

# Diversification Lives

There has long been a debate in phylogenetics about the possibility of estimating speciation and extinction rates from trees of extant species. Nee et al. (1994) showed that, mathematically, both rates could be estimated; Kubo and Iwasa (1995) showed that estimation of extinction rate had high uncertainty and that an infinite array of models with gradually changing speciation and/or extinction rates would be indistinguishable; Rabosky ( ) ***showed that a constant rate model could get bad estimates if the true model were one with rapidly changing speciation rates; Beaulieu and O’Meara ( ) showed that for more biologically plausible models and large enough trees that a constant rate model could still work. There are a wide variety of papers developing models that seek to fit changing speciation or extinction rates using modern trees ( ), some, given the empirical difficulties in estimating extinction rates, seek to estimate just their difference, the diversification rate ( ), often by assuming extinction rate is constant. There are also a wide array of models that examine the correlation of traits with diversification, speciation, and/or extinction rates ( ), collectively known as SSE models (for state speciation and extinction). There are over ,000 papers that cite these models.***

Louca and Pennell (2020), and the Pagel (2020) commentary on this, put all of this enterprise in doubt. Formally, Louca and Pennell (2020) present a substantially similar argument, though much more detailed, to the one presented by Kubo and Iwasa (1995) a quarter century ago that for a given phylogeny, there are an infinite array of models that have speciation and/or extinction rate smoothly vary through time that can equally well fit the tree. Their solution is not to give up all analyses of how diversification has changed through time from modern chronograms, but instead use the pulled diversification rate to examine patterns of variation through time: look at one parameter that does not map to a biological process, in the same way that effective population size is a useful parameter in population genetics that does not map directly to a readily observable biological estimate like census population size. However, the popular interpretation of their work, assisted by reviews such as Pagel (2020), is that any attempt to learn about diversification from modern chronograms is futile. This interpretation is not correct.

We focus here on SSE models. These models are known to have issues. Maddison and FitzJohn (2015) show that single changes in characters may be overinterpreted as showing strong evidence for a diversification pattern change. If a tree evolves under a process where diversification patterns may change on some branches, these methods may try to correlate these changes with a character if forced to choose between that and an incorrect model of homogeneous patterns (Rabosky and Golberg ( ); Beaulieu and O’Meara ( )), though this may reflect more misapplication of statistics by biologists than a flaw in the methods per se (Beaulieu and O’Meara ( )). There might need to be a lot of data to infer rates accurately (Davis ( )). There are approaches that can deal with some of these issues: check to see if there is a consistent diversification pattern across trait origins (Beaulieu achene), provide rich models that allow rate variation not to be shoehorned into available characters (hisse\_, **geohisse**), etc. But what does this new work suggest for these?

In fact, the possibility of trait-dependent diversification models still working is left unresolved (Louca and Pennell (2020), S.6) – the authors believe the chance of identifiability is slim but acknowledge this is not proven. It could be worth distinguishing the methods that are weeds from those that are useful before burning a field to the ground.

For models that do not change speciation and extinction over time, there is a well-behaved likelihood surface (Nee et al. 1994), complete with a peak.

```
## function (save = "default", status = 0, runLast = TRUE)
## .Internal(quit(save, status, runLast))
## <bytecode: 0x7fd7d25391c0>
## <environment: namespace:base>
```

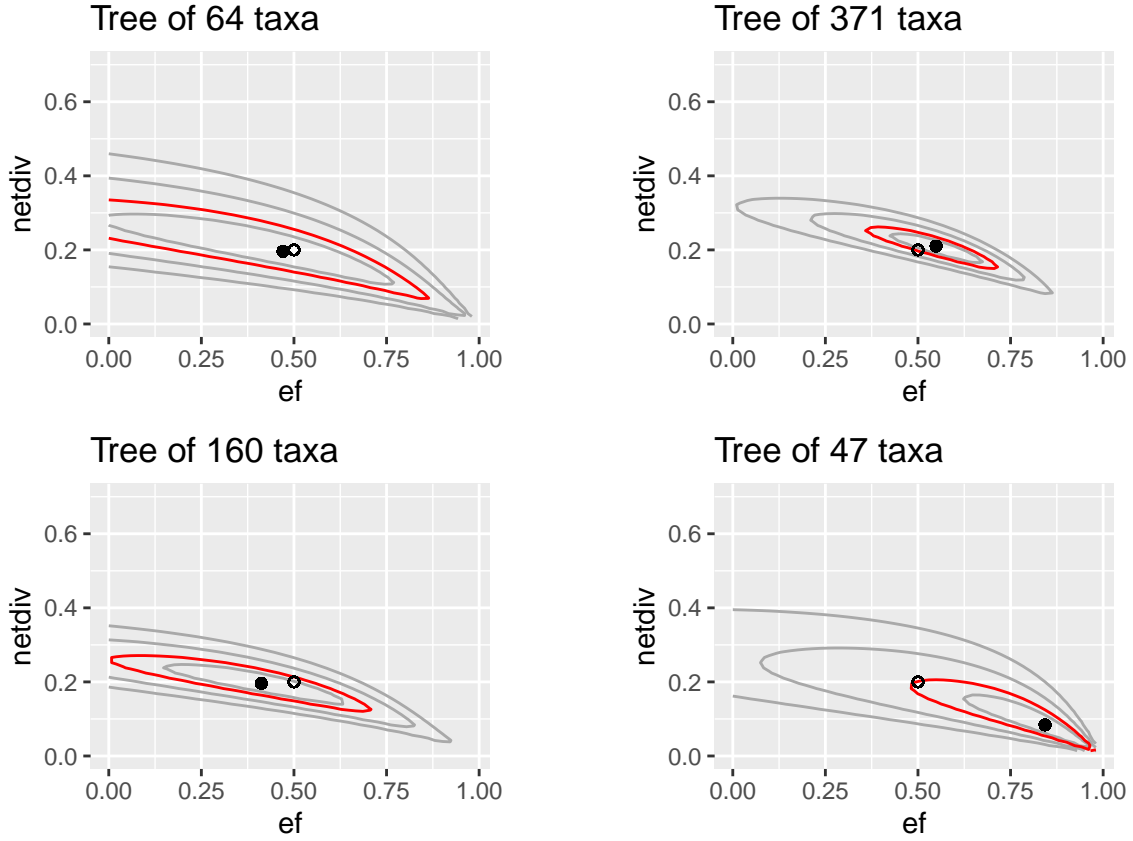


Figure 1: Likelihood surface for constant rate birth death model from four simulations with the same parameters. Birth rate = 0.4, death rate = 0.2, total time of 20. The black dot shows the maximum likelihood point; The open circle shows the true generating values; the red contour line shows all the points within 2 lnL units of the best point. Other contour lines are 1, 5, and 10 lnL units away from the optimum

As shown by the Fig. 1 (which mimics a similar one by Nee et al. 1994), there is a well-defined peak, albeit substantial uncertainty in extinction fraction: not surprising, given the size of these trees. The identifiability issues of models that allow speciation and/or extinction rates to gradually change have also long been known, but sadly ignored. Kubo and Iwasa (1995) showed that a gradual speciation rate change model with constant extinction rate could produce the same lineage through time curve as a constant speciation rate model with gradual extinction rate change, but, rather than just presenting this troubling special case, as their work was categorized by Louca and Pennell (2020), they explicitly state that “There are infinitely many cases intermediate between these two that also generate the same  $\ln(N_t)$  pattern, in which both the branching rate and the extinction rate change with time.”

Louca and Pennell (2020) expand on the mathematical treatment of Kubo and Iwasa (1995). Louca and Pennell (2020) show that two models are "congruent if and only if they have the same  $r_p$  and the same  $\lambda_p$  at some time point in the present or past (for example the same product  $\rho * \lambda_o$ )." Their pulled diversification rate is:

Their pulled diversification rate is (supplement equation 10):

$$r_p := \lambda - \mu + (d \lambda / d T) / \lambda$$

that is, the pulled diversification rate at a given time is equal to the normal estimate of diversification rate at that time plus the slope of how  $\lambda$  changes with time divided by the value of  $\lambda$  at that time.  $\lambda$  is the speciation rate at some time, which can change in their general model; in some equations this is represented as a single variable (i.e., supplement equation 10) but in other times a function of time (supplement equation 12). To make the derivation easier to follow we adopt the latter notation:  $\lambda(u)$  is the speciation rate at time  $u$  in the past; the present is time 0. Rewriting the above equation using this notation:

$$r_p(u) := \lambda(u) - \mu(u) + (d \lambda(u) / d T) / \lambda(u)$$

Take the simplest case, as in Nee et al. (1995), of a set of models where  $(d \lambda(u) / d T)$  is zero for all  $u$  (that is, where the speciation rate does not change with time), and assume  $\mu(u)$  is also constant. In that case,  $r_p(u) := \lambda(u) - \mu(u) + 0 / \lambda(u) = \lambda(u) - \mu(u)$ . Two such models are congruent if and only if  $r_p(u)_{\text{modelA}} == r_p(u)_{\text{modelB}}$  and if  $\lambda_p(u)_{\text{modelA}} == \lambda_p(u)_{\text{modelB}}$ . The first requirement can be satisfied by any two models with the same net diversification rate; for example, take a Yule model (where  $\mu(u)_{\text{modelA}} = 0$  for all  $u$ ) and a non-Yule model that may have the same net diversification rate (say,  $\mu(u)_{\text{modelB}} = K > 0$  for all  $u$ , so  $\lambda(u)_{\text{modelA}} = \lambda(u)_{\text{modelB}} - K$ ). However, at the tip,  $\rho * \lambda(u)_{\text{modelA}} != \rho * \lambda(u)_{\text{modelB}}$ , so these models are not congruent: they will have different likelihoods on the same tree. They may be practically indistinguishable as the parameters of the two models grow similar (i.e., as  $K$  approaches zero) and for small datasets, but mathematically, they are different models.

This is consistent with what we know from Nee et al. (1994), Stadler (\_\_\_\_) and others, including Figure 1: given knowledge of sampling fraction, and the branch lengths of a tree, one can estimate speciation and extinction rate.

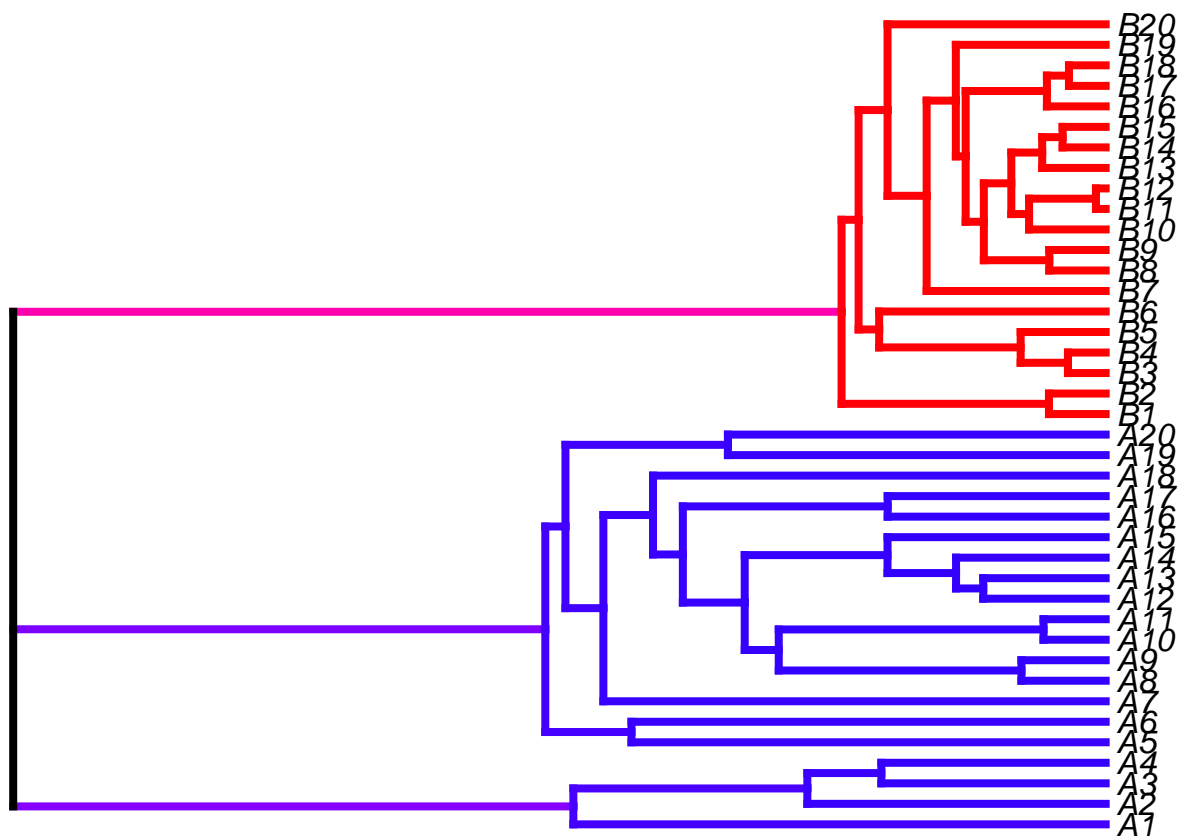
However, that is not all this requirement shows. If a speciation rate through time has a slope of zero but sudden jumps, i.e., a sudden burst of speciation from 65 to 55 MYA, then the pulled diversification rate equals the unpulled diversification rate in that time interval. However, here there can be some congruent, but distinct models. For example, if model\_A and model\_B have the same flat speciation and extinction rates through all intervals but this one, and in that interval model\_A is Yule and model\_B is not Yule, then they could have the same net diversification rate in that interval, still have the same pulled speciation rate at the present, and thus be congruent. Referring back to classic LTT plots may help. Assume for all time intervals but that  $\lambda(u)_{\text{modelA}} = \lambda(u)_{\text{modelB}} = 0.4$  and  $\mu(u)_{\text{modelA}} = \mu(u)_{\text{modelB}} = 0$ . In that time interval, modelA's rates don't change, but modelB has both its  $\lambda$  and  $\mu$  increase by 0.2. The slope of the lineage through time plot in that interval is still 0.4 (0.4 - 0 for modelA, 0.6 - 0.2 for modelB); the slope of the tangent to that plot is still 0.4 for them both at the present. So both models are congruent, fitting the same LTT plot.

Where do SSE models fall into this?

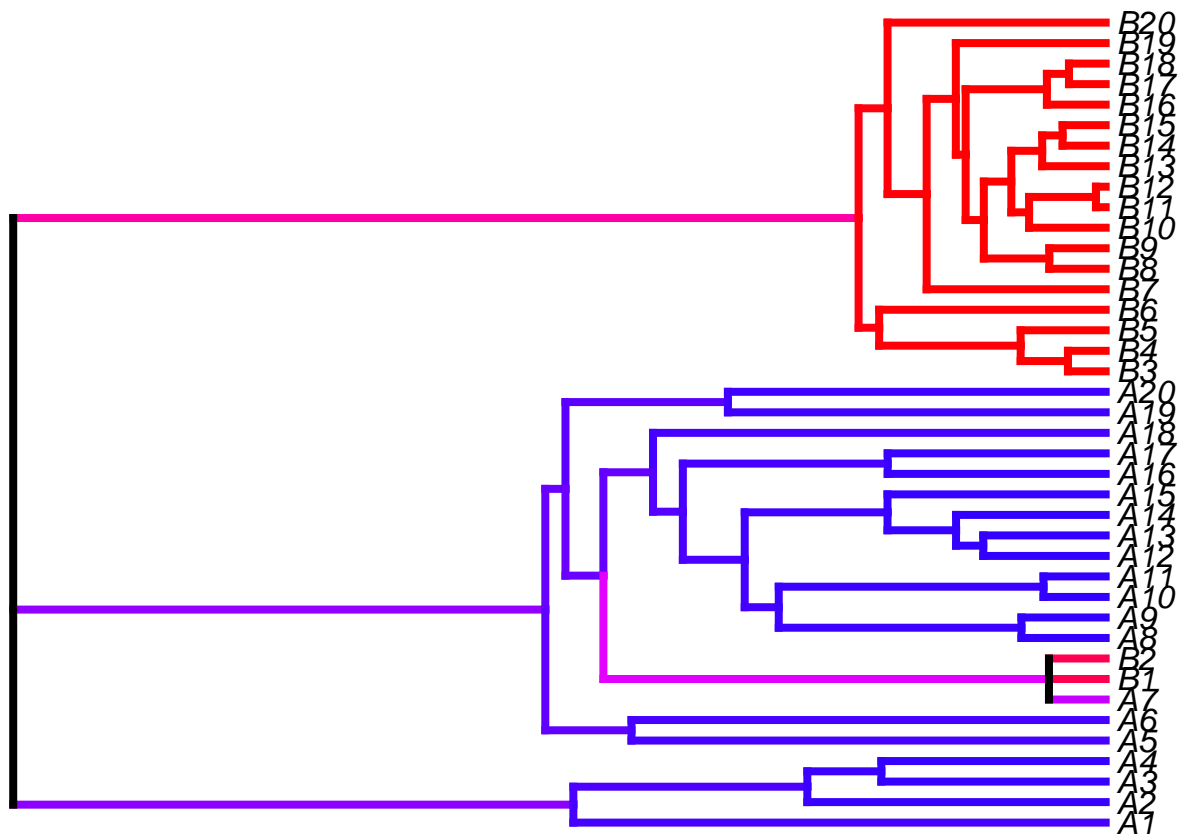
So, for the past quarter century, we should have known that attempting to estimate gradually changing speciation and/or extinction rates over trees leads to identifiability issues. And yet the field has merrily moved on with such methods (\_\_\_\_). How is this? One may be that this paper was largely overlooked: it has only 55 citations, and much citation seems to be regarding this paper as an early demonstration of the difficulty of estimating extinction rates on trees of extant taxa, which is indeed a main point. This paper is less pessimistic than Louca and Pennell (2020) or Pagel (2020), for example, showing that a discrete increase in speciation rate can be detected, perhaps another reason for ignoring its cautions. Another is that scientists have been willing to fix parameters to make models based on lineage through time data tractable, collapsing the model identifiability issue. Many early methods assumed extinction rate was zero (\_\_\_\_), **despite spectacular counterexamples (Irish elk)**. Later methods allowed extinction rate to be higher than zero, but constant over time (\_\_\_\_), or, less commonly, fixed speciation rate as constant but allowed extinction to vary (\_\_\_\_). This has led to a focus on net diversification or speciation since they are easier to estimate and more typically allowed to vary, even though there is ample evidence that extinction rate can vary dramatically and importantly.

However, all the caveats raised by Kubo and Iwasa (1995) and Louca and Pennell (2020) apply to models with at least one of speciation or extinction rates smoothly varying over the tree, with every taxon at a given time point experiencing the same rates. [Note that Kubo and Iwasa (1995), do caution about the practical difficulties of estimating extinction rate as well, even in the constant case]. We do know unchanging speciation and extinction rates are identifiable (\_\_\_\_). **Adding unknown sampling fraction (Stadler)**, makes it possible to estimate only two of speciation, extinction, or sampling fraction, but, of these, biologists may be most willing to use non-phylogenetic external estimates of sampling fraction. So, for a given tree with constant speciation and extinction rate and known sampling fraction, we can estimate these two rates.

What about two trees? If for a single tree we can estimate speciation and extinction rates, surely we can repeat this for a different tree and get its own estimates. But of course, no trees are completely independent: they are subtrees of a larger tree. If we assume these two trees are sister clades, we can estimate one set of rates for the left clade and one for the right clade.

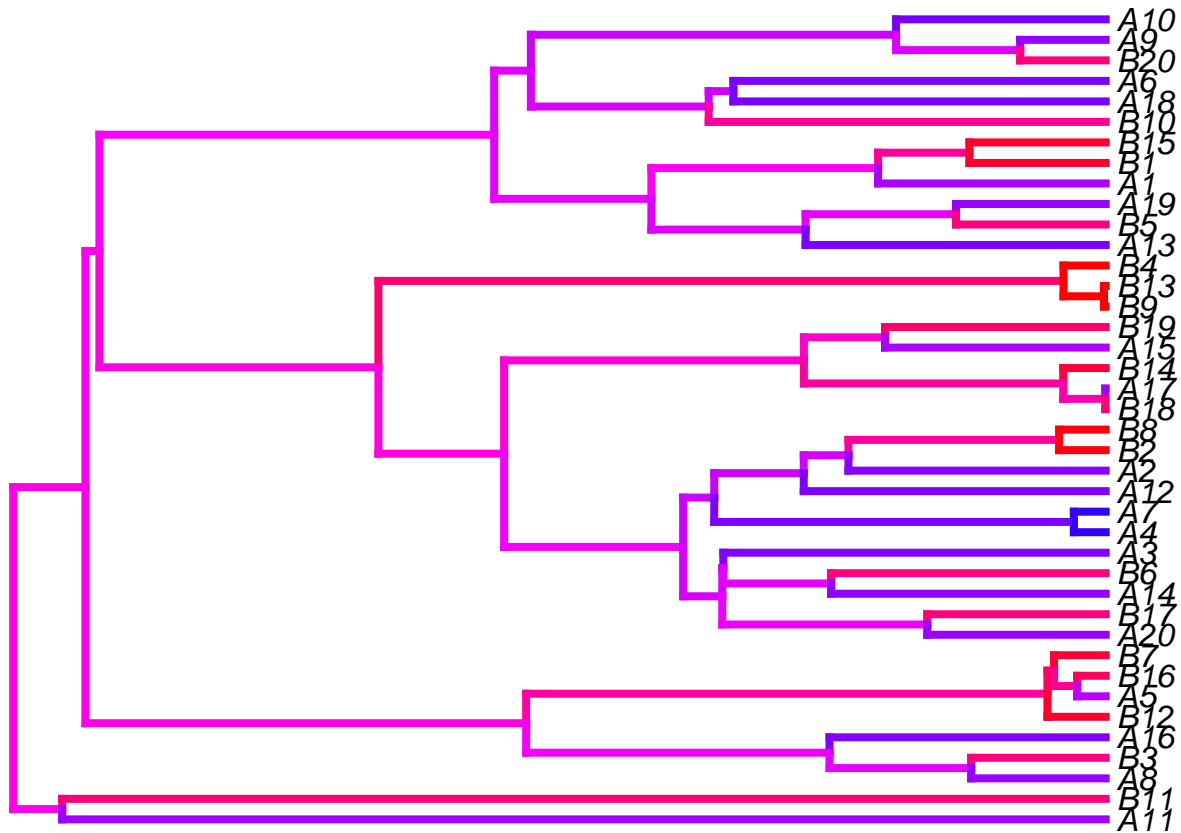


Well then, what if we can take a chunk of one clade and put it on the other clade?



In theory, if we know the mapping, we can still estimate the speciation and extinction rates on the red and blue parts of the tree.

Now imagine a more complex mapping of traits on a tree:



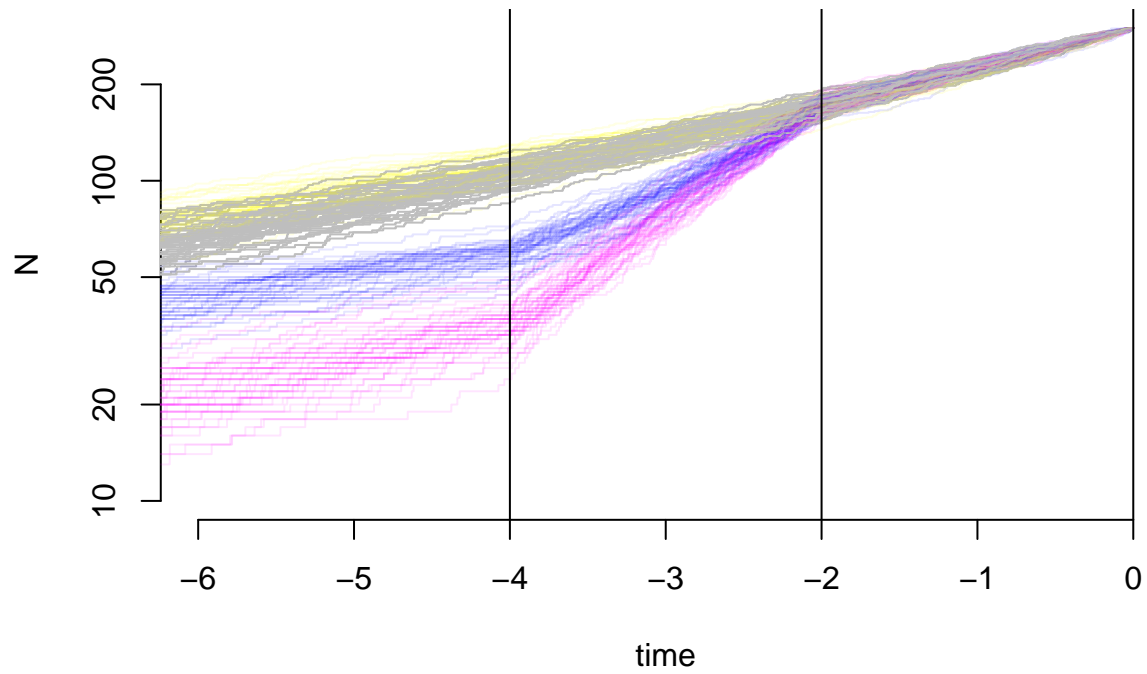
If we have a perfect mapping of which parts are from which tree, we could separately estimate the speciation and extinction rates for each.

This is almost BiSSE (Maddison et al. 2007), but it requires a character model, as well.

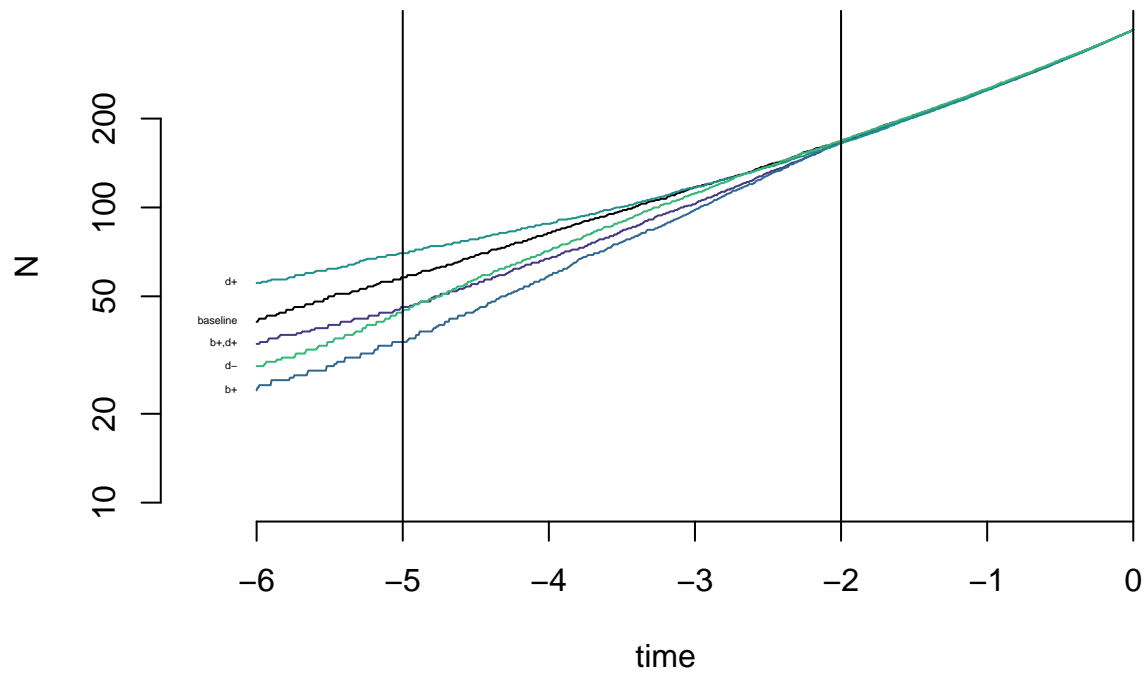
BiSSE cases:

Two clades, one with zero, one with 1 One tree with all tips zero, but some transitions to state 1 internally

```
## Warning in muD[2] <- muA + 0.19: number of items to replace is not a multiple of
## replacement length
```



```
##  lambdaA lambdaB muA muB divA divB
## 1    0.3    0.3 0.1 0.1  0.2  0.2
## 2    0.3    0.9 0.1 0.7  0.2  0.2
## 3    0.3    0.3 0.1 0.1  0.2  0.2
```



```
##  lambdaA lambdaB lambdaC lambdaD lambdaE muA muB muC muD muE
## 1    0.5    0.5    0.5    0.5    0.5 0.2 0.2 0.2 0.2 0.2
## 2    0.5    0.7    0.7    0.5    0.5 0.2 0.4 0.2 0.4 0.0
## 3    0.5    0.5    0.5    0.5    0.5 0.2 0.2 0.2 0.2 0.2
```

## Other stuff

Brownian motion is at the heart of many comparative methods: independent contrasts, ancestral state reconstruction, and more rely on this model. Evolutionary trends are long a topic of great interest in comparative methods: horses getting bigger with fewer toes, evolutionary arms races between predators and prey, and more. It is trivial to add a parameter to Brownian motion models to allow for the mean to evolve along a trend; the likelihood for such models given actual data is finite, and the simple no trend model is even nested within the trend model, so comparisons between a trend and no trend model is very easy. One could do a lot of great biology if you could compare these models on trees of modern taxa, but it's impossible: the likelihood of any trend model on a tree with equal root to top lengths for all taxa is exactly identical. So much as we might want to use this, these models are not identifiable for these kinds of trees.

However, it's a long way from saying these models aren't identifiable to saying any model using Brownian motion is impossible to use on trees of modern taxa. We can compare Brownian motion models with more complex models that are identifiable, such as Ornstein-Uhlenbeck models, Brownian models with more than one rate, models where the rate changes over time, and much more. So yes, Brownian motion with a trend models are unidentifiable on chronograms of modern taxa, but we do not say that any model that attempts to estimate rates of evolution on such trees are impossible. Some models in this space give the same likelihoods and cannot be distinguished, but many others can – this calls for care and analysis, not panic.