; λ: speciation rate; μ: extinction rate.

λ	μ	N						
		10	20	50	100	200	1000	random
0.1	0	0.692	0.817	0.900	0.930	0.942	0.943	0.935
	0.05	0.691	0.818	0.900	0.927	0.939	0.945	0.847
	0.09	0.691	0.817	0.900	0.927	0.939	0.942	0.680
0.2	0.1	0.690	0.817	0.899	0.926	0.937	0.931	0.933
	0.15	0.690	0.815	0.897	0.925	0.937	0.932	0.893
	0.19	0.691	0.818	0.899	0.925	0.938	0.932	0.787

Table 2: Proportion of points from the LTT plots falling below the predicted LTT line. See Table 1 for details.

λ	μ	N						
		10	20	50	100	200	1000	random
0.1	0	0.546	0.521	0.505	0.505	0.501	0.504	0.505
	0.05	0.548	0.525	0.504	0.512	0.509	0.501	0.513
	0.09	0.553	0.525	0.515	0.512	0.501	0.507	0.550
0.2	0.1	0.552	0.527	0.507	0.496	0.500	0.500	0.502
	0.15	0.551	0.522	0.510	0.506	0.500	0.499	0.516
	0.19	0.539	0.517	0.521	0.510	0.502	0.504	0.529

Table 3: Proportion of points from the LTT plots falling within the 95% prediction interval for two models of time-dependent diversification. The simulations were replicated 10,000 times. T was random except when N was random where $T = 50$ and $n_0 = 1$.

Model	N				
	20	50	100	200	random
Early burst	0.910	0.930	0.926	0.906	0.792
Recent extinction	0.916	0.897	0.858	0.811	0.890

Table 4: Proportion of points from the LTT plots falling below the predicted LTT line. See Table 3 for details.

Model	N				
	20	50	100	200	random
Early burst	0.353	0.358	0.343	0.310	0.716
Recent extinction	0.278	0.276	0.257	0.233	0.616

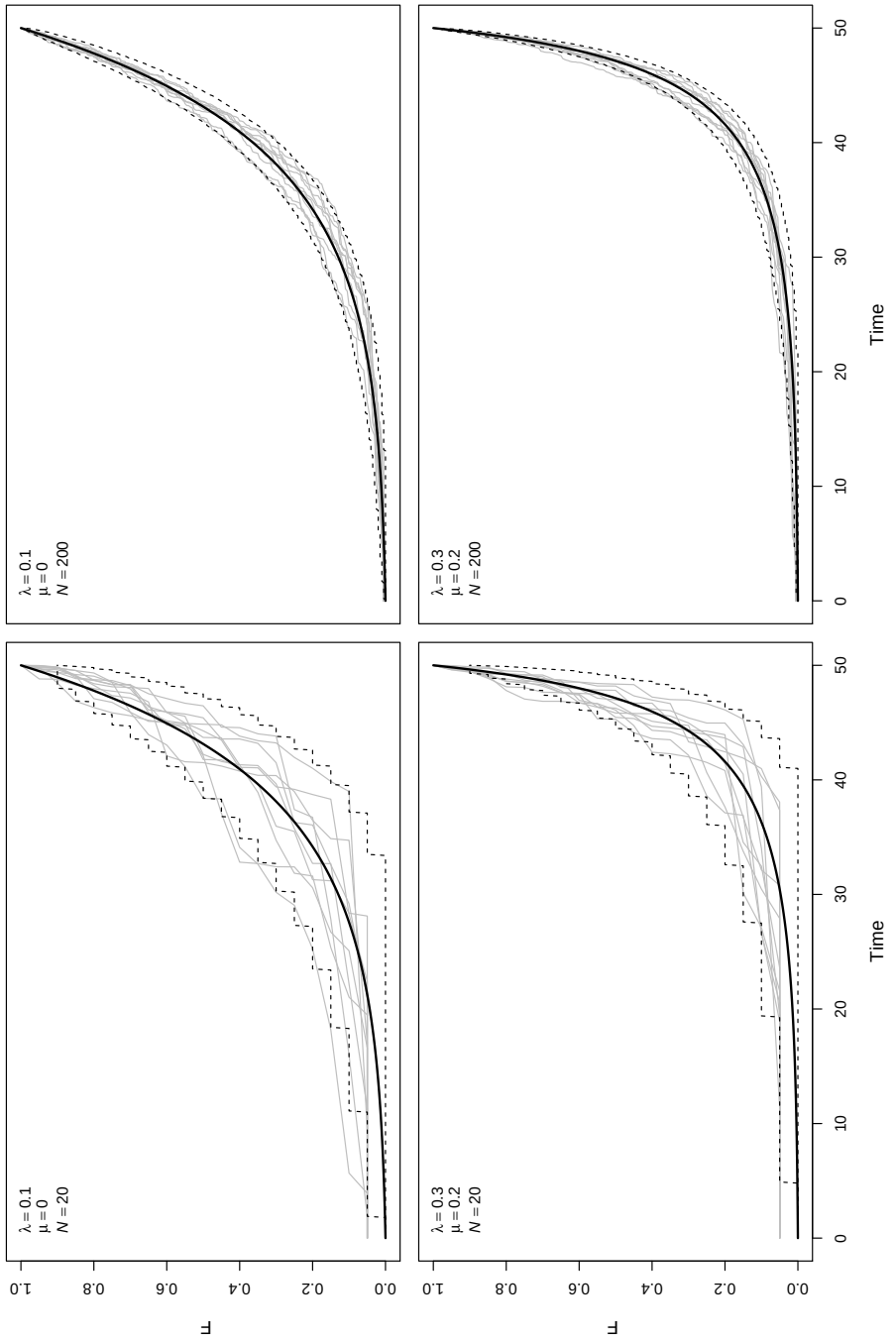

Figure 1: Lineage-through time plots (grey lines) for ten simulated trees with different values of speciation (λ) and extinction (μ) rates and tree size (N). The thick line shows the predicted LTT distribution, and the dashed lines delimit the 95% prediction intervals. The y-axes are scaled to one. Γ: cumulative proportions of lineages.

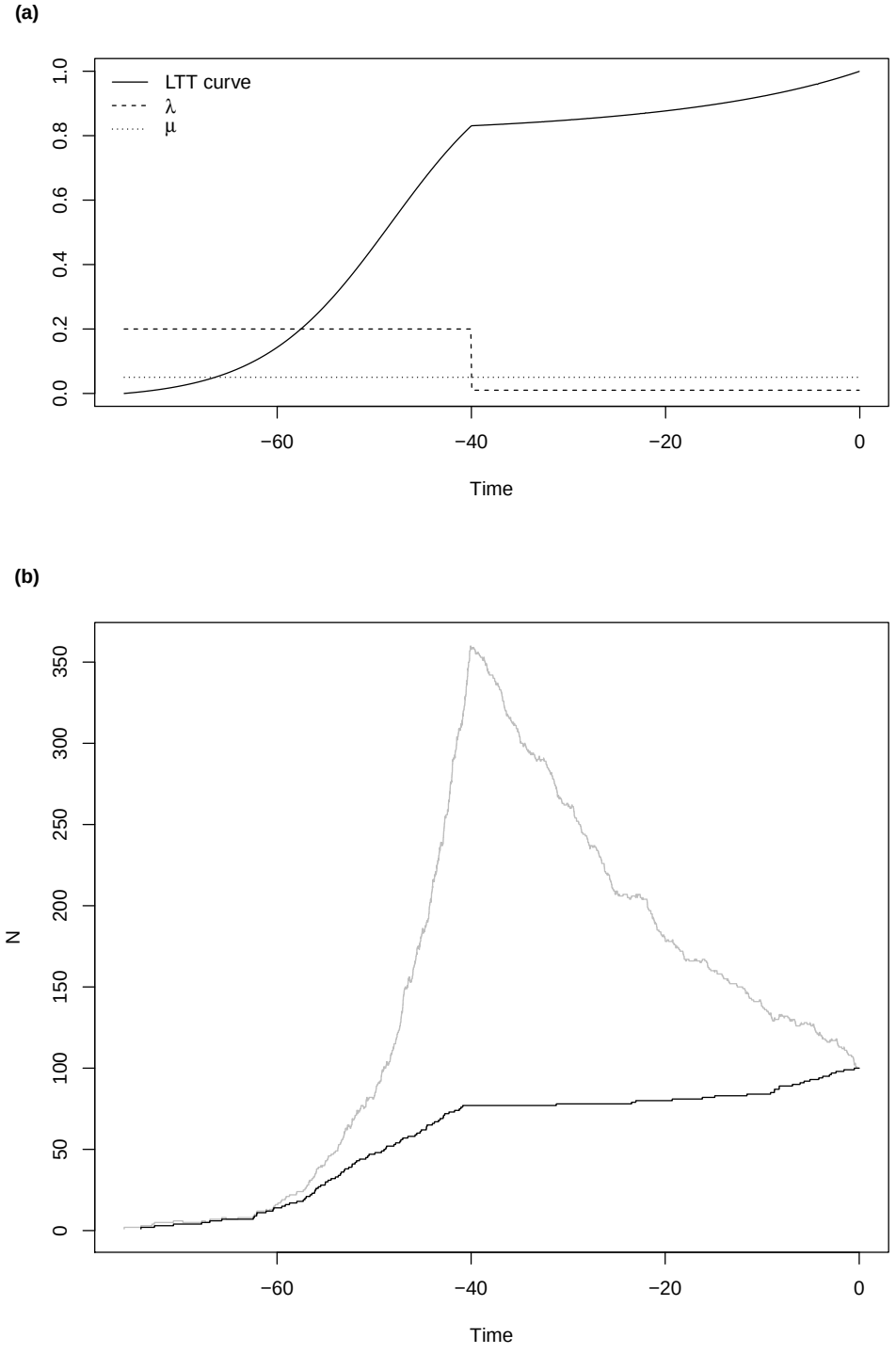

Figure 2: (a) The early burst model of speciation showing the parameters (λ and μ) and the predicted LTT curve. (b) The LTT plot from a phylogeny simulated with this model (black line) and the number of living lineages (grey line) from the same simulation.

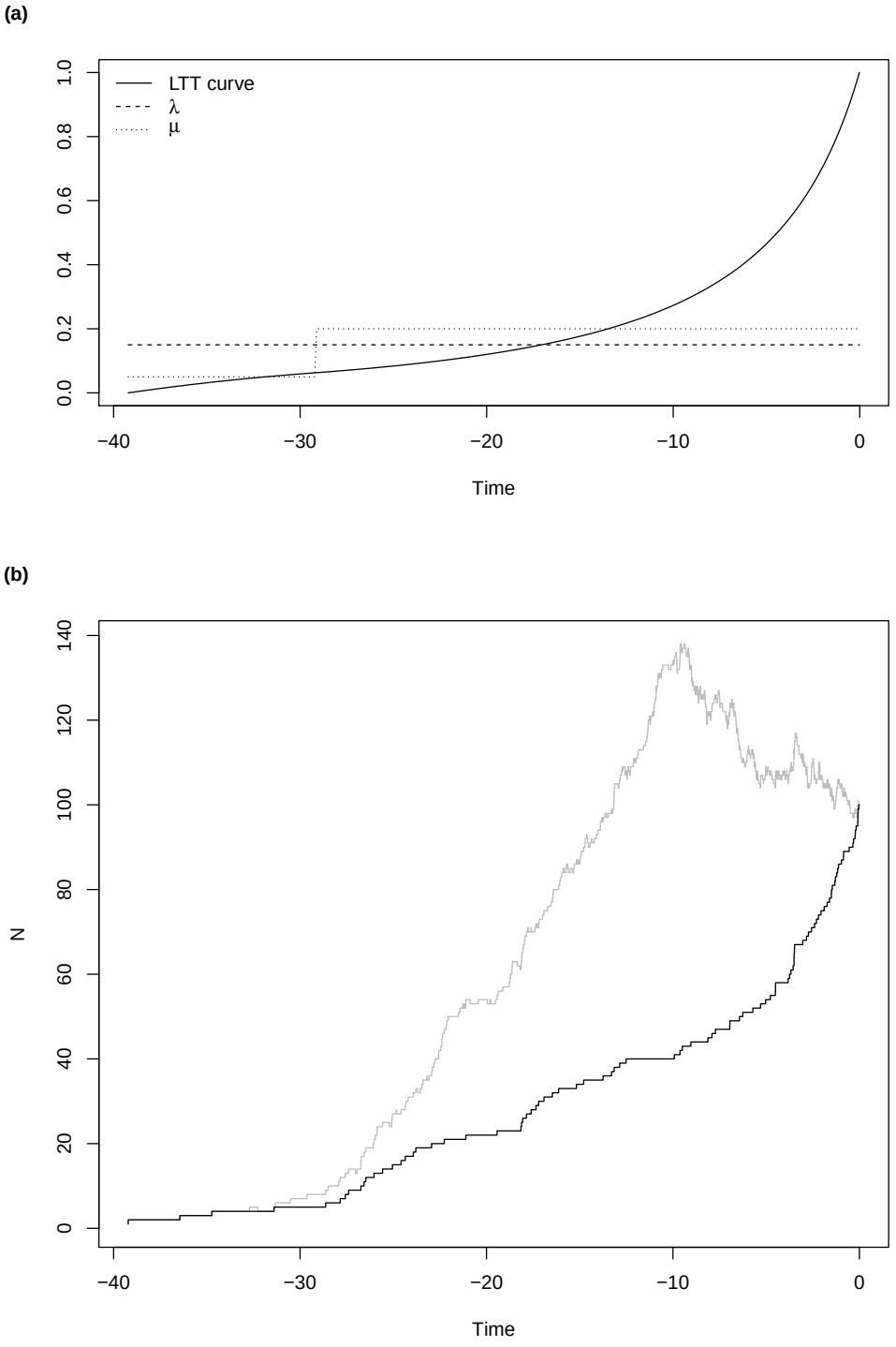

Figure 3: (a) A model of increasing extinction in recent time showing the parameters (λ and μ) and the predicted LTT curve. (b) The LTT plot from a phylogeny simulated with this model (black line) and the number of living lineages (grey line) from the same simulation.