

## Chapter 12: Beyond birth-death models

pdf version

### Section 12.1: Capturing variable evolution

There are times and places where the tree of life has grown more rapidly than others. For example, islands and island-like habitats are often described as “incubators” of new species, and diversification rates in such habitats can proceed at an extremely rapid pace. On a broader scale, many studies have shown that speciation rates are elevated and/or extinction rates depressed following mass extinctions. Finally, some clades seem to diversify much more rapidly than others. That is why, for instance, I can see so many birds out the window of this coffee shop, but no tuataras (Figure 12.1). All of these facts lead to the idea that simple, constant-rate birth-death models are not adequate to capture the complexity and dynamics of speciation and extinction across the tree of life. Speciation and extinction rates vary through time, across clades, and among geographic regions. We can sometimes predict this variation based on what we know about the mechanisms that lead to speciation and/or extinction.

Figure 12.1. Birds, like this mountain bluebird, are very diverse in the present day, with about 10,000 species. By contrast, tuataras (right) are very depauperate, with only one or two species (depending on taxonomy) in the present day. However, the stem age of tuataras is very much older than the stem age of birds.

In this chapter, I will explore some extensions to birth-death models that allow us to explore diversification in more detail. This chapter also leads naturally to the next, chapter 13, which will consider the case where diversification rates depend on species’ traits.

Key questions:

1. Do some clades diversify (or go extinct) at a higher rate than others?
2. Do diversification rates vary through time?
3. Are speciation rates diversity-dependent?
4. Do models of protracted speciation explain phylogenetic tree shapes?

### Section 12.2: Variation in diversification rates across clades

We know from analyses of tree balance that the tree of life is more imbalanced than birth-death models predict. We can explore this variation in diversification rates by allowing birth and death models to vary along branches in phylogenetic trees. The simplest scenario is when one has a particular prediction about diversification rates to test. For example, we might wonder if diversification

rates in one clade (clade A in Figure 12.2) is higher than in the rest of the phylogenetic tree. We can test this hypothesis by fitting a multiple-rate birth-death model.

The simplest method to carry out this test is by using model selection in a ML framework. To do this, we first fit a constant-rates birth-death model to the entire tree, and calculate a likelihood. We can then fit variable-rates birth-death models to the data, comparing the fit of these models using either likelihood ratio tests or AICc.

<< Figure 12.2 >>

Consider the example in Figure 12.2. We would like to know whether clade A has speciation and extinction rates,  $\lambda_A$  and  $\mu_A$ , that differ from the background rates,  $\lambda_B$  and  $\mu_B$  – we will call this a “variable rates” model. The alternative is a “constant rates” model where the entire clade has constant rate parameters  $\lambda$  and  $\mu$ . These two models are nested, since the constant-rates model is a special case of the variable rates model where  $\lambda_A = \lambda_B = \lambda$  and  $\mu_A = \mu_B = \mu$ . Calculating the likelihood for these two models is reasonably straightforward, although we have to account for the fact that one section of the tree has been pruned out of the background when calculating its rates.

For a real example, let’s look at the phylogenetic tree of vertebrates and evaluate the hypothesis that birds have distinct diversification rates from other major lineages (mammals, tuataras, squamates, turtles, and crocodiles; Figure 12.3).

<< Figure 12.3 >>

We can calculate the likelihood of the constant rates model, with two parameters  $\lambda$  and  $\mu$ , to a variable rates model with four parameters  $\lambda_{\text{bird}}$ ,  $\mu_{\text{bird}}$ ,  $\lambda_{\text{other}}$ , and  $\mu_{\text{other}}$ . For this example, we obtain the following results.

Model	Parameter estimates	ln-Likelihood	AIC
Constant rates	$\lambda = 0.022$ $\mu = 0.0$	-75.7	155.4
Variable rates	$\lambda_{\text{bird}} = 0.052$ $\mu_{\text{bird}} = 0.018$ $\lambda_{\text{other}} = 0.126$ $\mu_{\text{other}} = 0.102$	-73.5	155.0

We see from these results that there is effectively no difference between the two models, and so no reason (from this small dataset) to conclude that birds are significantly different from other amniotes (but see xxx).

Of course, more elaborate comparisons are possible. For example, one could compare the fit of four models, as follows: Model 1, constant rates; Model 2, speciation rate in clade A differs from the background; Model 3, extinction rate in clade A differs from the background; and Model 4, both speciation and extinction rates in clade A differ from the background. In this case, some of the pairs of models are nested – for example, Model 1 is a special case of Model 2, which is, in turn, a special case of Model 4 – but all four do not make a nested series. Here we benefit from using a model selection approach based on AICc. We can fit all four models and use their relative number of parameters to calculate AICc scores. We can then calculate AICc weights to evaluate the

relative support for each of these four models. (As an aside, it might be difficult to differentiate among these four possibilities without a lot of data!)

But what if you do not have an a priori reason to predict differential diversification rates across clades? Or, what if the only reason you think one clade might have a different diversification rate than another is that it has more species? (Such reasoning is circular, and will wreak havoc with your analyses!) In such cases, we can use methods that allow us to fit general models where diversification rates are allowed to vary across clades in the tree. Available methods are relatively recent and a bit beyond the scope of this book. Newer methods use Bayesian machinery (especially reversible-jump MCMC) to fit multi-rate models of diversification to comparative datasets (e.g. Alfaro et al. 2009, Rabosky et al. xxx, Ane paper).

### Section 12.3: Variation in diversification rates through time

In addition to considering rate variation across clades, we might also wonder whether birth and/or death rates have changed through time. For example, perhaps we think our clade is an adaptive radiation that experienced rapid diversification upon arrival to an island archipelago and slowed as this new adaptive zone got filled. This hypothesis is an example of density-dependent diversification, where diversification rate depends on the number of lineages that are present. Alternatively, perhaps our clade has been experiencing extinction rates that have changed through time, perhaps peaking during some time period of unfavorable climactic conditions. This is another hypothesis that predicts variation in diversification (speciation and extinction) rates through time.

We can fit time-dependent diversification models using likelihood equations that allow arbitrary variation in speciation and/or extinction rates through time. To figure out the likelihood we can first make a simplifying assumption: though diversification rates might change, they are constant across all lineages at any particular time point. In particular, this means that speciation (and/or extinction) rates slow down (or speed up) in exactly the same way across all lineages in an evolving clade. Our assumption also means that we can consider time-slices through the tree rather than individual branches, i.e. we can get all the information that we need to fit these models from lineage through time plots. This type of a model is called “Equal-Rates Markov” in the literature and predicts exactly the same distribution of tree balance statistics as constant-rate birth death models.

For the purposes of this chapter, I will consider only the simplest time-dependent model where diversification rates (speciation, extinction, or both) change linearly through time. In other words, for time-dependent speciation, we can consider the case where:

where  $\lambda_0$  is the initial speciation rate at time zero and  $m$  is the slope of the relationship between speciation rate and time. If  $m > 0$ , then speciation rate

speeds up through time, leading to a tree with an excess of young lineages and a lineage-through-time plot that bends upwards on a ln-scale; when  $m < 0$ , speciation rate slows through time, resulting in a tree with long tip branches and a lineage-through-time plot that bends down towards the present day.

<< Include likelihood equations >>

Likelihood equations for general time-dependent diversification are...

This sort of approach has become very popular, as time-dependent diversification models are consistent with many ecological models of how multi-species clades might evolve through time. For example, adaptive radiation models based on ecological opportunity predict that, as niches are filled and ecological opportunity “used up,” then we should see a declining rate of diversification through time. By contrast, some models predict that species create new opportunities for other species, and thus predict accelerating diversification through time. These are reasonable hypotheses, but there is a statistical challenge: in each case, there is a very different model that predicts the exact same pattern. In the case of decelerating diversification, the predicted pattern of a lineage-through-time plot that bends down towards the present day can also come from a model where lineages accumulate at a constant rate, but then are not fully sampled at the present day. In other words, if we are missing some living species from our phylogenetic tree and we don’t account for that, then we would mistake a constant-rates birth death model for a signal of slowing diversification through time. Methods have been developed that can account for this, either by simulation or analytical equations that account for randomly missing taxa. Some methods can even account for the fact that the missing taxa might be non-random, as missing taxa tend to be either rare or poorly differentiated from their sister lineages (e.g. often younger than expected by chance).

Likewise, a pattern of accelerating differentiation mimics the pattern caused by extinction. A phylogenetic tree with high but constant rates of speciation and extinction is impossible to distinguish from a tree with no extinction and speciation rates that accelerate through time.

Both of the above caveats are certainly worth considering when interpreting the results of tests of diversification from phylogenetic data. In many cases, adding fossil information will allow investigators to reliably distinguish between the stated alternatives, although methods that tie fossils and trees together are still relatively poorly developed. And various methods have been developed that will give ambiguous results when multiple models provide equivalent explanations for the data.

## Section 12.4: Protracted speciation

In all of the diversification models that we have considered so far, speciation happens instantly; one moment we have a single species, and then immediately

true. But this is not biologically plausible. Speciation takes time, as evidenced by the increasing numbers of partially distinct populations that biologists have identified in the natural world. Furthermore, the fact that speciation takes time can have a profound impact on the shapes of phylogenetic trees. Because of this, it is worth considering diversification models that explicitly account for the fact that the process of speciation has a beginning and an end.

The most successful models to tackle this question have been models of protracted speciation (xxx citations). In such models, speciation begins by the formation of an incipient species. This represents a “partial” species; one can imagine, for example, that this is a population that has split off from the main range of the species, but has not yet evolved full reproductive isolation. The incipient species only becomes a “full” species if it survives some time interval,  $\tau$ , that represents the time it takes to evolve full reproductive isolation (Figure xxx).

Because speciation takes time, the main impact of this model is that we predict fewer very young species in our tree – that is, the nodes closest to the tips of the tree are not as young as they would be compared to pure-birth or birth-death models without protracted speciation (Figure xxx). As a result, protracted speciation models produce lineage through time plots that can mimic the properties often attributed to diversity-dependence, even without any interactions among lineages (ref)!

Likelihood approaches are available for this model of protracted speciation. The way they work is that... (xxx)

So far, models of protracted speciation remain mostly in the realm of ecological neutral theory. However, I think models that treat speciation as a process that takes time – rather than something instantaneous – will be an important addition to our macroevolutionary toolbox in the future.

## Section 12.5: Summary

In this chapter I discussed models that go beyond constant rate birth-death models. We can fit models where speciation rate varies across clades or through time (or both; see Rabosky xxx). In some cases, very different models predict the same pattern in phylogenetic trees, warranting some caution until direct fossil data can be incorporated. I also described a model of protracted speciation, where speciation takes some time to complete. This latter model is potentially better connected to microevolutionary models of speciation, and could point towards fruitful directions for the field. We know that simple birth-death models do not capture the richness of speciation and extinction across the tree of life, so these models that range beyond birth and death are critical to the growth of comparative methods.

## References