

## Supplementary Methods and Results

### Simulated Trees for Figure 1

For comparison with the observed  $ED_{t1} \sim ED_{t0}$  relationship, simulated trees were generated representing various scenarios of evolutionary distinctness biased evolution. Simulated tree-growth models and scenarios were based on methods of [3]. Four scenarios were simulated:

- *Null*: Evolutionary Distinctness has no impact on birth and death rates.
- *Panchronic* (Pan.): Evolutionary distinct tips have lower rates of speciation and extinction.
- *Evolutionary Relict* (Rel.): Evolutionary distinct tips have higher rates of extinction and lower rates of speciation.
- *Phylogenetic Fuse* (P.F.): Evolutionary distinct tips have higher rates of speciation and lower rates of extinction.

For each scenario 100 trees were simulated containing both extinct and extant tips. A random tree of 10 tips was used as a seed from which to simulate. Initially, the birth-death rate was 3-1 but after a burn-in period of 10 iterations the birth rate was reduced to 1. Speciation and extinction probabilities were determined directly from the ED values of the species in a tree for each iteration. For example, in a tree of four species each with ED values of 1, 1, 1 and 2 then the latter species, under a panchronic scenario where high ED leads to lower speciation and extinction, will have probabilities of speciating and going extinct as its relative ED value:  $\frac{2}{5}$  or 0.4. Under an evolutionary relict scenario the same species would have the same probability of speciation but a higher probability of extinction, its inverse relative ED:  $1 - \frac{2}{5}$  or 0.6. Trees were then simulated until a total 1000 tips (both representing extinct and extant) were represented in the tree. Simulations were restarted for tree where all tips went extinct before 1000 tips had been reached. Each simulated tree was then sliced into 10 equally spaced time units representing the complete root-to-tip distance. To avoid any influence of the simulation seed tree and burn-in period the first 5 time units were excluded from subsequent analysis. Using the same methods as described in the main methods,  $\log(ED_{t1})$  and  $\log(ED_{t0})$  for all simulated clades were calculated from these time slices (figure S9).

The branch lengths of all the trees were rescaled by changing the age of the simulated trees from 1 to 164 to produce values equivalent the mammalian tree of life. Linear models were then estimated from these datasets for each scenario using R's "lm" function. A process similar to that described in the main methods was then used to identify linear polynomial models that best described the  $\log(ED_{t1}) \sim \log(ED_{t0})$  relationship of each scenario.

All of the scripts for the analysis described in this section are available at the main author's GitHub repository (<https://github.com/DomBennett/Project-karenina>) under the folder "additional\_analysis".

## Mixed-Effects Models

Linear mixed-effects models (LMEMs) [36, 37] are an extension of generalised linear models and were originally developed to model relationships in experimental situations where important factors that influence the response variable are outside of the experimenter's control. In LMEM terminology the variables outside of control are referred to as "random-effects" whereas the variables to be modelled are termed "fixed-effects". In our case, the response variable is  $ED_{t1}$  and the main factor outside of control is *epoch*. The epoch transitions are outside of our control, as they do not represent the same length of time-step, they have different number of taxa and have different tree ages (i.e. the age of the Mammalian tree in the Jurassic was younger than it is today). By specifying epoch in a model, we are able to model future ED as a function of past ED ( $ED_{t1} \sim ED_{t0}$ ) while estimating the intercepts and slopes separately for each epoch. Functionally, this is equivalent to modelling each epoch separately but with the added benefits that the overall relationship between past and future ED can be estimated; issues of multiple significance tests are avoided; more than just epoch can be considered as a random-effect; and the influence of the random-effects can be better controlled. On this latter point, the LMEM approach allows us to control whether or how we estimate the intercepts and slopes of the random effects. For example, in our schematic, the model formula  $ED_{t1} \sim ED_{t0} + (1/epoch)$  indicates modelling  $ED_{t1}$  as a function of  $ED_{t0}$  while considering epochs separately, in which case an independent intercept will be estimated for each epoch. Overall, however, the slope of the model will be estimated across all epochs. To also control for the influence of the random-effect in trend, not just scale, we can estimate random slopes for each epoch by changing the previous formula's random-effects structure to  $ED_{t1} \sim ED_{t0} + (ED_{t0}/epoch)$ .

The random effects formula structure can be used to add additional random effects. In addition to controlling for epoch transitions, we used LMEMs to also control for non-independence of taxonomic groups as we expect closely related clades to have similar EDs, e.g. multiple genera of Rodents may have low average ED values while multiple genera of Afrotheria may have high average

ED values. The additional taxonomic information can be specified in the random effects structure by sub-setting by orders, e.g.  $ED_{t1} \sim ED_{t0} + (1/epoch) + (1/order)$ , or species/clade IDs, e.g.  $ED_{t1} \sim ED_{t0} + (1/epoch) + (1/id)$ . The random effects structure can also allow hierarchical categories with which taxonomic ranks can be specified, e.g.  $ED_{t1} \sim ED_{t0} + (1/epoch) + (1/order/genus/id)$ . For each extra level in the hierarchy, however, there is a non-linear gain in computation time, limiting the maximum number of levels that can be run.

Our modelling approach was not to include all these terms into a single model from the start. Instead we began with the most basic model that consisted of only the response variable and the explanatory variable and then added these extra terms and random-effects structures to create increasingly complex models in order to identify the simplest model that can explain the most observed variance. In tables 1 to 3 in the main text, each successive row shows a more complex model. We tested for significant differences in explained variance between these models using ANOVA and the Akaike Information Criterion (AIC) [38]. AIC is a measure of a model's likelihood weighted by the number of estimated parameters, the lower its value the better the model fit.

### **Comparing the *Real* and the *Random***

Before modelling the relationship of  $ED_{t0}$  and  $ED_{t1}$ , we first tested whether the *real* distribution of trees (where fossils are added to trees based on fossil age and taxonomy) and their ED values differed significantly from the *random* distribution (where fossils are added randomly). For the random placement, a mean  $19,027 \pm 1$  fossil tips were added to the original 4,510 tipped mammalian supertree for each iteration. Initially, we determined whether the estimates of ED of the *real* and *random* had different distributions by comparing the mean and variance of the estimated values using the t-test and F-test, respectively. Secondly, we tested whether ED estimates of shared nodes was greater or smaller for the *real* and *random* distributions. *Shared nodes* are the nodes, which appear multiple times across the iterations. Because fossil placement in the *random* is not informed by taxonomy, shared nodes in

the *random* are either due to their existing already within the tree or chance. If taxonomic information informs the stochastic adding process, we should expect ED estimates of shared nodes across the iterations of the *real* to vary less than those in the *random*. To test this, we calculated the standard deviation of ED values estimated for shared nodes across the iterations, and tested whether the *real* had a lower variance using a t-test.

The *random* dataset was much bigger (213,427) than the *real* (115,810) and the distribution of shared nodes was more even (1 - 0%, 5 - 25%, 23 - 50%, 49 - 75%, 100 - 100%). Like the *real*, the *random* distribution also showed a positive non-linear relationship between  $ED_{t0}$  and  $ED_{t1}$  and differences between the epoch-to-epoch transitions were great, particularly at low ED values (figure S4). As for the *real*, we removed JU-CL and CL-CU transitions from all subsequent analysis.

The estimated ED values of  $t1$  and  $t0$  ( $\Delta ED$ , calculated as  $ED_{t0} - ED_{t1}$ ) for all nodes across all iterations differed significantly between the *real* and *random*. The mean  $\Delta ED$  for the *real* was lower (-0.47) than that of the *random* (-0.41) (t-test,  $t = -58.268$ ,  $p < 0.001$ ). The spread also differed significantly. The variance of the *real*  $\Delta ED$  distribution (0.125) was significantly greater than the *random* (0.042) (F-test,  $F = 0.3385$ , variance ratio = 0.340,  $p < 0.001$ ). Additionally, the standard deviation of the mean  $\Delta ED$  for shared nodes across the iterations was much greater for the *random* (0.401) than the *real* (0.327) (t-test,  $t = 100.11$ ,  $p < 0.001$ ). (See figure S5.)

### **Comparing Model Outcomes between the *Real* and *Random***

Due to the possibility that the non-linear observed relationship, as described in the main text, is a consequence of the random placement of fossils in the iterated trees, we modelled the same relationship as for the *real* with the *random*. Although we found a similar relationship for the *random* iterated tree distribution using an equivalent model to *obs2*, we found the predicted values for  $ED_{t1}$  at high  $ED_{t0}$  were not as extreme (figure S7). Additionally, we compared the trend between  $ED_{t0}$  and  $ED_{t1}$  using a

General Additive Model for species/clades that were shared 50 or fewer times across the *real* iterations to those shared more than 50 times. The species/clades shared more than 50 times showed a stronger nonlinear relationship (figure S8). Therefore the points of which we can be more confident indicate a stronger nonlinear relationship indicating that any error in the stochastic fossil-adding process is likely to have *underestimated* the nonlinear nature of the relationship rather than caused it.

The similarity of results between the *real* and *random* datasets may be attributable to their both sharing the same base tree and the same ranges of fossil taxonomies and age ranges. This latter reason may have constrained the *random* process to placing tips in similar positions to the *real*. Furthermore, despite the taxonomic constraint there were large numbers of species/clades that were effectively placed in the *real* distribution randomly, as their positions were unique (i.e. not repeated in any other iteration). This finding applies particularly to early fossils, which had few branches on which to be added, due to the paucity of closely related species in the Recent. Where there are few branches and many fossil records, new tree sections, unique to each iteration, are generated as fossil records combine in novel ways. This difference in the spread of shared nodes may be due to the taxonomic constraint causing the consistent placement of uncertain nodes in unique points while also mapping certain nodes to the same locations. Additionally, the finding that so large a proportion of the *real* dataset was essentially added randomly indicates that increasing the number of iterations beyond one hundred is unlikely to have changed these results.

## Supplementary Figures

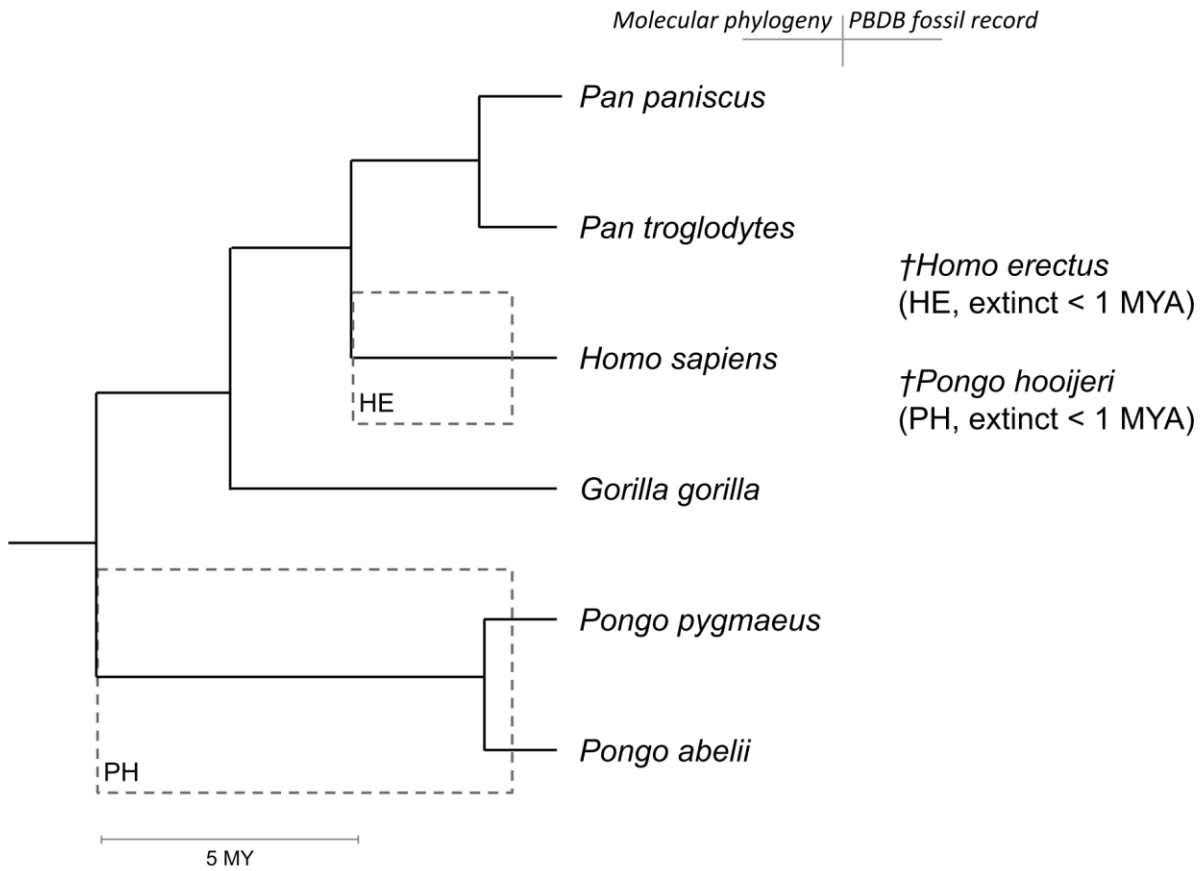


FIGURE S1. Fossil placement using treeman's pinTips(). This example uses a phylogenetic tree of extant apes, and two fossil ape species. Comparing the taxonomy of the tree and the fossil records we can identify branches the fossils could potentially branch from; these windows of placement are then limited by the expected occurrence of the fossil. In the case of *Homo erectus*, its genus is *Homo* and there is only one extant member of the *Homo* genus in the tree. As such, a *H. erectus* branch can begin, within the HE hatched box, anywhere along the pending edge of *H. sapiens* and end at any point within the HE hatched box. In the case of *Pongo hooijeri*, the branch can begin, within the PH hatched box, anywhere along the parent branch of *P. pygmaeus* and *P. abelii* or either of the pendant edges of *P. pygmaeus* and *P. abelii*.

## Eocene – Present

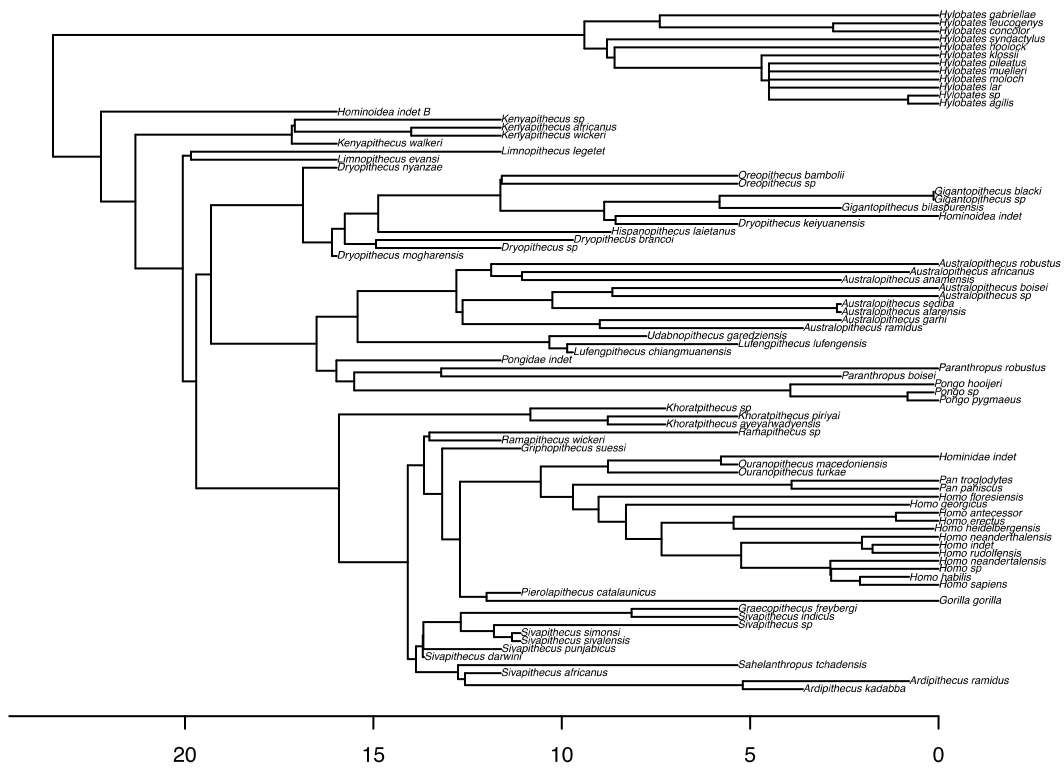


FIGURE S2.1 Example output of taxonomically constrained stochastic fossil placement: extinct and extant apes (Hominioidea) from the Eocene to the present.



### Miocene (14 MYA)

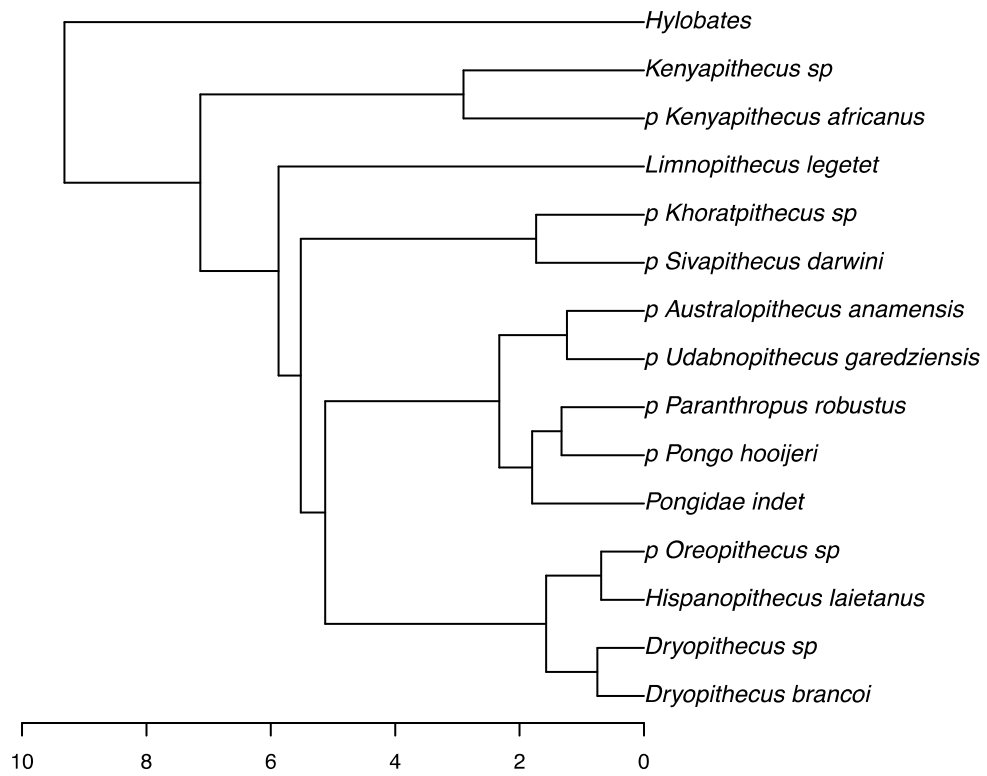


FIGURE S2.2 Example output of taxonomically constrained stochastic fossil placement: a reconstructed tree of extant apes during the Miocene. The tree appears ultrametric as it is a time slice taken from the tree in figure S.1.

**Pleistocene (1.3 MYA)**

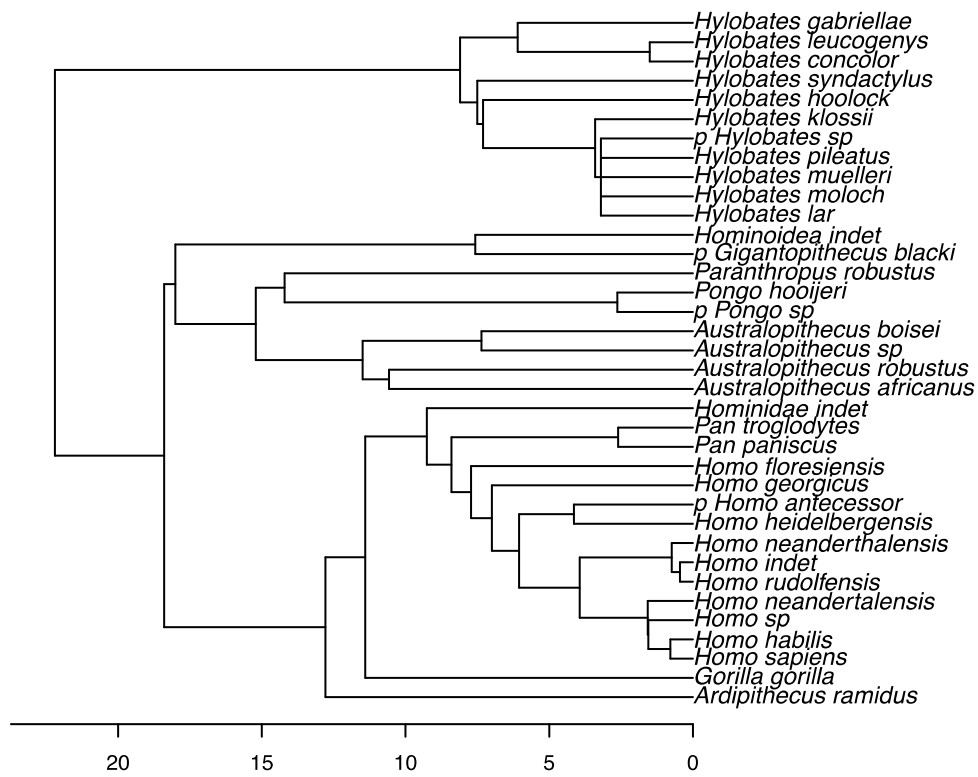


FIGURE S2.3 Example output of taxonomically constrained stochastic fossil placement: reconstructed tree of extant apes during the Pleistocene. The tree appears ultrametric as it is a time slice taken from the tree in figure S.1.

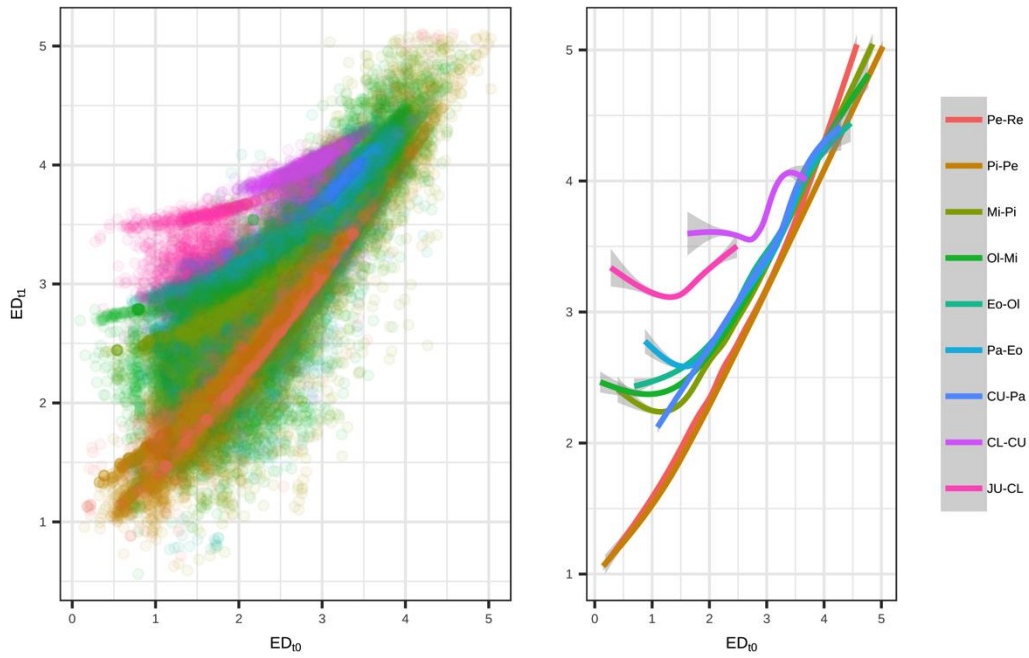


FIGURE S3. ED values of clades in an epoch ( $ED_{t0}$ ) against the following epoch ( $ED_{t1}$ ) for estimates generated from the *real* distribution of molecular-fossil trees. Left, points of all clade/species coloured by specific epoch-epoch transition. Right, for visual purposes, estimated General Additive Models by epoch. Colours indicate: Pleistocene to Recent (Pe-Re), Pliocene to Pleistocene (Pi-Pe), Miocene to Pliocene (Mi-Pi), Oligocene to Miocene (Oi-Mi), Eocene to Oligocene (Eo-Oi), Paleocene to Eocene (Pa-Eo), Cretaceous Upper to Paleocene (CU-Pa), Cretaceous Lower to Cretaceous Upper (CL-CU) and Jurassic to Cretaceous Lower (JU-CL).

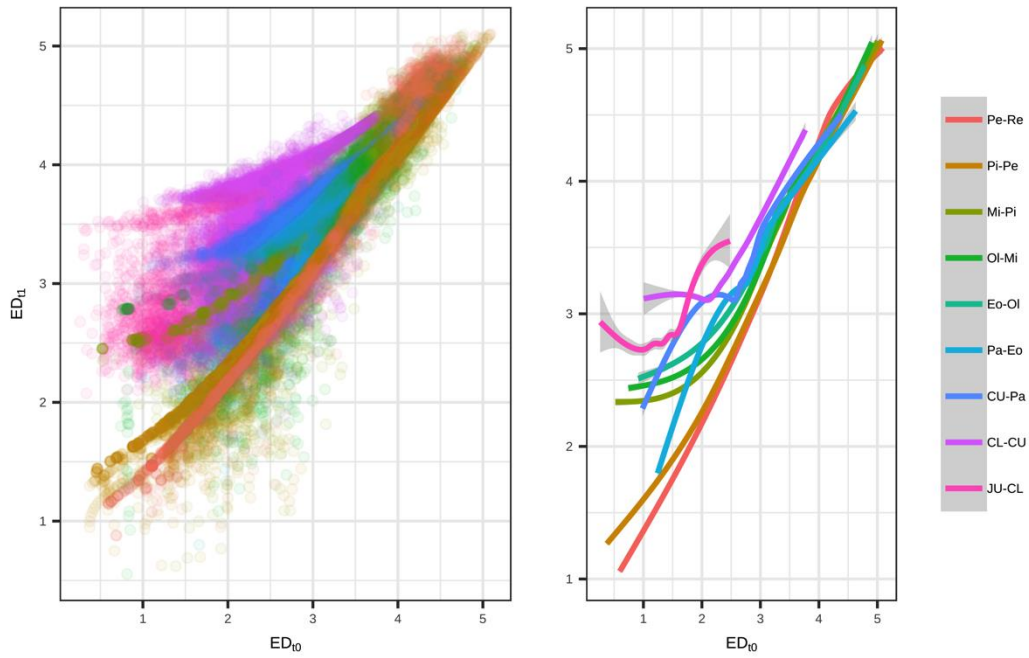


FIGURE S4. ED values of clades in an epoch ( $ED_{t_0}$ ) against the following epoch ( $ED_{t_1}$ ) for estimates generated from the *random* distribution of molecular-fossil trees. Left, points of all clade/species coloured by specific epoch-epoch transition. Right, for visual purposes, estimated General Additive Models by epoch. Colours indicate: Pleistocene to Recent (Pe-Re), Pliocene to Pleistocene (Pi-Pe), Miocene to Pliocene (Mi-Pi), Oligocene to Miocene (Oi-Mi), Eocene to Oligocene (Eo-Oi), Paleocene to Eocene (Pa-Eo), Cretaceous Upper to Paleocene (CU-Pa), Cretaceous Lower to Cretaceous Upper (CL-CU) and Jurassic Upper to Cretaceous Lower (JU-CL).

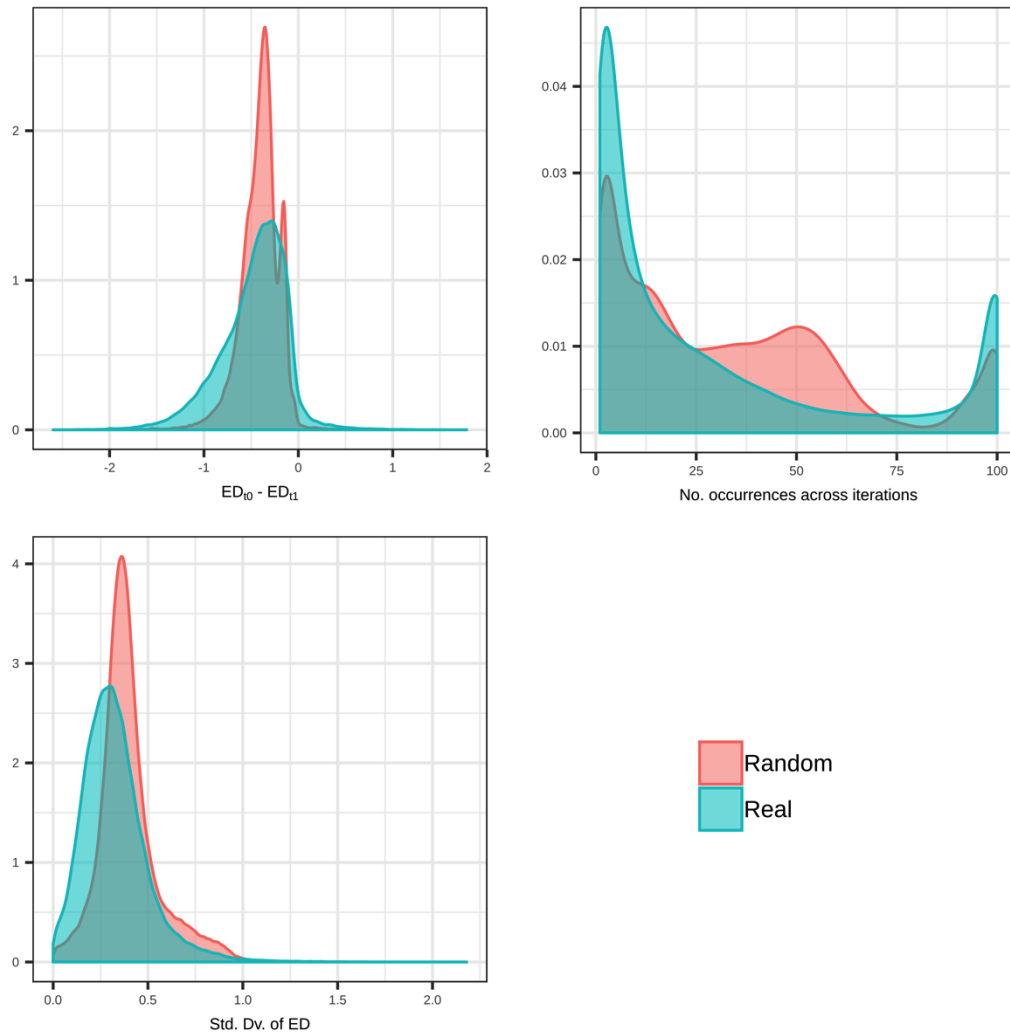


FIGURE S5. Comparing the *real* and *random* tree distributions; the real distribution has lower ED values and less variance. Top-left, the *real* molecular-fossil tree distribution has a greater range of mean  $\Delta ED$  values calculated across iterations than the *random* distribution. Top-right, proportion of times each node is shared across iterations for both *real* and *random*. Bottom-left, the *real* distribution shows less variance of  $\Delta ED$  for identifiable clades (shared nodes) across the distribution than does the *random*.

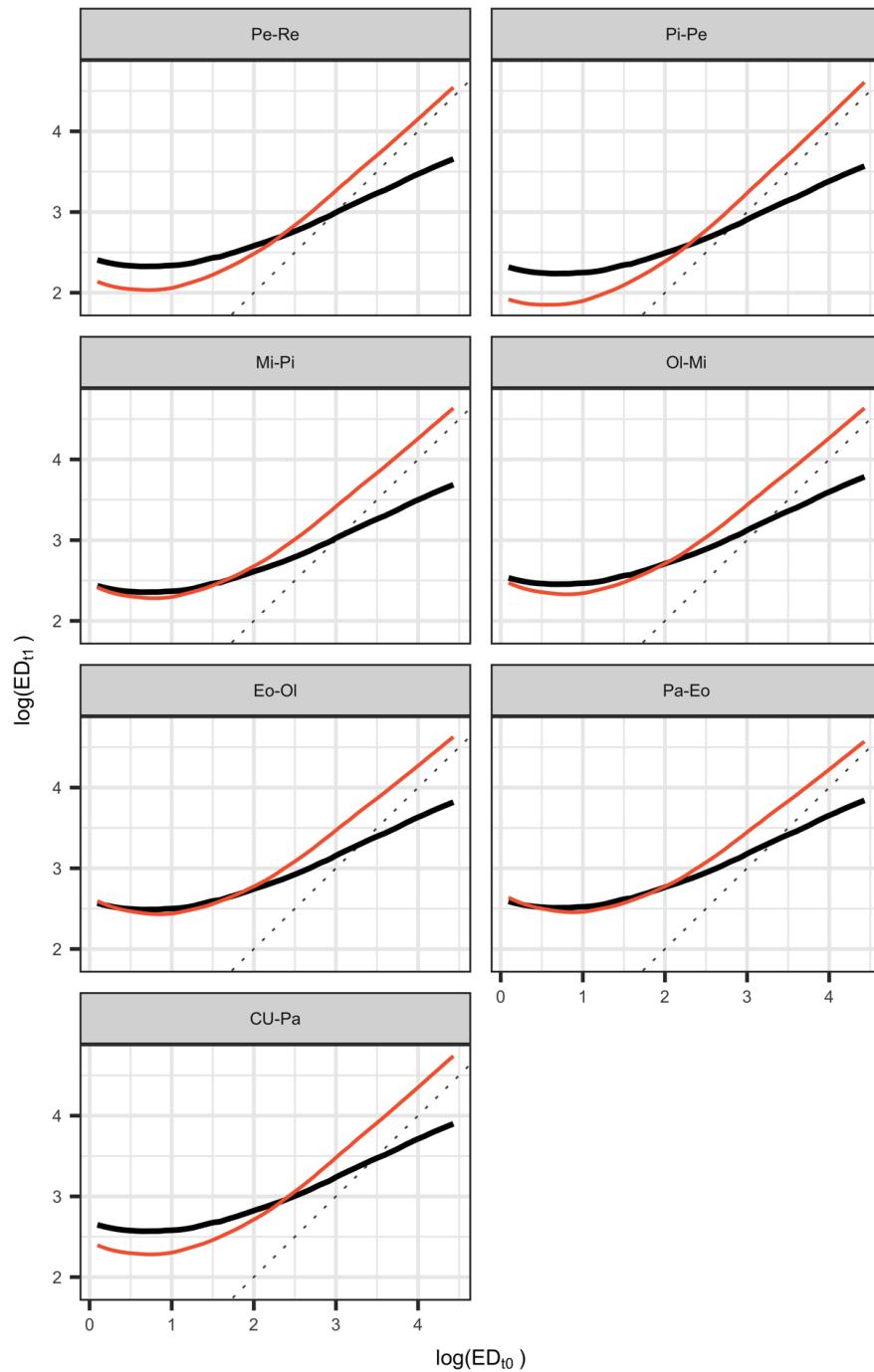


FIGURE S6. Predicted  $\log(ED_{t1})$  values generated from the observed non-linear model (*m3b*, solid red line) and the expected model (*nlg*, solid black line) for a representative dataset of a range of  $\log(ED_{t0})$  values and a random subset of one hundred genera across the different epoch-to-epoch transitions. The dotted line indicates perfect linear relationship. For plotting,  $\log(ED_{t1})$  estimates across the different genera were median averaged. Epochs: Pleistocene to Recent (Pe-Re), Pliocene to Pleistocene (Pi-Pe), Miocene to Pliocene (Mi-Pi), Oligocene to Miocene (Oi-Mi), Eocene to Oligocene (Eo-Oi), Paleocene to Eocene (Pa-Eo) and Cretaceous Upper to Paleocene (CU-Pa).

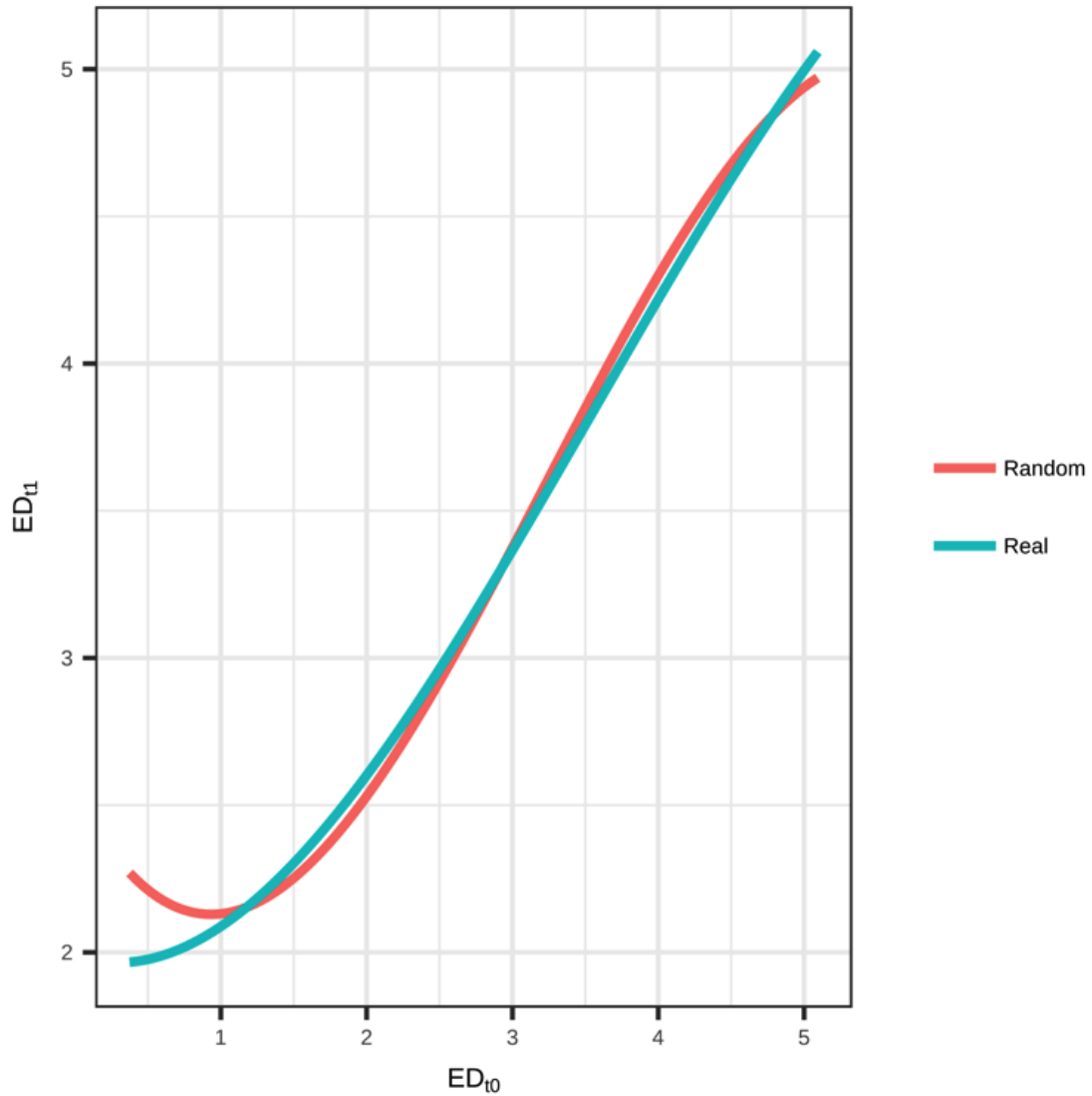


FIGURE S7. Predicted  $ED_{t1}$  values generated from trinomial models with random effect structure of (1|genus), a comparable model to *m3b* but which can be generated for both *real* and *random* molecular-fossil distributions. Although both lines are similar, the *real* line has higher  $ED_{t1}$  estimates for high  $ED_{t0}$  and lower  $ED_{t1}$  estimates for low  $ED_{t0}$ , indicating that the observed  $ED_{t1}$ - $ED_{t0}$  relationship may be more conservative than reality due to errors in the stochastic fossil-adding process. Values are generated for a representative dataset of a range of  $ED_{t0}$  values, a random subset of one hundred genera and all the epoch-to-epoch transitions. For plotting,  $ED_{t1}$  estimates across the different genera and epoch-to-epoch transitions were median averaged.

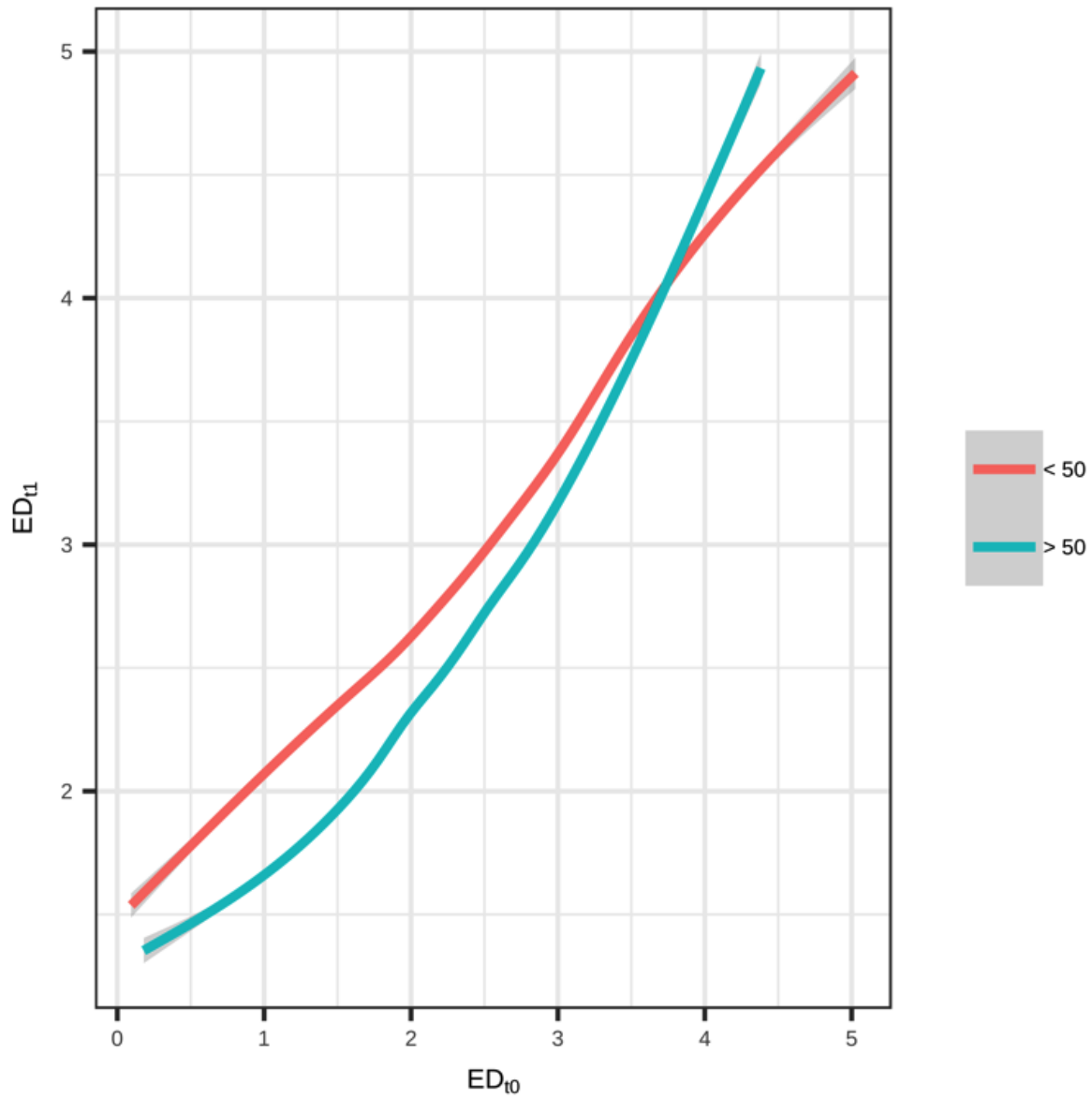


FIGURE S8. Generative Additive Models to explore the  $ED_{t0}$  and  $ED_{t1}$  relationship using subsets of the *real* dataset. Blue/green line indicates clades/species points that were shared over 50 times across the iterations of the *real* (high confidence nodes). Red/orange line indicates clades/species points that were shared 50 or fewer times across the iterations of the *real* (low confidence nodes). Points in which we have higher confidence show a stronger nonlinear relationship, with higher estimates for  $ED_{t1}$  at high  $ED_{t0}$ .



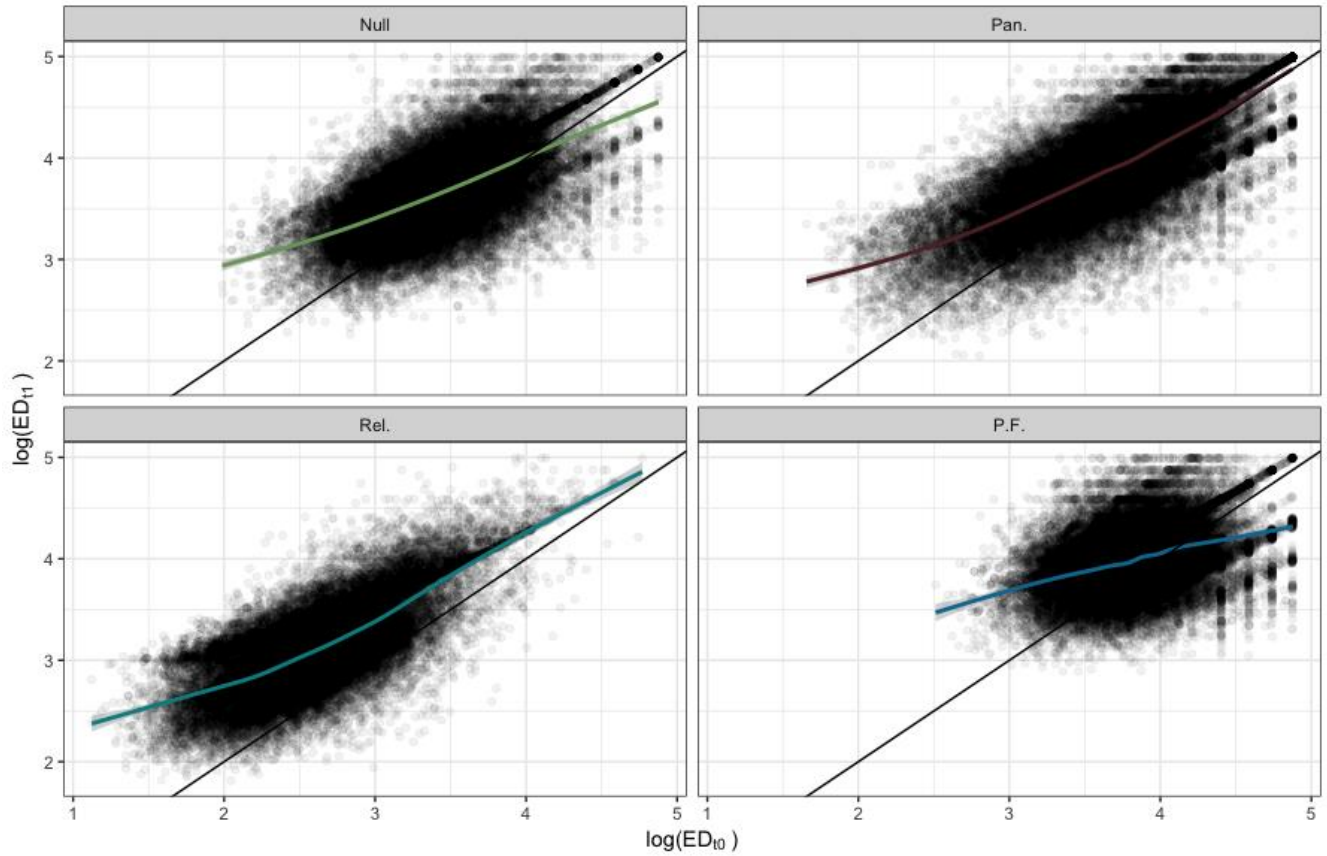


Figure S9. Future ED as a function of past ED for four simulated scenarios: non-biased birth-death model (Null), evolutionary distinct tips have lower rates of extinction and speciation (Pan.), evolutionary distinct tips have higher rates of extinction and lower rates of speciation (Rel.) and evolutionary distinct tips have lower rates of extinction and higher rates of speciation (P.F.). Solid, coloured lines represent General Additive Models, solid black line represents  $\log(ED_{t0}) = \log(ED_{t1})$ .

## Supplementary Tables

TABLE S1. Epoch-to-epoch mid-point estimates taken from the International Chronostratigraphic Chart (2013) used for estimating changes in evolutionary distinctnesses from the inferred molecular-fossil mammalian phylogenetic tree.

<b>Period</b>	<b>Code</b>	<b>Span (MYA)</b>	<b>Time</b>
Pleistocene - Recent	Pe-Re	1.30 - 0.00	1.30
Pliocene - Pleistocene	Pi-Pe	3.96 - 1.30	2.66
Miocene - Pliocene	Mi-Pi	14.18 - 3.96	10.22
Oligocene - Miocene	Ol-Mi	28.47 - 14.18	14.28
Eocene - Oligocene	Eo-Ol	44.95 - 28.47	16.49
Paleocene - Eocene	Pa-Eo	61.00 - 44.95	16.05
Cretaceous Upper - Paleocene	CU-Pa	83.25 - 61.00	22.25
Cretaceous Lower - Cretaceous Upper	CL-CU	122.75 - 83.25	39.50
Jurassic Upper - Cretaceous Lower	JU-CL	154.25 - 122.75	31.50