

## Appendix S1 – Details of likelihood computation

Here we provide further details for the likelihood computation of a phylogeny (see the Box in the main text) and discuss two extensions: the inclusion of extant species missing in the phylogeny and the computation of the number of species through time.

### Topology of a phylogeny

The diversity-dependent birth-death model specifies the rates at which species speciate and become extinct. The rates are such that if there are  $n$  species at the time a speciation (extinction) event occurs, then each of these  $n$  species has the same probability to speciate (become extinct). As a consequence, the probability that a certain phylogeny is predicted by the diversity-dependent model is independent of its topology. The probability of a phylogeny is a function of its branching times only. This property is illustrated in Figure S1. The three phylogenies have the same branching times but a different topology. Hence, they have the same probability.

We should point out that the term “probability of a phylogeny” is a misnomer for two reasons. First, we are dealing with a probability density rather than with a probability, because the branching times are continuous variables. Strictly speaking, the probability of a phylogeny is zero. Second, what we call the probability of a phylogeny is really the probability of a vector of branching times, thus using the common terminology for the constant-rate birth-death model. Note that, because of the property mentioned above, the difference between the two probabilities is a combinatorial factor that does not depend on the model parameters, and hence does not affect the parameter estimation.

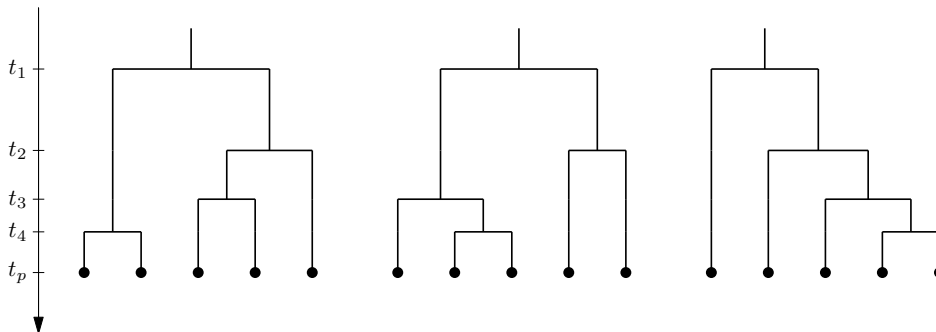


Figure S1: Phylogenies, branching times and topologies. Three different phylogenies are shown for five extant species (indicated by dots) at the present time  $t_p$ . Our definition of the branching times  $t_1, t_2, \dots$  is illustrated on the left axis. The first branching time  $t_1$  corresponds to the crown node; the second branching time  $t_2$  corresponds to the initiation of a third lineage; the third branching time  $t_3$  corresponds to the initiation of a fourth lineage; etc. As a result, there are two lineages between times  $t_1$  and  $t_2$ ; there are three lineages between times  $t_2$  and  $t_3$ ; etc. Time increases from the past towards the present, i.e.,  $t_1 < t_2 < t_3 < t_4 < t_p$ . The three phylogenies have the same branching times, but their topologies, i.e., the order in which lineages undergo branching events, differ. These phylogenies (same branching times, different topology) have the same likelihood under the diversity-dependent birth-death model.

### Likelihood of a phylogeny

As shown in the Box in the main text, the likelihood of a phylogeny can be obtained by integrating  $q - 1$  systems (B2) of linear ordinary differential equations (ODEs). Each system corresponds to the dynamics between two successive branching points  $t_{k-1}$  and  $t_k$ , and transforms a vector of probabilities  $Q_n(t_{k-1})$  into another vector of probabilities  $Q_n(t_k)$ . We denote this (linear) transformation by  $A_k(t_k - t_{k-1})$ . At the branching points  $t_k$  the final

condition of one system is linked with the initial condition of the next system by multiplying  $Q_n(t_k)$  by  $k \lambda_n$ . We denote this (linear) transformation of the vector of probabilities  $Q_n(t_k)$  by  $B_k$ .

The initial condition at the first branching time  $t_1$  can be represented as a vector  $v_1$  of probabilities  $Q_n(t_1)$ . All components of  $v_1$  are zero except for the second component which is equal to one. The final condition at the present time  $t_p$  can be represented as the scalar product of a vector  $v_p$  with the vector of probabilities  $Q_n(t_p)$ . All components of  $v_p$  are zero except for the  $q$ -th component which is equal to one. As a result, the probability  $P(\vec{t})$  of a phylogeny with branching times  $\vec{t} = (t_1, t_2, t_3, \dots, t_{q-1})$  is given by

$$P(\vec{t}) = v_p \cdot A_q(t_p - t_{q-1}) B_{q-1} A_{q-1}(t_{q-1} - t_{q-2}) B_{q-2} \dots A_4(t_4 - t_3) B_3 A_3(t_3 - t_2) B_2 A_2(t_2 - t_1) v_1. \quad (\text{S1})$$

with  $v \cdot w$  the scalar product of the vectors  $v$  and  $w$ .

We use a conditional version of the probability  $P(\vec{t})$  as the likelihood. We condition on the survival of the two lineages initiated at the crown node (otherwise, this node would not be the crown node). The probability  $P_c(t_1, t_p)$  that these two lineages survive can be computed as

$$P_c(t_1, t_p) = v_c \cdot A_2(t_p - t_1) v_1, \quad (\text{S2})$$

with the vector  $v_c$  having the first component equal to zero and all other components equal to one. Finally, the likelihood  $L(\vec{t})$  is defined as

$$L(\vec{t}) = \frac{P(\vec{t})}{P_c(t_1, t_p)}. \quad (\text{S3})$$

We have checked numerically that Eq. (S3) corresponds to known likelihood formulas in two special cases: (a) the linear birth-death model, i.e., when using diversity-independent rates,  $\lambda_n = \lambda$  and  $\mu_n = \mu$ , and (b) the diversity-dependent pure birth model, i.e., when setting the extinction rate to zero,  $\mu_n = 0$ .

## Phylogeny with missing species

Formula (S1) is still valid when including species extant at the present time  $t_p$ , but missing in the phylogeny. It suffices to modify the final vector  $v_p$ . Without missing species, the vector  $v_p$  extracts  $Q_q(t_p)$  from the probability vector  $Q_n(t_p)$ . With, say  $m$ , missing species, the vector  $v_p$  should select the probability  $Q_{q+m}(t_p)$ . Hence, all components of  $v_p$  are equal to zero except for component  $q + m$  which is equal to one.

## Number of species through time

We present an algorithm to compute the (expected) total number of species through time. Note that the total number of species is different from the number of lineages, that is, the number of species with descendants until the present time  $t_p$ . We will present an algorithm for the computation of the (expected) number of lineages through time in Appendix S2.

We define  $P(r(t)|\vec{t})$  as the probability of having  $r$  species at time  $t$  (with  $t_1 < t < t_p$ ) conditional on the phylogeny, which is given by the branching times  $\vec{t}$ . To compute  $P(r(t)|\vec{t})$  we use an approach similar to the likelihood computation (S1), except that, in addition, we condition on the number of species  $r$  at time  $t$ . This conditioning can be accommodated in the computation using a projection operator  $C_r$ . This operator sets all components of  $Q_r(t)$  to zero except for component  $r$ .

Assuming that  $t_{k-1} < t < t_k$ , the probability  $P(r(t), \vec{t})$  of observing the phylogeny  $\vec{t}$  and  $r$  species at time  $t$  is given by

$$\begin{aligned} P(r(t), \vec{t}) = & v_p \cdot A_q(t_p - t_{q-1}) B_{q-1} A_{q-1}(t_{q-1} - t_{q-2}) B_{q-2} \dots \\ & A_{k+1}(t_{k+1} - t_k) B_k A_k(t_k - t) C_r A_k(t - t_{k-1}) B_{k-1} \dots \\ & A_4(t_4 - t_3) B_3 A_3(t_3 - t_2) B_2 A_2(t_2 - t_1) v_1. \end{aligned} \quad (\text{S4})$$

Dividing by the probability  $P(\vec{t})$  of the phylogeny  $\vec{t}$ , we get the conditional probability  $P(r(t)|\vec{t})$ ,

$$P(r(t)|\vec{t}) = \frac{P(r(t), \vec{t})}{P(\vec{t})}. \quad (\text{S5})$$

Formula (S4) can be evaluated by integrating the master equation (B2). However, this approach is computationally demanding: after having integrated from  $t_1$  to  $t$ , the integration from  $t$  to  $t_p$  has to be repeated for all values for  $r$ . Efficiency can be gained by combining the master equation (B2) with its corresponding backward equation,

$$\frac{d}{dt} Q_n^{\text{back}}(t) = \mu_n (n - k) Q_{n+1}^{\text{back}}(t) + \lambda_n (n + k) Q_{n-1}^{\text{back}}(t) - (\mu_n + \lambda_n) n Q_n^{\text{back}}(t), \quad (\text{S6})$$

with  $Q_n^{\text{back}}(t)$  the probability that a realization of the diversification process has  $n$  species at time  $t$  and is consistent with the phylogeny from  $t_p$  back to  $t$ . We have

$$P(\vec{t}) = Q_q(t_p) = Q_2^{\text{back}}(t_1) = \sum_r Q_r(t) Q_r^{\text{back}}(t),$$

illustrating that the master equation (B2) and the backward equation (S6) are interchangeable. We can integrate the master equation (B2) from two species at time  $t_1$  to  $q$  species at time  $t_p$  (first equality), or integrate the backward equation (S6) from  $q$  species at time  $t_p$  back to two species at time  $t_1$  (second equality), or integrate the master equation (B2) from two species at time  $t_1$  to  $r$  species at time  $t$  and the backward equation (S6) from  $q$  species at time  $t_p$  back to  $r$  species at time  $t$ , and take the sum over  $r$  (third equality). In the latter case, we also have

$$P(r(t), \vec{t}) = Q_r(t) Q_r^{\text{back}}(t),$$

so that a single integration of (B2) and (S6) suffices to compute  $P(r(t), \vec{t})$  for all values of  $r$ .

## Further extensions

Further extensions of (S1) based on the master equation (B2) are possible. For example, we can account for species present at the crown node but extinct at the present time. Although these species are not represented in the phylogeny, they do contribute to the diversity-dependent diversification process. It suffices to modify the initial vector  $v_1$  in (S1): with  $m$  additional species, the vector  $v_1$  has a one at position  $m + 2$  and zeros elsewhere.

However, a problem arises for the conditioning (S2). We should guarantee that the additional species at the first branching time  $t_1$  become extinct before the present time  $t_p$ . However, we cannot separate the descendants of these additional species (which should all become extinct) from the descendants of the two crown species (which need not become extinct). In Appendix S2 we show that this problem can be resolved by using a differential equation for a two-dimensional probability distribution, rather than the one-dimensional distribution of the master equation (B2).

## Appendix S2 – Extensions of likelihood computation

Here we discuss further extensions of our computational approach as presented in Appendix S1. We present the computation of the number of lineages through time, and explain how additional species at the crown node can be taken into account.

### Combinatorial properties

We prove two combinatorial properties of the diversification process which we will need below.

#### Distribution of descendants over ancestors

Suppose that there are  $k$  ancestor species at time  $t_a$ , and that each of them has descendant species at a later time  $t_b$ . Denote by  $n_i$  the number of descendants of ancestor  $i$ , and by  $n$  the total number of descendants at time  $t_b$ . We claim that any configuration  $\vec{n} = (n_1, n_2, \dots, n_k)$  with  $n_i \geq 1$  and  $\sum_i n_i = n$  is equiprobable. Note that there are  $\binom{n-1}{k-1}$  such configurations, so that their probability equals  $1/\binom{n-1}{k-1}$ .

To prove this property, we consider the effect of single (speciation or extinction) event on the configuration probabilities. For example, consider  $n$  species and  $k$  ancestors, so that all configurations  $\vec{n}$  have the same probability. Assume a speciation event occurs. The probability  $P^+(\vec{n})$  for a configuration  $\vec{n}$  immediately after the event can be expressed in terms of the probabilities  $P^-(\vec{n})$  immediately before the event:

$$\begin{aligned} P^+(\vec{n}) &= \sum_{i=1}^k \frac{n_i - 1}{n - 1} P^-(\vec{n} - \vec{e}_i) \\ &= \sum_{i=1}^k \frac{n_i - 1}{n - 1} \frac{1}{\binom{n-2}{k-1}} = \frac{n - k}{n - 1} \frac{1}{\binom{n-2}{k-1}} = \frac{1}{\binom{n-1}{k-1}}, \end{aligned} \quad (\text{S7})$$

with  $\vec{e}_i$  a vector with a one at position  $i$  and zeros elsewhere. Hence, all configurations are equiprobable after the speciation event. A similar computation holds for extinction events (the computation uses the probability of lineage extinction, see below). This proves the property because the dynamics between times  $t_a$  and  $t_b$  consist of a series of speciation and extinction events.

#### Probability of lineage extinction

An extinction event of one of the  $n$  species can induce the extinction of one of the  $k$  lineages. Assuming that all configurations of the  $n$  species over the  $k$  lineages are equiprobable, we compute the probability  $P_{\text{ext}}(n, k)$  that a species extinction induces a lineage extinction.

We first compute a related but different probability, namely the probability that a randomly chosen lineage has only one species, and denote it by  $P_{\text{ext}}^*(n, k)$ . A lineage has one species if and only if the remaining  $n - 1$  species belong to the remaining  $k - 1$  lineages. Hence,

$$P_{\text{ext}}^*(n, k) = \frac{\binom{n-2}{k-2}}{\binom{n-1}{k-1}} = \frac{k-1}{n-1}. \quad (\text{S8})$$

On average there are  $k P_{\text{ext}}^*(n, k)$  lineages with one species, which is also the average number of species without other species in its lineage. Hence,

$$P_{\text{ext}}(n, k) = \frac{k P_{\text{ext}}^*(n, k)}{n} = \frac{k(k-1)}{n(n-1)}. \quad (\text{S9})$$

## Number of lineages through time

Here we present an algorithm for the computation of the (expected) number of lineages through time for the diversity-dependent birth-death model. That is, at each time  $t$  we count the number of species with descendants until the present time  $t_p$ . Note that this number is a fraction of the total number of species. We presented an algorithm to compute the (expected) total number of species through time in Appendix S1.

We denote the first branching time (crown node) by  $t_1$  and the present time by  $t_p$ . We compute the number of lineages  $\ell(t)$  at time  $t$  (with  $t_1 < t < t_p$ ) surviving until time  $t_p$ . The computation consists of integrating two systems of ordinary differential equations (ODEs), one from  $t_1$  to  $t$  and another from  $t$  to  $t_p$ .

### From $t_1$ to $t$

First, we integrate the dynamics for the probability of a configuration  $\vec{n}$  starting from the two crown species. Due to the equiprobability of configurations, it suffices to track the probability of any pair  $(n, k)$  with  $n$  the total number of species and  $k$  the number of surviving lineages. We are interested in  $k = 2$  because the crown lineages should survive until the present time  $t_p$ . Hence, we have to compute the probability  $P(n; s)$  that there are  $n$  species at time  $s$  and both crown lineages survive.

We construct a system of ODEs for the probabilities  $P(n; s)$ . A speciation event occurs with rate  $\lambda_n n$ , increases  $n$  by one and leaves  $k = 2$  unchanged. An extinction event occurs with rate  $\mu_n n$ , decreases  $n$  by one, and either keeps  $k = 2$  or gives  $k = 1$ . We are interested in the first possibility, which has probability  $1 - P_{\text{ext}}(n, 2) = 1 - \frac{2}{n(n-1)}$ . Hence, the dynamics for  $P(n; s)$  are

$$\begin{aligned} \frac{d}{ds} P(n; s) &= \lambda_{n-1} (n-1) P(n-1; s) \\ &\quad + \mu_{n+1} (n+1) \left(1 - \frac{2}{(n+1)n}\right) P(n+1; s) \\ &\quad - (\lambda_n + \mu_n) n P(n; s). \end{aligned} \tag{S10}$$

The initial condition at the first branching time  $t_1$  is given by  $P(2; t_1) = 1$ , and all other  $P(n; t_1) = 0$ . We integrate this system of ODEs till time  $t$ . Note that  $1 - \sum_{n \geq 2} P(n; t)$  equals the probability that one of the crown lineages becomes extinct before time  $t$ .

### From $t$ to $t_p$

Second, we consider the  $r$  species present at time  $t$  as ancestor species with descendants at a later time  $s \geq t$ . We integrate the dynamics for the probability of a configuration  $\vec{n}$ . Due to the equiprobability of configurations, it suffices to track the probability of any pair  $(n, k)$  with  $n$  the total number of species and  $k$  the number of surviving lineages at time  $s$  out of the  $r$  lineages at time  $t$ . Denote this probability by  $P(n, k, t; s)$ . As the initial condition for this second step we use the probabilities  $P(r; t)$  obtained in the first step.

We construct a system of ODEs for the probabilities  $P(n, k, t; s)$ . A speciation event occurs with rate  $\lambda_n n$ , increases  $n$  by one and leaves  $k$  unchanged. An extinction event occurs with rate  $\mu_n n$ , decreases  $n$  by one, and either leaves  $k$  unchanged or decreases  $k$  by one. The first possibility has probability  $1 - P_{\text{ext}}(n, k) = 1 - \frac{k(k-1)}{n(n-1)}$ , whilst the second one has probability  $P_{\text{ext}}(n, k) = \frac{k(k-1)}{n(n-1)}$ . In the latter case, we have to guarantee that the extinction does not induce the extinction of a crown lineage. This happens with probability  $1 - P_{\text{ext}}(k, 2) =$

$1 - \frac{2}{k(k-1)}$ . Hence, the dynamics for  $P(n, k, t; s)$  are

$$\begin{aligned} \frac{d}{ds} P(n, k, t; s) &= \lambda_{n-1} (n-1) P(n-1, k, t; s) \\ &+ \mu_{n+1} (n+1) \left(1 - \frac{k(k-1)}{(n+1)n}\right) P(n+1, k, t; s) \\ &+ \mu_{n+1} (n+1) \frac{(k+1)k}{(n+1)n} \left(1 - \frac{2}{(k+1)k}\right) P(n+1, k+1, t; s) \\ &- (\lambda_n + \mu_n) n P(n, k, t; s). \end{aligned} \quad (\text{S11})$$

The initial condition at time  $t$  is given by  $P(k, k, t; t) = P(k; t)$ , and  $P(n, k, t; t) = 0$  if  $n \neq k$ . We integrate this system of ODEs till time  $t_p$ .

Note that  $1 - \sum_{n \geq k \geq 2} P(n, k, t; t_p)$  equals the probability that one of the crown lineages becomes extinct before time  $t_p$ . As a consequence, conditioning that this does not happen corresponds to normalizing the probabilities  $P(n, k, t; t_p)$  with  $n \geq k \geq 2$ . The resulting distribution  $\tilde{P}(n, k, t; t_p)$  gives the probability of a pair  $(n, k)$  with  $n$  the total number of species at time  $t_p$  and  $k$  the number of lineages present at time  $t$  and surviving until time  $t_p$ . Finally, the number of lineages  $\ell(t)$  is given by

$$\ell(t) = \sum_{n \geq k \geq 2} k \tilde{P}(n, k, t; t_p). \quad (\text{S12})$$

We have compared this result with simulations and obtained an excellent correspondence.

## Additional species at crown node

In Appendix S1 we have explained how to take into account additional species at the first branching time  $t_1$ . We argued that the conditioning step requires a system of ODEs for a two-dimensional probability distribution. Indeed, we have to simultaneously guarantee that the two crown lineages survive until  $t_p$  and that all descendants of the additional species at  $t_1$  become extinct before  $t_p$ .

We consider the two crown species and the  $m$  additional species as ancestor species with descendants at a later time  $t > t_1$ . Due to the equiprobability of configurations, it suffices to track the probability of any pair  $(n, k)$  with  $n$  the total number of species and  $k$  the number of surviving lineages at time  $t$  out of the  $m+2$  lineages at time  $t_1$ . Denoting this probability by  $P(n, k; t)$ , the dynamics are

$$\begin{aligned} \frac{d}{dt} P(n, k; t) &= \lambda_{n-1} (n-1) P(n-1, k; t) \\ &+ \mu_{n+1} (n+1) \left(1 - \frac{k(k-1)}{(n+1)n}\right) P(n+1, k; t) \\ &+ \mu_{n+1} (n+1) \frac{(k+1)k}{(n+1)n} \left(1 - \frac{2}{k+1}\right) P(n+1, k+1; t) \\ &- (\lambda_n + \mu_n) n P(n, k; t). \end{aligned} \quad (\text{S13})$$

Note the difference with (S11) in the third line. The initial condition at  $t_1$  is given by  $P(m+2, m+2; t_1) = 1$  and all other  $P(n, k; t_1) = 0$ . The probability we are interested in at time  $t_p$  is  $\sum_{n \geq 2} P(n, 2; t_p)$ .

The same approach can be used to compute the likelihood of a phylogeny with additional species at the crown node and with species extant at time  $t_p$  but missing in the phylogeny. If we want to guarantee that all additional species at the crown node become extinct before time  $t_p$ , then we can use a system of ODEs for a two-dimensional probability distribution  $P(n, k; t)$  analogous to (S13). The variable  $k$  tracks how many of the  $m+2$  species at the crown node have descendants at a later time  $t$ , so that we can impose that none of the  $m$  additional species at the crown node have descendants at the present time  $t_p$ .

## Appendix S3 – Bias in maximum likelihood estimation

Figure S2 shows lineage through time (LTT) plots of the DDL+E model for different parameter combinations. The parameters used in panels a, c, e and g correspond to those in Fig. 1a. The parameters used in panels b, d, f and h correspond to the 50th percentiles of the maximum likelihood estimates calculated for trees simulated with the parameters from panels a, c, e and g, respectively (see Table 1). By comparing panel a with panel b (or c with d and so on), we can see that the geometric mean LTT dynamics are similar between the original parameters (a) and the maximum likelihood estimated parameters (b). However, we see that the range of LTT dynamics that arise under the original parameters (a, c, e and g) is much broader, especially toward the present. The observation that the average LTT expectation is almost identical under quite different DDL+E parameters but that the variance is much greater under the original parameters may explain why we observe bias in the maximum likelihood estimates.

## References

- [1] Harmon L. J., Weir J. T., Brock C., Glor R. E., Challenger W. 2008 GEIGER: investigating evolutionary radiations. *Bioinformatics* **24**, 129–31.

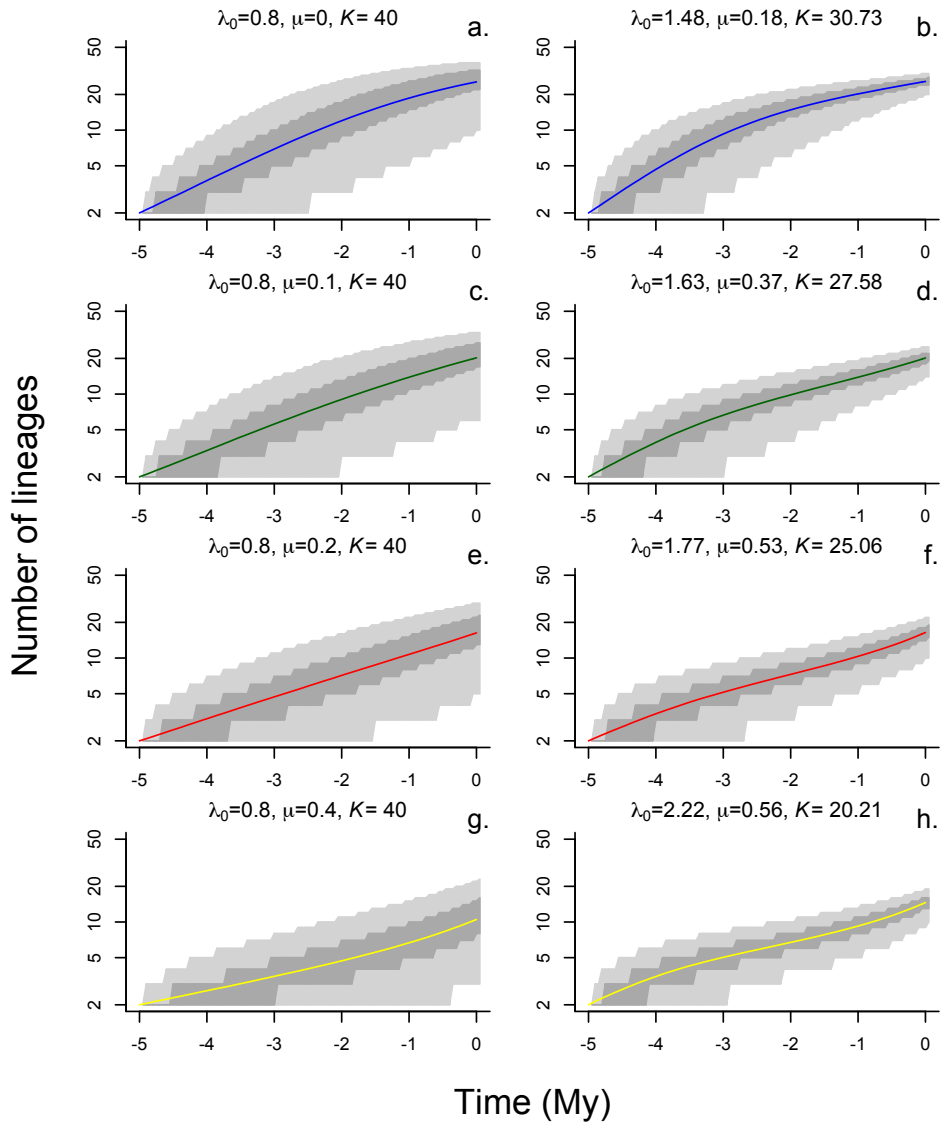


Figure S2: Lineage through time (LTT) plots showing the geometric mean (solid line), interquartile range (dark grey) and 95% confidence interval (light grey) from 10,000 simulations of the DDL+E model under different combinations of parameters. The parameters used in panels a, c, e and g correspond to those in Fig. 1a. The parameters used in panels b, d, f and h correspond to the 50th percentiles of the maximum likelihood estimates calculated for trees simulated with the parameters from panels a, c, e and g, respectively (see Table 1). Tree simulations were conducted using a modified version of the R code in geiger [1].