

Chapter 11: Fitting birth-death models

pdf version

Introduction: Diversity hotspots

The number of species on the Earth remains highly uncertain, but ranges somewhere above 10 million (Mora et al. 2011). As far as we know, all of those species are descended from a single common ancestor that lived some 4.2 billion years ago (Hedges and Kumar 2009). All species on Earth, then, formed by the process of speciation, the process by which one species splits into two (or more) descendants (Coyne and Orr 2004). Imbalance in diversity across the tree of life tells us that speciation is much more common in some lineages than others (Mooers and Heard 1997). Moreover, numerous studies have argued that certain habitats are “hotbeds” of speciation (e.g. Hutter et al. 2017, Miller and Wiens (2017)). For example, the high Andes ecosystem called the Páromo - a peculiar landscape of alien-looking plants and spectacled bears - harbors the highest speciation rates on the planet (Madriñán et al. 2013).



Figure 1: Figure 11.1 Páromo ecosystem, Chingaza Natural National Park, Colombia. Photo taken by the author.

In this chapter we will explore how we can fit birth-death models to data and, in the process, learn about speciation and extinction rates. Birth-death models can be applied to clade ages and diversities or to patterns of branching times in phylogenetic trees. We will explore both maximum likelihood and Bayesian methods to do this.

Key questions:

1. How can we calculate speciation and extinction rates using clade ages and diversities?
2. How can we fit a birth-death model to the pattern of branching times on a phylogenetic tree?

Clade age and diversity

If we know the age of a clade and its current diversity, then we can calculate the net diversification rate for that clade. The simplest way follows Magallón and Sanderson (2001), who give an equation for the estimate of net diversification rate as:

(eq. 11.1)

$$\hat{r} = \frac{\ln(n)}{t_{stem}}$$

where t_{stem} is the stem group age and $r = \lambda - \mu$. Alternatively, one can use t_{crown} , the crown group age:

(eq. 11.2)

$$\hat{r} = \frac{\ln(n) - \ln(2)}{t_{crown}}$$

The two equations differ because at the crown group age one is considering the clade's diversification starting with two lineages rather than one (Figure 11.2).

These two equations give ML estimates for r only if there is no extinction. If there has been extinction in the history of the clade, then these estimates will be biased. Under a scenario with extinction, one can define $\epsilon = \mu/\lambda$ and use the following method-of-moments estimators from Rohatgi (1976, following the notation of Magallon and Sanderson (2001)):

(eq. 11.3)

$$\hat{r} = \frac{\log[n(1 - \epsilon) + \epsilon]}{t_{stem}}$$

for stem age, and

(eq. 11.4)

$$\hat{r} = \frac{\ln\left[\frac{n(1-\epsilon^2)}{2} + 2\epsilon + \frac{(1-\epsilon)\sqrt{n(n\epsilon^2 - 8\epsilon + 2n\epsilon + n)}}{2}\right] - \ln(2)}{t_{crown}}$$

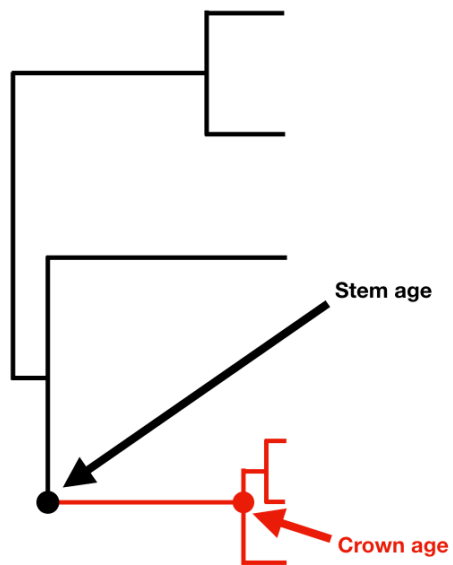


Figure 2: Figure 11.2. Crown and stem age of a clade.

for crown age. (Note that eq. 11.3 and 11.4 reduce to 11.1 and 11.2, respectively, when $\epsilon = 0$). Magallón and Sanderson (2001), following Strathmann and Slatkin (1983), describe how to use eq. 10.13 and 10.15 to calculate confidence intervals for the number of species at a given time.

For example, consider these data, which summarize the ages and diversities of a number of plant lineages in the Páramo (from Madriñán et al. 2013). For each lineage, I have calculated the pure-birth estimate of speciation rate (from equation 11.2, since these are crown ages), and net diversification rates under three scenarios for extinction ($\epsilon = 0.1$, $\epsilon = 0.5$, and $\epsilon = 0.9$).

Lineage	n	Age	\hat{r}_{pb}	$\hat{r}_{\epsilon=0.1}$	$\hat{r}_{\epsilon=0.5}$	$\hat{r}_{\epsilon=0.9}$
Aragoa	17	0.42	5.10	5.08	4.53	2.15
Arcytophyllum	11	10.96	0.18	0.18	0.16	0.07
Berberis	32	3.80	0.73	0.73	0.66	0.36
Calceolaria	67	2.50	1.39	1.39	1.28	0.78
Draba	55	3.05	1.09	1.08	1.00	0.59
Espeletia	20	4.04	1.01	1.01	0.94	0.62
Festuca	36	4.28	0.68	0.67	0.61	0.34
Jameson	32	7.60	0.36	0.36	0.33	0.18
+						
Eriosorus						
Lupinus	66	1.47	2.38	2.37	2.19	1.34
Lysipomus	27	8.96	0.29	0.29	0.26	0.14
Oreobolus	5	3.01	0.30	0.30	0.26	0.09
Puya	46	0.80	3.92	3.91	3.58	2.07
Valeriana	53	14.58	0.22	0.22	0.21	0.12

Table 11.1. Estimates of net diversification rates for Páramo lineages using equations 11.2 and 11.4.

We can also estimate birth and death rates for clade ages and diversities using ML or Bayesian approaches. We already know the full probability distribution for birth-death models starting from any standing diversity $N(0) = n_0$ (see equations 10.13 and 10.15). We can use these equations to calculate the likelihood of any particular combination of N and t (either t_{stem} or t_{crown}) given particular values of λ and μ . We can then find parameter values that maximize that likelihood. Of course, with data from only a single clade, we cannot estimate parameters reliably; in fact, we are trying to estimate two parameters from a single data point, which is a futile endeavor. (It is common, in this case, to assume some level of extinction and calculate net diversification rates based on that, as we did in Table 11.1 above).

One can also assume that a set of clades have the same speciation and extinction rates and fit them simultaneously, estimating ML parameter values. This is the approach taken by Magallón and Sanderson’s (2001) paper on diversification

rates across angiosperms. When we apply this approach to the Paramo data, shown above, we obtain ML estimates of $\hat{r} = 0.27$ and $\hat{\epsilon} = 0$. If we were forced to estimate an overall average rate of speciation for all of these clades, this might be a reasonable estimate. However, the table above also suggests that, perhaps, some of these clades might be diversifying faster than others. We will return to the issue of variation in diversification rates across clades in the next chapter.

Another approach is to use a Bayesian approach to calculate posterior distributions for birth and death rates based on clade ages and diversities. This approach has not, to my knowledge, been implemented in any software package, although the method is straightforward (for similar approaches, see xxx B Moore). To do this, we will modify the basic algorithm for Bayesian MCMC (see Chapter 2) as follows:

1. Sample a set of starting parameter values, r and ϵ , from their prior distributions. For this example, we can set our prior distribution for both parameters as exponential with a mean and variance of $\lambda_p = 1$ (note that this choice might depend on the units you are using, especially for r). Note the potentially confusing notation here - λ_p is the single parameter of the prior distribution, and we will use it as a prior on both speciation and extinction rates. We then select starting r and ϵ from their priors.
2. Given the current parameter values, select new proposed parameter values using the proposal density $Q(p'|p)$. For both parameter values, we can use a uniform proposal density with width w_p , so that $Q(p'|p) \propto U(p - w_p/2, p + w_p/2)$. We can either choose both parameter values simultaneously, or one at a time (the latter is typically more effective).
3. Calculate three ratios:
 - a. The prior odds ratio. This is the ratio of the probability of drawing the parameter values p and p' from the prior. Since we have exponential priors for both parameters, we can calculate this ratio as:

(eq. 11.5)

$$R_{prior} = \frac{\lambda_p e^{-\lambda_p p'}}{\lambda_p e^{-\lambda_p p}} = e^{\lambda_p (p - p')}$$

- b. The proposal density ratio. This is the ratio of probability of proposals going from p to p' and the reverse. We have already declared a symmetrical proposal density, so that $Q(p'|p) = Q(p|p')$ and $R_{proposal} = 1$.
- c. The likelihood ratio. This is the ratio of probabilities of the data given the two different parameter values. We can calculate these probabilities from equations 11.3 or 4 (depending on if the data are stem ages or crown ages). Then we just use equation 2.26.
4. Find R_{accept} as the product of the prior odds, proposal density ratio, and the likelihood ratio. In this case, the proposal density ratio is 1, so:

(eq. 11.6)

$$R_{accept} = R_{prior} \cdot R_{likelihood}$$

5. Draw a random number u from a uniform distribution between 0 and 1. If $u < R_{accept}$, accept the proposed value of both parameters; otherwise reject, and retain the current value of the two parameters.
6. Repeat steps 2-5 a large number of times.

When we apply this technique to the Páromo (from Madriñán et al. 2013), we obtain posterior distributions for both r (mean = 0.497, 95% CI = 0.08-1.77) and ϵ (mean = 0.36, 95% CI = 0.02-0.84; Figure 11.3).

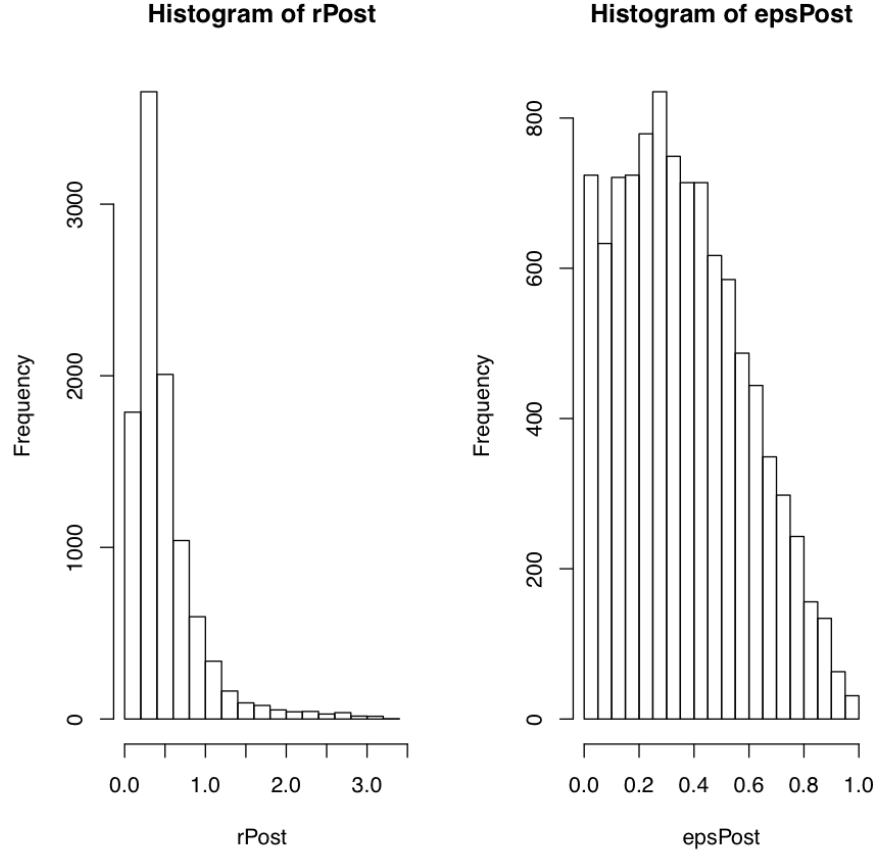


Figure 3: Figure 11.3. Posterior distributions for r and ϵ for Páromo clades (Madriñán et al. 2013).

Thus, we can estimate diversification rates from data on clade ages and diversities. If we have a whole set of such clades, we can (in principal) estimate both speciation and extinction rates, so long as we are willing to assume that all of the clades share equal diversification rates. However, as we will see in the next section, this assumption is almost always dubious!

Tree Balance

We can use the concept of tree balance to evaluate the fit of constant-rate birth-death models to phylogenetic trees. Tree balance is based on comparing the number of species descended from each pair of sister lineages in the tree. For single nodes, we already know that the distribution of sister taxa species richness is uniform over all possible divisions of N_n species into two clades of size N_a and N_b . We can use this fact to derive a simple test of whether the distribution of species between two sister clades is unusual compared to the expectation under a birth-death model. This test can be used, for example, to test whether the diversity of exceptional clades, like passerine birds, is higher than one would expect when compared to their sister clade. This approach traces back to Raup and colleagues, who applied stochastic birth-death models to paleontology in a series of influential papers in the 1970s (e.g. Raup et al. 1973, Raup and Gould (1974)). Slowinsky and Guyer (1993) developed a test based on calculating a P-value for a division at least as extreme as seen in a particular comparisons of sister clades. We consider N_n total species divided into two sister clades of sizes N_a and N_b , where $N_a < N_b$ and $N_a + N_b = N_n$. Then:

(eq. 11.7)

If $N_a \neq N_b$:

$$P = \frac{2N_a}{N_n - 1}$$

If $P > 1$ then set $P = 1$

For example, we can assess diversification in the Andean representatives of the legume genus *Lupinus* (Hughes and Eastwood 2006). In particular, this genus includes a young radiation of 81 Andean species, spanning a wide range of growth forms. The likely sister clade to this spectacular Andean radiation is a clade of *Lupinus* species in Mexico that includes 46 species (Drummond et al. 2012). In this case $N_a = 81 - 46 = 35$, and we can then calculate a P-value testing the null hypothesis that both of these clades have the same diversification rate:

(eq. 11.8)

$$P = \frac{2N_a}{N_n - 1} = \frac{2 \cdot 35}{81 - 1} = 0.875$$

We cannot reject the null hypothesis. Indeed, later work suggests that the actual increase in diversification rate for *Lupinus* occurred deeper in the phylogenetic tree, in the ancestor of a more broadly ranging New World clade (Drummond et al. 2012, Hughes and Eastwood (2006)).

Often, we are interested in testing whether a particular trait - say, dispersal into the Páramo - is responsible for the increase in species richness that we see in some clades. In that case, a single comparison of sister clades may be unsatisfying, as sister clades almost always differ in many characters, beyond just the trait of interest. Even if the clade with our putative “key innovation” is more diverse, we still might not be confident in inferring a correlation from a single observation. To address this problem, many studies have used natural replicates across the tree of life, comparing the species richnesses of many pairs of sister clades that differ in a given trait of interest. Following Slowinsky and Guyer’s logic above, we could calculate a p-value for each clade, and then combine those p-values into an overall test. In this case, one clade (with diversity N_1) has the trait of interest and the other does not (N_0), and our formula is half of equation 11.5 since we will consider this a one-tailed test:

(eq. 11.9)

$$P = \frac{N_0}{N_n - 1}$$

When analyzing replicate clade comparisons - e.g. many sister clades, where in each case one has the trait of interest and the other does not - Slowinsky and Guyer (1993) recommended combining these p-values using Fisher’s combined probability test, so that:

(eq. 11.10)

$$P_{total} = -2 \sum \ln(P_i)$$

Follow-up work showed that this test, though, can be very sensitive to outliers - that is, clades with extreme differences in diversity - and can, in some cases with two characters, show that both characters significantly increase diversity (Vamosi and Vamosi 2005)! Fortunately, there are a number of improved methods that can be used that are similar in spirit to the original Slowinsky and Guyer test but more statistically robust (e.g Paradis 2012).

Finally, we can assess the overall balance of an entire phylogenetic tree using tree balance statistics. There are a relatively large number of such statistics, and different indices capture different aspects of diversification (Mooers and Heard 1997). Since the test statistics are based on descriptions of patterns in trees rather than particular processes, the relationship between imbalance and evolutionary processes can be difficult to untangle! But all tree balance indices

allow one to reject the null hypothesis that the tree was generated under a birth-death model. Actually, the expected patterns of tree balance are absolutely identical under a broader class of models called “Equal-Rates Markov” (ERM) models (Harding 1971, Mooers and Heard (1997)). ERM models specify that diversification rates (both speciation and extinction) are equal across all lineages for any particular point in time. However, those rates may or may not change through time. If they don’t change through time, then we have a constant rate birth-death model, as described above - so birth-death models are ERM models. But ERM models also include, for example, models where birth rates slow through time, or extinction rates increase through time, and so on. All of these models predict exactly the same pattern of tree balance.

Typical steps for using tree balance indices to test the null hypothesis that the tree was generated under an ERM model are as follows:

1. Calculate tree balance using a tree balance statistic.
2. Simulate pure birth trees to generate a null distribution of the test statistic. We are considering the set of ERM models as our null, but since pure-birth is simple and still ERM we can use it to get the correct null distribution.
3. Compare the actual test statistic to the null distribution. If the actual test statistic is in the tails of the null distribution, then your data deviates from an ERM model.

Step 2 is unnecessary in cases where we know null distributions for tree balance statistics analytically (e.g. Blum and François 2006). There are also some examples in the literature of considering null distributions other than ERM. For example, Mooers and Heard (1997) consider two other null models, PDA and EPT, which consider different statistical distributions of tree shapes (but both of these are difficult to tie to any particular evolutionary process).

Typically, phylogenetic trees are more imbalanced than expected under the ERM model. In fact, this is one of the most robust generalizations that one can make about macroevolutionary patterns in phylogenetic trees. This deviation means that diversification rates vary among lineages in the tree of life. We will discuss how to quantify and describe this variation in later chapters. These tests are all similar in that they use multiple non-nested comparisons of species richness in sister clades to calculate a test statistic, which is then compared to a null distribution, usually based on a constant-rates birth-death process (reviewed in Vamosi and Vamosi 2005, Paradis (2012)).

XXX example

Fitting birth-death models to branching times

Another approach that uses more of the information in a phylogenetic tree involves fitting birth-death models to the distribution of branching times in a phylogenetic tree. This approach traces all the way back to Yule (1925), who

first applied stochastic process models to the growth of phylogenetic trees. More recently, a series of papers by Raup and colleagues (Gould et al. 1977, Raup and Gould (1974), Raup (1985), Raup et al. (1973)) spurred modern approaches to quantitative macroevolution by demonstrating how variable clades grown under simple birth-death models can be.

Most modern approaches to fitting birth-death models to phylogenetic trees use the intervals between speciation events on a tree - the “waiting times” between successive speciation - to estimate the parameters of birth-death models. Figure 10.2 shows these waiting times. Frequently, information about the pattern of species accumulation in a phylogenetic tree is summarized by a lineage-through-time (LTT) plot, which is a plot of the number of lineages in a tree against time (see Figure 10.9). As I introduced in Chapter 10, the y-axis of LTT plots is log-transformed, so that the expected pattern under a constant-rate pure-birth model is a straight line. Note also that LTT plots ignore the relative order of speciation events. Stadler (2013b) calls models justifying such an approach “species-exchangable” models - we can change the identity of species at any time point without changing the expected behavior of the model. Because of this, approaches to understanding birth-death models based on branching times are different from - and complementary to - approaches based on tree topology.

As discussed in the previous chapter, even though we often have no information about extinct species in a clade, we can still (in theory) infer the presence of extinction from an LTT plot. The signal of extinction is an excess of young lineages, which is seen as the “pull of the recent” in our LTT plots (Figure 10.10). Statistical approaches can capture this pattern in a more rigorous way.

Likelihood of waiting times under a birth-death model

In order to use ML and Bayesian methods for estimating the parameters of birth-death models from comparative data, we need to write down the likelihoods of the waiting times between speciation events in a tree. There is a little bit of variation in notation in the literature, so I will follow Stadler (Stadler 2013b) to maintain consistency. We will assume that the clade begins at time t_1 with a pair of species. Most analyses condition the process as starting at this time t_1 , the node at the root of the tree, since we rarely have information on the stem age of our clade. We will also condition on both of these initial lineages surviving to the present day, as this is a requirement to obtain a tree like what we have [e.g. Stadler (2013a) equation 5].

Speciation and extinction events occur at various times, and the process ends at time 0 when the clade has n extant species - that is, we measure time backwards from the present day. Extinction will result in species that do not extend all the way to time 0. For now, we will assume that we only have data on extant species. We will refer to the phylogenetic tree that shows branching times leading to the extant species as the reconstructed tree (Nee xxx). For a reconstructed tree

with n species, there are $n - 1$ speciation times, which we will denote as $t_1, t_2, t_3, \dots, t_{n-1}$. The leaves of our ultrametric tree all terminate at time 0.

Note that in this notation, $t_1 > t_2 > \dots > t_{n-1} > 0$, that is, our speciation times are measured from the tips and constantly decreasing (this is an important notational difference between Stadler (2013a) and Nee (1994 and others), who considers the time intervals between speciation events, e.g. $t_1 - t_2$). For now, we will assume complete sampling; that is, all n species alive at the present day are in the tree.

We can now write down the likelihood of observing the set of speciation times t_1, t_2, \dots, t_{n-1} given the extant diversity of the clade, n , and our birth-death model parameters λ and μ . Conditioning on both initial lineages surviving to the present day, we then have:

(eq. 11.11)

$$L(t_i|\lambda, \mu) = \left(\frac{p_1(t_1)}{1 - p_0(t_1)} \right)^2 \prod_{i=2}^{n-1} \lambda p_1(t_i)$$

Where $p_0(t_i)$ and $p_1(t_i)$ are the probabilities of observing 0 and 1 species, respectively, after sampling a birth-death tree of age t , and can be calculated as:

(eq. 11.12)

$$p_0(t) = 1 - \frac{\lambda - \mu}{\lambda - \mu e^{-(\lambda - \mu)t}}$$

(eq. 11.13)

$$p_1(t) = \frac{(\lambda - \mu)^2 e^{-(\lambda - \mu)t}}{(\lambda - \mu e^{-(\lambda - \mu)t})^2}$$

Note that equation 11.12 is algebraically equivalent to the equation for a in equation 10.14.

Using maximum likelihood to fit a birth-death model

Given equation 11.11 for the likelihood, we can estimate birth and death rates using both ML and Bayesian approaches. For the ML estimate, we maximize equation 11.11 over λ and μ . For a pure-birth model, that is when $\mu = 0$, the maximum likelihood estimate of λ can be calculated analytically as:

(eq. 11.14)

$$\lambda = \frac{n-2}{s_{branch}}$$

where s_{branch} is the sum of branch lengths in the tree,

(eq. 11.15)

$$s_{branch} = \sum_{i=1}^{n-1} it_i$$

This is the Kendal-Moran estimator of the speciation rate (Nee 2006).

For a birth-death model, we can use numerical methods to maximize the likelihood over λ and μ .

EXAMPLE

Using Bayesian MCMC to fit a birth-death model

We can also estimate birth and death rates using a Bayesian MCMC. We can use exactly the method spelled out above for clade ages and diversities, but substitute equation xxx for the likelihood, thus using the waiting times derived from a phylogenetic tree to estimate model parameters.

EXAMPLE.

Sampling and birth-death models

It is important to think about sampling when fitting birth-death models to phylogenetic trees. If any species are missing from your phylogenetic tree, they will lead to biased parameter estimates. This is because missing species are disproportionately likely to connect to the tree on short, rather than long, branches. If we randomly sample lineages from a tree, we will end up badly underestimating both speciation and extinction rates (and wrongly inferring slowdowns; see chapter 12).

Fortunately, the mathematics for incomplete sampling of reconstructed phylogenetic trees has also been worked out. We can substitute in equations that include ρ , the proportion of sampled species, for equations 11.12 and 11.13 above:

(eq. 11.15)

$$p_0(t) = 1 - \frac{\rho(\lambda - \mu)}{\rho\lambda + (\lambda(1 - \rho) - \mu)e^{-(\lambda - \mu)t}}$$

(eq. 11.16)

$$p_1(t) = \frac{\rho(\lambda - \mu)^2 e^{-(\lambda - \mu)t}}{(\rho\lambda + (\lambda(1 - \rho) - \mu)e^{-(\lambda - \mu)t})^2}$$

EXAMPLE

Chapter Summary

In this chapter, I described how to estimate parameters from birth-death models using data on species diversity and ages, and how to use patterns of tree balance to test hypotheses about changing birth and death rates. I also described how to calculate the likelihood for birth-death models on trees, which leads directly to both ML and Bayesian methods for estimating birth and death rates. Next, we will explore elaborations on birth-death models, and discuss models that go beyond constant-rates birth-death models to analyze the diversity of life on Earth.

References

- Blum, M. G. B., and O. François. 2006. Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance. *Syst. Biol.* 55:685–691.
- Coyne, J. A., and H. A. Orr. 2004. *Speciation*. Sinauer, New York.
- Drummond, C. S., R. J. Eastwood, S. T. S. Miotto, and C. E. Hughes. 2012. Multiple continental radiations and correlates of diversification in lupinus (leguminosae): Testing for key innovation with incomplete taxon sampling. *Syst. Biol.* 61:443–460. academic.oup.com.
- Gould, S. J., D. M. Raup, J. J. Sepkoski, T. J. M. Schopf, and D. S. Simberloff. 1977. The shape of evolution: A comparison of real and random clades. *Paleobiology* 3:23–40. Cambridge University Press.
- Harding, E. F. 1971. The probabilities of rooted tree-shapes generated by random bifurcation. *Adv. Appl. Probab.* 3:44–77. Cambridge University Press.
- Hedges, B. S., and S. Kumar. 2009. *The timetree of life*. Oxford University Press, Oxford.
- Hughes, C., and R. Eastwood. 2006. Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the andes. *Proc. Natl. Acad. Sci. U. S. A.* 103:10334–10339.
- Hutter, C. R., S. M. Lambert, and J. J. Wiens. 2017. Rapid diversification and time explain amphibian richness at different scales in the tropical andes, earth’s

- most biodiverse hotspot. *Am. Nat.* 190:828–843.
- Madriñán, S., A. J. Cortés, and J. E. Richardson. 2013. Páramo is the world’s fastest evolving and coolest biodiversity hotspot. *Front. Genet.* 4:192.
- Magallon, S., and M. J. Sanderson. 2001. ABSOLUTE DIVERSIFICATION RATES IN ANGIOSPERM CLADES. *Evolution* 55:1762–1780.
- Miller, E. C., and J. J. Wiens. 2017. Extinction and time help drive the marine-terrestrial biodiversity gradient: Is the ocean a deathtrap? *Ecol. Lett.* 20:911–921.
- Mooers, A. O., and S. B. Heard. 1997. Inferring evolutionary process from phylogenetic tree shape. *Q. Rev. Biol.* 72:31–54.
- Mora, C., D. P. Tittensor, S. Adl, A. G. B. Simpson, and B. Worm. 2011. How many species are there on earth and in the ocean? *PLoS Biol.* 9:e1001127. Public Library of Science.
- Nee, S. 2006. Birth-Death models in macroevolution. *Annu. Rev. Ecol. Evol. Syst.* 37:1–17. Annual Reviews.
- Nee, S., R. M. May, and P. H. Harvey. 1994. The reconstructed evolutionary process. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 344:305–311.
- Paradis, E. 2012. SHIFT IN DIVERSIFICATION IN SISTER-CLADE COMPARISONS: A MORE POWERFUL TEST. *Evolution* 66:288–295. Wiley Online Library.
- Raup, D. M. 1985. Mathematical models of cladogenesis. *Paleobiology* 11:42–52.
- Raup, D. M., and S. J. Gould. 1974. Stochastic simulation and evolution of Morphology-Towards a nomothetic paleontology. *Syst. Biol.* 23:305–322. Oxford University Press.
- Raup, D. M., S. J. Gould, T. J. M. Schopf, and D. S. Simberloff. 1973. Stochastic models of phylogeny and the evolution of diversity. *J. Geol.* 81:525–542.
- Rohatgi, V. K. 1976. An introduction to probability theory and mathematical statistics. Wiley.
- Slowinski, J. B., and C. Guyer. 1993. Testing whether certain traits have caused amplified diversification: An improved method based on a model of random speciation and extinction. *Am. Nat.* 142:1019–1024.
- Stadler, T. 2013a. How can we improve accuracy of macroevolutionary rate estimates? *Syst. Biol.* 62:321–329.
- Stadler, T. 2013b. Recovering speciation and extinction dynamics based on phylogenies. *J. Evol. Biol.* 26:1203–1219.
- Strathmann, R. R., and M. Slatkin. 1983. The improbability of animal phyla

with few species. *Paleobiology*. cambridge.org.

Vamosi, S. M., and J. C. Vamosi. 2005. Endless tests: Guidelines for analysing non-nested sister-group comparisons. *Evol. Ecol. Res.* 7:567–579. Evolutionary Ecology, Ltd.

Yule, G. U., and Others. 1925. II.—A mathematical theory of evolution, based on the conclusions of dr. JC willis, FR S. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 213:21–87. The Royal Society.