

Birth-Death Models in Macroevolution

Sean Nee

Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, West Mains Road, Edinburgh, EH9 3JT, United Kingdom;
email: sean.nee@ed.ac.uk

Annu. Rev. Ecol. Evol. Syst. 2006. 37:1–17

First published online as a Review in Advance on September 1, 2006

The *Annual Review of Ecology, Evolution, and Systematics* is online at
<http://ecolsys.annualreviews.org>

This article's doi:
10.1146/annurev.ecolsys.37.091305.110035

Copyright © 2006 by Annual Reviews.
All rights reserved

1543-592X/06/1201-0001\$20.00

Key Words

fossil record, paleobiology, phylogenetics

Abstract

Birth-death models, and their subsets—the pure birth and pure death models—have a long history of use for informing thinking about macroevolutionary patterns. Here we illustrate with examples the wide range of questions they have been used to address, including estimating and comparing rates of diversification of clades, investigating the “shapes” of clades, and some rather surprising uses such as estimating speciation rates from data that are not resolved below the level of the genus. The raw data for inference can be the fossil record or the molecular phylogeny of a clade, and we explore the similarities and differences in the behavior of the birth-death models when applied to these different forms of data.

INTRODUCTION

Mathematical modeling of the dynamical process of speciation and extinction—more generally “birth-death” models—can be used to address many questions in macroevolution. For example, a simple mathematical model of the growth of a clade through speciation can be used to ask the straightforward question, What is the rate of speciation in the clade? This may be asked, for example, to discover if there was anything unusual about the rate of speciation during the Cambrian explosion (Lieberman 2001), the explosion referring, of course, to body plans. Birth-death models can also be used to address less obvious questions in macroevolution. How bad would the fossil record have to be to explain the discrepancy between the fossil ages of major radiations and their molecular ages (Foote et al. 1999)? For another example, species numbers were very low at both the start of the Cambrian and the period after the Late Permian mass extinction. In the subsequent radiations the number of higher taxa appearing was much higher in the former than the latter—why? (Nee & May 1997) Is it “strange” to have so many animal phyla with a very small number of species (Strathman & Slatkin 1983)?

The modern period of the use of birth-death processes to inform thinking about macroevolution began with the work of Raup and colleagues in 1973 (Raup et al. 1973) when computers were just starting to become a readily used tool. They modelled a scenario in which the species in a clade had equal probabilities of either speciating to create a second species or going extinct, and the overall clade size was kept roughly constant. They discovered that this purely random process could produce apparent trends and patterns resembling those in the fossil record, such as “adaptive radiations.” As Maynard Smith put it (Maynard Smith 1989), they showed that “...it is fatally easy to read a pattern into stochastically generated data.” This is obviously an important realization and familiarity with the “patterns” that random processes create is an essential piece of a scientist’s mental furniture.

We can extend the use of such models back further to 1924 to the work of the statistician Yule (1924) who modelled a clade growing according to the pure birth process in which extinction does not occur (see below). The Willis in his paper’s title (“A mathematical theory of evolution based on the conclusions of Dr. J.C. Willis, FRS”) was an expert on angiosperms who had formed what were perceived to be anti-Darwinian views on the distributions of numbers of subtaxa per taxon, e.g., the numbers of species in genera. The paper illustrates the sorts of taxon sizes that are generated by the random birth process—and, so, shows us the kind of patterns that random processes can generate. However, Yule went further and used his model for the second important use to which such models are put—parameter estimation. He used the birth process to estimate the rate of cladogenesis of the angiosperms. We will see more contemporary examples below.

Branching processes (Gilinsky & Good 1991) and purely statistical models of random walks (e.g., Cornette & Lieberman 2004)—as opposed to models with explicit births and deaths—are also models that have informed macroevolutionary thinking, although their mathematical structure differs from that of the models to be discussed here. But they are little used compared to birth-death models in the technical sense and will not be discussed here. Before leaving them entirely, we will note one

interesting result. Apart from the past 75 Myr, the diversity fluctuations seen in Sepkoski's celebrated 540 Myr fossil compendium (Sepkoski 1982) cannot be distinguished from a random walk (Cornette & Lieberman 2004). The past 75 Myr have seen diversity climb higher than random expectations and this is probably not an artifact of a better fossil record, the so-called pull of the recent (Jablonski et al. 2003).

Phylogenetic information is central to inferences about macroevolution as we are interested in the fates of *clades*, i.e., groups of related organisms. After the work of Raup and his colleagues, the second wave of interest in birth-death models and macroevolutionary inference was stimulated by the increasing availability of molecular phylogenies (Hey 1992, Nee et al. 1992). Fossil family trees describe the times of appearance of taxa as well as the genealogical relationships among them. So, too, molecular phylogenies also come equipped with a temporal dimension provided by the molecular "clock." Increasingly sophisticated statistical advances allow the dating (possibly only relative) of the nodes in a molecular phylogeny even in the absence of a universal metronomic molecular time piece (e.g., Aris-Brosou & Yang 2003, Sanderson 1997). Of course, dated molecular phylogenies can be the subject of controversy (e.g., Aris-Brosou & Yang 2003, Benton & Ayala 2003), but such controversy is premised on their existence. In this second phase of interest in birth-death models the emphasis has been primarily on using these models for statistical inference rather than familiarizing us with the sort of patterns that can be generated by random processes.

We will now methodically go through the various kinds of birth-death models that have been used and see the sorts of macroevolutionary inferences that have been drawn from them as well as discuss their general properties. The mathematics of these models will not be presented—this has been done many times before. An important source of original results for the general birth-death process with varying birth and death rates is Kendall (1948). Useful summaries of various results, as well as a few new results, can be found in Raup (1985), Foote (2001), and Nee et al. (1994). Statistical analysis of macroevolutionary data is not addressed here because such analyses are constantly evolving and frequently need to be tailored for the idiosyncrasies of particular data sets. The studies discussed in this review are the best guides to how to perform such analyses.

THE PURE BIRTH AND PURE DEATH MODELS

The Pure Birth Model

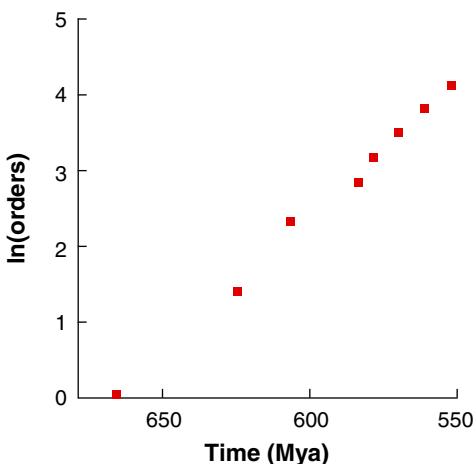
In this model each species has a constant probability, b , of producing a new species at each point in time and extinction never occurs. In this case, starting with $N(0)$ species at the start of the process, the number of species in the clade, $N(t)$, is expected to grow exponentially over time t :

$$N(t) = N(0)e^{bt}. \quad 1.$$

Of course, this model may be applied to other entities in addition to species—higher taxa for example. It immediately follows from the exponential growth that a

Figure 1

Cumulative increase in the natural logarithm of the number of metazoan orders over the Vendian and Lower Cambrian. The use of the natural logarithm allows one to assess the fit of the birth model from the linearity of the plot and estimate the rate of diversification from the slope of the points. Data from Sepkoski (1978).



plot of the logarithm of the number of species against time should be linear and the slope of the plot provides an estimate of b , the per-capita birth rate, as long as it is understood that this particular calculation assumes natural logarithms. Equation 1 can be interpreted as a purely deterministic model of exponential growth, as it was in Stanley (1975) and Sepkoski (1978). It is also the expected number of species of the stochastic model (e.g., Nee 2001).

Figure 1 shows the increase in the number of metazoan orders as we cross into the Cambrian (Sepkoski 1978) and shows that, for this data set at least, ordinal diversification can be well-described by a model of exponential growth. The rate of growth, b , can be roughly computed from the Figure as 0.033/Myr, although Sepkoski was more interested in the constancy of the diversification rate rather than its actual value.

Figure 2 illustrates the same sort of analysis carried out on a molecular phylogeny (Baldwin & Sanderson 1998). The phylogeny is of the plants in the Hawaiian silversword alliance, a group with remarkable morphological diversity (trees, shrubs, vines, etc.) and habitat heterogeneity. Although there is an apparent slight curvature suggesting a slight increase in diversification rate over time, this is not so dramatic as to obviate the use of the pure birth model to estimate speciation rates as $0.56 \pm 0.17/\text{Myr}$. This is a remarkably high rate, exceeding that of continental radiations. For example, the Old World monkeys had the highest speciation rate of the four main groups of primate—0.342/Myr—the other groups being New World monkeys, Madagascar primates, and apes (Purvis et al. 1995). The latter three groups had similar rates of diversification that were statistically significantly smaller than that of the Old World monkeys. Based on fossil data, Alroy reports that origination rates of mammals during the Paleocene averaged 0.43/Myr, but this was accompanied by high extinction rates resulting in a rather slower net accumulation rate (Alroy 1999). To make these numbers meaningful consider that a clade diversifying at a rate of 0.5/Myr would grow to ≈ 150 after 10 Myr and $\approx 270,000$ after 25 Myr.

Paleontologists will be aware that Sepkoski and others did not imagine they were simply estimating origination rates. In fact, they were interested in net rates, r , of

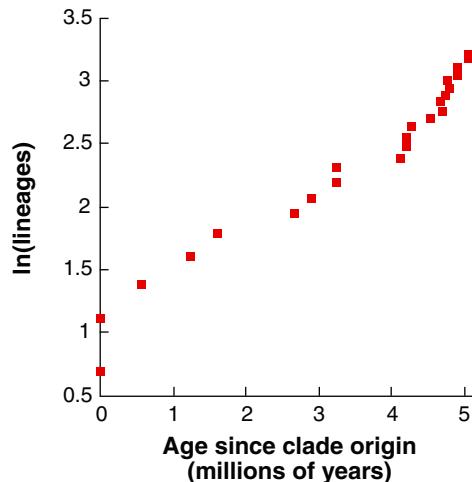


Figure 2

Cumulative increase in the logarithm of the number of lineages in a dated molecular phylogeny of the Hawaiian silversword alliance. Data from Baldwin & Sanderson (1998).

diversification, origination minus extinction, $b-d$, so the b in Equation 1 should be interpreted as this rate r for paleontological studies. We have included macroevolutionary paleontological inference in this section simply because of the interest in a single parameter.

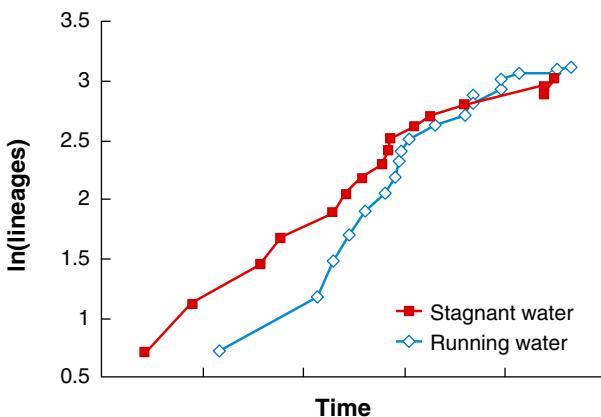
However, it is not possible to have this more general interpretation of the parameter when dealing with molecular phylogenies. As discussed in the next section, introducing the death rate d qualitatively changes the behavior of the model and the apparent growth of the clade through time is no longer log-linear. However, although it is unrealistic to assume that no extinction occurs, this is not actually relevant. If, as in the case of the silversword alliance diversification discussed above, the model is appropriate for the data, and the rough log-linearity of **Figure 2** says that it is, then it is appropriate for estimation of a diversification rate even if, technically, this is purely a speciation rate.

So far, we have seen the pure birth model used to make inferences about the rate of diversification and its constancy, as evidenced by log-linearity. Other questions can be addressed. For example, Ribera and colleagues used molecular phylogenies to investigate whether there exist differences in diversification rates of aquatic beetle clades that inhabit running versus stagnant water bodies (Ribera et al. 2001). As can be seen from the roughly parallel slopes of the two clades in **Figure 3** there is no substantial difference. A comparison based on fossil data suggests balanoid barnacles diversified more rapidly than molluscs (Stanley & Newman 1980). (Of course, the next step is to investigate possible causes of such differences—or lack thereof—but that is beyond the scope of this review.)

Many authors have suggested that speciation rates may have increased during the Pleistocene (roughly two million to 10,000 years ago) because of recurring ice

Figure 3

Cumulative increase in the logarithm of the number of lineages in a molecular phylogeny of two species of aquatic beetle, one living in stagnant water and the other in running water. The time axis is in units of genetic distance and is not labeled because it is not calibrated. Data from Ribera et al. (2001).



ages (e.g., Hewitt 1999). Though a molecular phylogenetic study of birds (Zink & Slowinski 1995) actually suggested that rates declined, another such study of North American tiger beetles (Barraclough & Vogler 2002) found some, albeit weak, evidence for a Pleistocene increase. Much stronger molecular evidence for an increase is found for Holarctic damselflies, which also display a geographical signal: a more northerly clade experienced a 24-fold increase in the rate of cladogenesis compared to a sixfold increase in a more southerly clade (Turgeon et al. 2005).

In fact, the primate analyses of Purvis et al. (1995) and the damselfly analysis of Turgeon et al. (2005) used a birth-death model (see below) but the best estimate for the death rate was zero, so they are discussed in this section. A best estimate of $d = 0$ is frequently the case (e.g., Losos & Schlüter 2000). It is unclear why this is the case.

The Pure Death Model

The pure death model (Raup 1975, van Valen 1973) is only useful for paleontological data. Suppose we have data on the longevities of a number of species. On the assumption that at any point in time each species has the same probability of going extinct, d , then we expect an exponential decay in the number of species that survive to age t , $N(t)$:

$$N(t) = N(0)e^{-dt}.$$

A semilog plot of the number of species surviving to age t is the analogue of our previous semilog plots and, once again, the linearity of the plot informs us of the adequacy of the model and the slope of the plot informs us of the extinction rate d .

On the basis of such analyses, van Valen proposed an evolutionary law (Van Valen 1973). In the words of Raup (Raup 1975, p. 82), "...within an ecologically homogeneous taxonomic group, extinction occurs at a stochastically constant rate." To explain this law, van Valen proposed a Red Queen view of evolution in which evolutionary advances in one species are seen as environmental deterioration by others, leading to perpetual coevolution and, somehow, constant rates of extinction. However

dubious the merits of the model (Maynard Smith 1989), there remains the empirical question of the constancy of extinction rates. To not observe such constancy would be surprising—and interesting—as it would imply either that species senesce, or the converse, that the longer a species survives, the more likely it is to survive even longer. I think we would expect the future lifetime of a species to be independent of how long it has been around already, and that is what the pure death process postulates.

THE BIRTH-DEATH MODEL

We now introduce extinction by allowing species to not only speciate, with constant rate b , but to also die, with constant rate d . This introduces no radically new features for the interpretation of paleontological data: Clades are expected to grow exponentially at a rate $b-d$, the net rate of diversification, and this is the interpretation that paleontologists have always put on the single parameter of the model. But adding extinction has a dramatic impact on the interpretation of molecular phylogenies.

We saw in the previous section that the semilog representation of the growth of a molecular phylogeny is expected to be linear under a pure birth process. **Figure 4** illustrates that this is no longer the case when the extinction rate is nonzero. In fact, over much of the history of the clade the plot is expected to be linear with a slope of $b-d$, just as it would be with paleontological data, but as we approach the present the slope is expected to increase, asymptotically approaching b , thus creating the illusion of an accelerating rate of cladogenesis. This happens because the molecular phylogeny is based entirely on data from extant species, and species that have originated more recently in the past have had less time to go extinct. One consequence of this behavior is that a clade that we might naively think to be in rude health, with an accelerating rate of diversification, is, in fact, subject to high rates of extinction. This is discussed in the context of the phylogeny of plethodontid salamanders in Nee et al. (1995).

Intuition, based on this behavior of the lineage-through-time plot, suggests that we have two items of information about the growth of the clade (b , $b-d$) and this

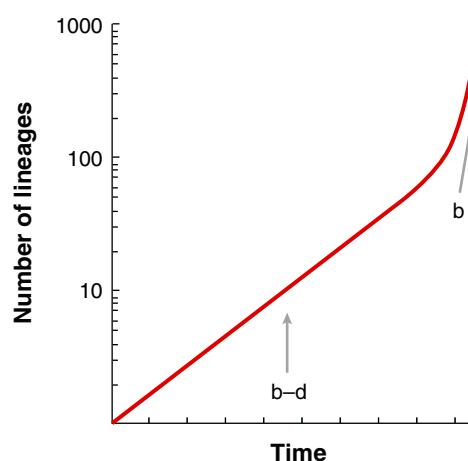


Figure 4

Theoretical plot showing the expected cumulative increase in the logarithm of the number of lineages in a molecular phylogeny growing according to a birth-death process. Theory from Nee et al. (1994).

suggests the possibility that we may be able to estimate speciation and extinction rates separately. This is, in fact, the case (Nee et al. 1994).

The stochastic theory is such that it is more natural to consider the composite parameters $b-d$ and d/b . The first is, of course, the net rate of diversification and the second controls the degree to which the growth of the clade departs in its behavior from that of a pure birth process—visually, the larger d/b the greater degree of curvature in **Figure 4**. A crucial probability relevant for both paleontological and molecular phylogenetic data is the probability that a lineage that arose at some time in the past still exists at time t later, i.e., it has at least one progeny. This is given by:

$$\Pr(t) = \frac{1 - \frac{d}{b}}{1 - \frac{d}{b} e^{-(b-d)t}}. \quad 2.$$

Probability Equation 2 is central because a lineage will not appear in a molecular phylogeny unless it has at least one extant descendant. The importance of the probability in paleontology is that it forms the basis of cohort analysis of, say, the survival of genera through time, which can be used to estimate origination and extinction rates of the species (Raup 1975, 1978; Foote 1988, 2001). Foote's analysis returned a high speciation rate for Cambrian trilobites of 0.4/Myr.

So, we have two surprises from the birth-death model. First, we can estimate speciation and extinction rates from molecular phylogenies even though they do not contain information from extinct species. Second, we can estimate per-species speciation and extinction rates even from fossil data that are not resolved to a level below that of, say, the genus.

Van Valen's empirical data in support of his law included data on the survivorship of taxa higher than species. Raup (1975) made the observation that semilog survivorship plots are not expected to be linear under a birth-death model if the taxa are higher than species: $\log N(t)$, $N(t)$ from Equation 2 as $N(0)$ $\Pr t$ is no longer a linear function of time. This underlies the nonlinearity of **Figure 4** as well.

Equation 2 forms the basis of an interesting argument that life arose repeatedly on Earth, even though all current life is descended from a single ancestor (Raup & Valentine 1983). I invite the reader to consult this admirably speculative paper for details.

Although it is possible in principle to estimate both composite parameters (and so, if you wish, recover b and d), in fact our estimates of $b-d$ are generally much more precise than that of d/b . This is not obvious from Equation 2. This lack of obviousness is good, because it is not necessarily true! If, for example, we had as our data the fact that after a time t one genus was extinct and one other extant, then a maximum likelihood analysis will return an estimate of $d/b = 0.5$ and any number $b-d$ from very small to infinity so the probability in Equation 2 becomes 1/2. As soon as the exponential term in Equation 2 gets small then any value of $b-d$ will generate an estimated probability of 0.5, which is what it should be on the basis of these data. It is only as we accumulate more data stretched over time that the parameter $b-d$ becomes of crucial importance in fitting our estimated probability to the data. The

imprecision of our estimates of d/b compared to $b-d$ was noted in paleontology by Foote (1988) in a study of trilobite genera survivorship and by Nee et al. in a study of *Drosophila* (Nee et al. 1995). A related fact is that although it is possible in principle to detect mass extinctions in molecular phylogenies, which is remarkable, the signal is expected to be weak in the kinds of data sets we can realistically expect to observe (Harvey et al. 1994).

THE MORAN PROCESS

The previous models are suitable for radiations that have not hit any limits to diversity. At the other extreme, we require a model for clades that have reached a plateau, as, for example, metazoan orders appear to have done in the Ordovician and Silurian and remained at the plateau for about 200 million years (Sepkoski 1978). A useful model for this was first introduced in population genetics by Moran (1958): at each point in time, each lineage (or higher taxon) has a probability of going extinct and when a lineage does go extinct it is replaced by the progeny of another lineage chosen at random. So in this model the clade is kept at a constant size deterministically. There are many analytical results available for this process in the population genetics literature, particularly that branch known as coalescence theory (e.g., Hudson 1990), and these tell us what a molecular phylogeny of the extant members of a clade that has arisen from this process should look like. As might be, to some extent, anticipated from the behavior of the birth-death process, the Moran process generates a clade with an apparently accelerating rate of cladogenesis over its entire history.

There is no substantial difference between the Moran process and one of the simulation algorithms studied by, among others, Raup et al. (1973) and Gould et al. (1977), nor between this and the algorithm employed by Sepkoski (1978) and Sepkoski & Kendrick (1993), once the plateau has been reached for some time. (In Sepkoski 1978 the clade grows logistically to the plateau.) In the simulations of Raup et al. (1973) and Gould et al. (1977), they decided on a ceiling diversity and set $b = d$ at this ceiling. When the diversity dropped below the ceiling, they set $b > d$ to get it back there: similarly, they set $b < d$ when diversity rises above.

The Moran process was introduced into the study of molecular trees by Hey (1992), although not identified as such. In an analysis of eight phylogenetic trees (of salamanders, for example), Hey found that a pure birth process model ($d=0$) provided a statistically superior description of the data than the Moran model ($d=b$) and argued on this basis for the merits of the pure birth process model as a model of macroevolution. However, Nee (2001) argues on methodological grounds that the Moran process should remain an integral component of our arsenal.

Random Walk

There is a very big difference between the Moran process and a birth-death process in which $b=d$ and these rates are kept constant. In this model, the total size of the clade follows a random walk. This model has been studied extensively in the context of fossil data with particular reference to the shape of clades (Gould et al. 1977, Kitchell

& MacLeod 1988, Uhen 1996), and not at all, to my knowledge, with reference to molecular data. This is understandable when we realize that extinction is inevitable for clades following this random walk: In the fossil record, clades come and go but in molecular trees of extant species, clades only come.

THE SHAPES OF CLADES

Paleontologists have long been interested in the shapes of clades, i.e., their diversity paths as they first appear in the fossil record, reach their peak diversity level, and ultimately dwindle to extinction (see **Figure 5**). So, for example, are clades symmetrical, with their rise to peak diversity being the mirror image of their decline to extinction? This is what would be expected from a birth-death process with $b = d$: A random walk is symmetrical with respect to time. As mentioned in the Introduction, the ability of birth-death simulation processes to produce clades that look like actual ones as they appear in the fossil record was one of the first surprising results in the modern period of the use of these models for macroevolutionary inference (Gould et al. 1977, Raup et al. 1973, Stanley et al. 1981).

A remarkable result was obtained by Gould et al. (1987). In a study of eight major groups of marine invertebrates they found that clades arising early in time (Cambrian and Ordovician) tended to be bottom-heavy (as in **Figure 5**), whereas those arising in later periods tended, on average, to be symmetrical. This means that the temporal direction of diversification has a signal and that diversification is asymmetrical with respect to time, unlike the simple random walk model. They developed macroevolutionary hypotheses for this phenomenon in terms of adaptive radiations into empty niche space.

A simulation study of birth-death processes casted doubt on this result by noting that the average bottom-heaviness Gould et al. observed in the early period—as measured by a center of gravity, CG, a statistic they had devised—was quite likely

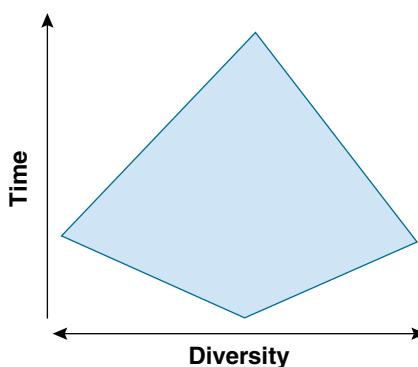


Figure 5

A spindle diagram showing the increase and decline of a clade over time. The width of the diagram at any point in time indicates its diversity at that time. The particular clade shown here is bottom heavy.

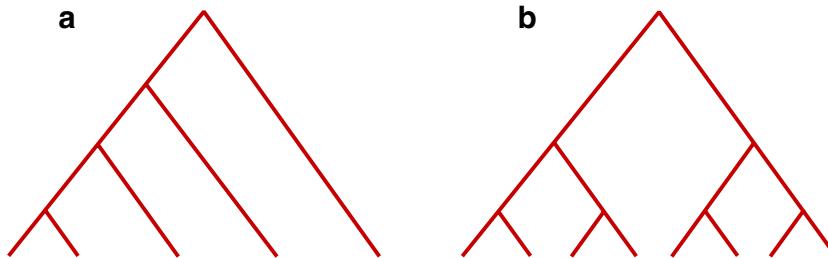


Figure 6

Two phylogenetic topologies. Typology *a* is completely unbalanced and typology *b* is completely balanced.

to arise by chance alone (Kitchell & MacLeod 1988). This critique is not entirely definitive, however. It is true that the original paper did not do a statistical analysis of the CG difference between the early and later periods—an omission remedied by this paper. But the paper also included a regression analysis of the CG of mammalian genera as a function of time since the base of the Tertiary (≈ 65 Mya) and observed a highly significant tendency for clades to become more top-heavy over time.

Alas, this result seems extremely sensitive to the data used. In a reanalysis using an updated data set, exactly the opposite pattern is found (Uhen 1996). Uhen concluded that it is pointless to generate macroevolutionary speculations for observations that are clearly highly volatile as new data are accumulated. It is a shame that Gould's arrow of time joins the 26-million-year extinction periodicity and its associated Death Star (Muller 1988) in the bin of exciting paleontological might-have-beens.

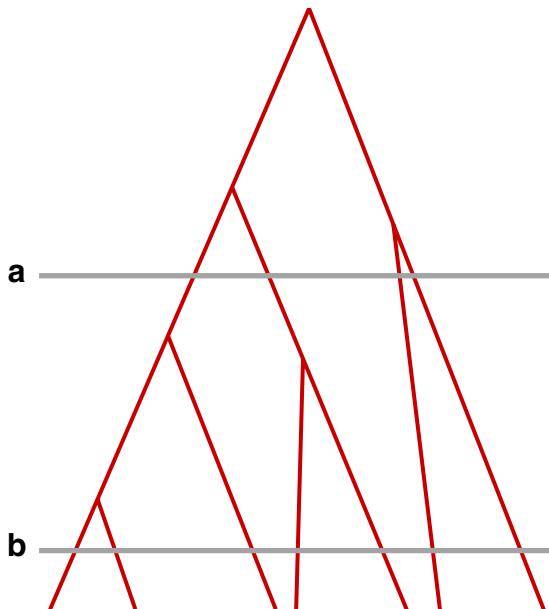
The analogue of the analysis of clade shape in the study of trees of extant species is the investigation of tree balance (see **Figure 6**). A variety of statistics and null models for studying tree balance have been suggested (Moors & Heard 1997). There is really only one null model that is meaningful in a macroevolutionary context, however. This is commonly referred to as the “equal rates Markov” (ERM) model. Thus far we have not discussed tree topology. Under the pure birth process, each lineage is as likely to give rise to a new lineage as any other, so if we are tracking topology as well as internode intervals in the growing tree, whenever a new lineage appears it arises from an existing lineage that is chosen at random. This is the usual description of the ERM model. Analysis of tree shape consists of (*a*) defining a tree balance statistic—several are defined by Kirkpatrick & Slatkin (1993); (*b*) calculating the distribution of the statistic under the ERM model, and (*c*) comparing the statistic’s value in a real tree with this distribution to see if it is unusual.

The general result that has emerged is that real trees are more unbalanced than expected under the ERM model (Moors & Heard 1997, Pinelis 2003) and the conclusion is that there is heterogeneity among lineages in their propensity to diversify.

Though the birth process requires the ERM model of topology, the model itself is not particularly wedded to the birth process: It is simply an algorithm for generating tree topologies under the hypothesis of lineage equivalence. If the tree has a temporal dimension (so far in this topology discussion we have not assumed that it does), we can do a different sort of analysis that is clearly tied to the birth process and, perhaps, reveals more than the summary-statistic approach to tree balance studies.

Figure 7

We ask the question: For each lineage that crosses line *a* in the phylogeny, how many daughter lineages does it have at the time of line *b*?



Consider **Figure 7**. If the tree has grown according to the pure birth process, then the number of daughter lineages for each parental lineage is expected to have a one parameter geometric distribution. This result is very general and does not require, for example, that the birth rate be constant—it can change arbitrarily (Nee et al. 1994). All that matters is that it be the same for each lineage at each point in time.

Figure 8 illustrates the results of such an analysis carried out on Sibley and Ahlquist's molecular phylogeny of birds. As is readily apparent the Passeri (song birds) and Ciconiiformes (shore birds, waders, flamingos, gulls, and such like) were identified as statistical outliers from the fitted geometric distribution. One can speculate as to why these groups radiated so exceptionally: It is surely no coincidence that the Ciconiiform radiation occurred at the time of the breakup of Gondwanaland (Cotgreave & Harvey 1994). Analyzing these two radiations on their own displays an excellent fit to a pure birth process (Nee et al. 1992). This kind of analysis also identified Old World monkeys as a significant radiation, in the sense of being anomalous under the null model, as did our previous analyses of rates of cladogenesis.

What we have just done, of course, is look at the numbers of subtaxa per taxon, and this kind of analysis is meaningful for both molecular and fossil data. It has been known since the work of Willis in the 1920s that the frequency distributions answering such questions—how many species per genus or how many families per order, etc., are of the “hollow curve” variety, with a mode of monotypic taxa and a long tail, such as seen in **Figure 8**. Sepkoski (Sepkoski 1978) was interested in such distributions in the context of the question of whether or not studies of diversity that are only resolved to, say, ordinal level are informative about diversification at the species level. He concluded that they are.

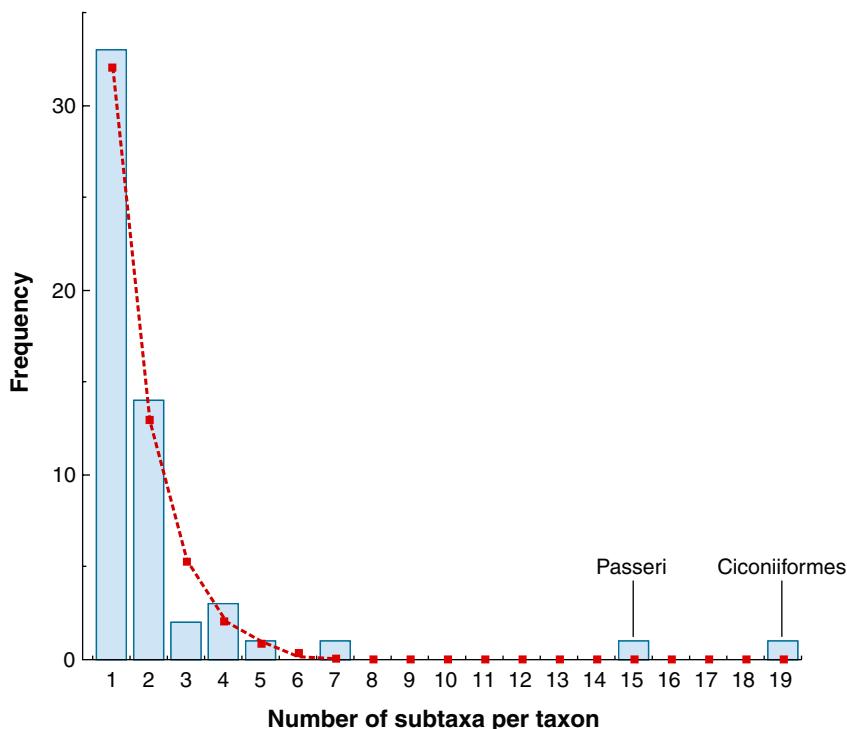


Figure 8

Frequency distribution of the number of daughter lineages per parental lineage for the sort of analysis described in **Figure 7** applied to actual data: in this case the Sibley and Ahlquist phylogeny of the birds. For details see Nee et al. (1992).

SAMPLING AND PARAPHYLY

So far, we have been implicitly assuming that we have all the members of a clade in our analysis. Though there has been great progress in producing more-or-less complete phylogenies of large groups with a temporal dimension (Bininda-Emonds et al. 1999, Purvis 1995), it is often going to be the case that molecular phylogenies will only be a sample of a clade. Paleontological data are often incomplete as well and the exclusion of a lineage and all its descendants from a clade on the basis of the possession of some characteristic is known as paraphyly.

Paraphyly stalks the compendia of fossil data and much thought has been given to how it may bias results. Stung by criticism that the large numbers of paraphyletic taxa in his compendium (Sepkoski 1982) constituted “taxonomic noise” (Smith & Patterson 1988), Sepkoski & Kendrick carried out simulations of evolutionary history using every sort of model discussed so far and threw in the odd mass extinction as well (Sepkoski & Kendrick 1993). They then sampled taxa from the known evolutionary history and asked how readily they allowed that history to be inferred. They concluded that although monophyletic taxa, as expected, perform best, nonetheless paraphyletic taxa do indeed contain a good signal.

On the basis of intuition, it has been thought that paraphyly may make clades appear more bottom-heavy than they actually are (Gould et al. 1987, Smith 1994). In fact, simulation shows that paraphyly is more likely to make clades appear top-heavy

(Uhen 1996). Although Uhen does discuss why this is the case, I am unable to distill his explanation into a succinct, intuitive morsel and encourage any interested reader to consult his paper.

Molecular phylogenies that are based on simply a random sample of a clade that has grown according to the birth or birth-death process will give the misleading impression that the rate of cladogenesis has been slowing down, so the upward curvature in **Figure 4** becomes downward instead (Nee et al. 1994). This effect arises because lineages that have arisen in the recent past are likely to have fewer progeny than older lineages and are, as a result, less likely to have any progeny lineages in the sample. This effect is particularly striking in the phylogenetic trees of viruses that have been spreading exponentially, such as HIV, which have effectively a star-burst phylogeny (e.g., Nee et al. 1995, Slatkin & Hudson 1991) (remember that, for viruses, decades are a macroevolutionary timescale). Sampling has no qualitative effect on trees that have grown according to the Moran process.

CONCLUSION

We will finish off where we began, with a visit to some less obvious uses of birth-death processes for macroevolutionary inference. The two papers to be mentioned here both exploit the concept of evolutionary history, which, in the present context, simply means the sum total of the branch lengths in the phylogeny of a clade (Nee & May 1997). Foote et al. (1999) coupled a birth model—to generate an evolutionary history potentially being laid down in rock—with a stochastic model of actual preservation. The purpose was to ask about the probability that eutherian mammals arose much earlier than the fossil record suggests. They concluded that it is highly unlikely that they arose very long before they actually left a trace of their existence, contrary to molecular evidence.

Nee & May (1997) asked how much evolutionary history would remain in a clade after a mass extinction using birth-death models for analytical results and simulations. They concluded that a surprising amount could be preserved—in a sound-bite, 80% of the evolutionary history of a clade could remain after the extinction of 95% of the species. Essentially this is because the deep branches of the clade are likely to survive even a substantial “pruning of the twigs.” Erwin suggested that this could account for why the biodiversity recovery after the Late Permian mass extinction produced so few higher taxa such as phyla: They were mainly likely to have survived and maintained their position in niche space.

Birth-death models have a distinguished and continuing history in macroevolutionary inference. Occasionally, improved models are recommended. These are usually premised on the idea that more realistic models are to be preferred or that a model that better fits observed patterns is a step forward. In my opinion, there is no substitute for models that have a simple interpretation and lead to easy inference, and whose failure is often of more interest than the ability to fit an observed pattern. I expect the sort of models discussed here to maintain their pre-eminence for a long time to come.

LITERATURE CITED

- Alroy J. 1999. The fossil record of North American mammals: evidence for a Paleocene evolutionary radiation. *Syst. Biol.* 48:107–18
- Aris-Brosou S, Yang Z. 2003. Bayesian models of episodic evolution support a Late Precambrian explosive diversification of the Metazoa. *Mol. Biol. Evol.* 20:1947–54
- Baldwin BG, Sanderson MJ. 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *Proc. Natl. Acad. Sci. USA* 95:9402–6
- Barraclough T, Vogler AP. 2002. Recent diversification rates in North American tiger beetles estimated from a dated mtDNA phylogenetic tree. *Mol. Biol. Evol.* 19:1706–16
- Benton MJ, Ayala FJ. 2003. Dating the tree of life. *Science* 300:1698–700
- Bininda-Emonds ORP, Gittleman JL, Purvis A. 1999. Building large trees by combining phylogenetic information: a complete phylogeny of the extant Carnivora (Mammalia). *Biol. Rev. Camb. Philos. Soc.* 74:143–75
- Cornette JL, Lieberman BS. 2004. Random walks in the history of life. *Proc. Natl. Acad. Sci. USA* 101:187–91
- Cotgreave P, Harvey PH. 1994. Associations among biogeography, phylogeny and bird species diversity. *Biodivers. Lett.* 2:46–55
- Foote M. 1988. Survivorship analysis of Cambrian and Ordovician trilobites. *Paleobiology* 14:258–71
- Foote M. 2001. Evolutionary rates and the age distributions of living and extinct taxa. In *Evolutionary Patterns—Growth, Form and Tempo in the Fossil Record*, ed. JBC Jackson, S Lidgard, FK McKinney, pp. 245–95. Chicago: Univ. Chicago Press
- Foote M, Hunter JP, Janis CM, Sepkoski JJ Jr. 1999. Evolutionary and preservational constraints on origins of biological groups: divergence times of eutherian mammals. *Science* 283:1310–14
- Gilinsky NL, Good IJ. 1991. Probabilities of origination, persistence and extinction of families of marine invertebrate life. *Paleobiology* 17:145–66
- Gould SJ, Gilinsky NL, German RZ. 1987. Asymmetry of lineages and the direction of evolutionary time. *Science* 236:1437–41
- Gould SJ, Raup DM, Sepkoski JJ Jr, Schopf TJM, Simberloff DS. 1977. The shape of evolution: a comparison of real and random clades. *Paleobiology* 3:23–40
- Harvey PH, May RM, Nee S. 1994. Phylogenies without fossils. *Evolution* 48:523–29
- Hewitt GM. 1999. Post-glacial recolonization of European biota. *Biol. J. Linn. Soc.* 68:87–112
- Hey J. 1992. Using phylogenetic trees to study speciation and extinction. *Evolution* 46:627–40
- Hudson RR. 1990. Gene genealogies and the coalescent process. In *Oxford Surveys in Evolutionary Biology*, ed. D Futuyma, J Antonovics, 7:1–49. Oxford, UK: Oxford Univ. Press
- Jablonski D, Roy K, Valentine JW, Price RM, Anderson PS. 2003. The impact of the pull of the recent on the history of marine diversity. *Science* 300:1133–35
- Kendall DG. 1948. On the generalized birth-and-death process. *Ann. Math. Stat.* 19:1–15

- Kirkpatrick M, Slatkin M. 1993. Searching for evolutionary pattern in the shape of a phylogenetic tree. *Evolution* 47:1171–81
- Kitchell JK, MacLeod N. 1988. Macroevolutionary interpretations of symmetry and synchronicity in the fossil record. *Science* 240:1190–93
- Lieberman BS. 2001. A test of whether rates of speciation were unusually high during the Cambrian radiation. *Proc. R. Soc. London Ser. B* 268:1707–14
- Losos JB, Schlüter D. 2000. Analysis of an evolutionary species-area relationship. *Nature* 408:847–50
- Maynard Smith J. 1989. The causes of extinction. *Philos. Trans. R. Soc. London Ser. B* 325:241–52
- Mooers AO, Heard SB. 1997. Inferring evolutionary process from phylogenetic tree shape. *Q. Rev. Biol.* 72:31–55
- Moran PAP. 1958. Random processes in genetics. *Proc. Camb. Philos. Soc.* 54:60–71
- Muller R. 1988. *Nemesis: The Death Star*. New York: Weidenfeld & Nicolson
- Nee S. 2001. Inferring speciation rates from phylogenies. *Evolution* 55:661–68
- Nee S, Holmes EC, May RM, Harvey PH. 1995. Estimating extinction from molecular phylogenies. In *Extinction Rates*, ed. JL Lawton, RM May, pp. 164–82. Oxford, UK: Oxford Univ. Press
- Nee S, May RM. 1997. Extinction and the loss of evolutionary history. *Science* 278:692–94
- Nee S, May RM, Harvey PH. 1994. The reconstructed evolutionary process. *Philos. Trans. R. Soc. London Ser. B* 344:305–11
- Nee S, Mooers AO, Harvey PH. 1992. Tempo and mode of evolution revealed from molecular phylogenies. *Proc. Natl. Acad. Sci. USA* 89:8322–26
- Pinelis I. 2003. Evolutionary models of phylogenetic trees. *Proc. R. Soc. London Ser. B* 270:1425–31
- Purvis A. 1995. A composite estimate of primate phylogeny. *Philos. Trans. R. Soc. London Ser. B* 348:405–21
- Purvis A, Nee S, Harvey PH. 1995. Macroevolutionary inferences from primate phylogeny. *Proc. R. Soc. London Ser. B* 260:329–33
- Raup DM. 1975. Taxonomic survivorship curves and Van Valen's law. *Paleobiology* 1:82–96
- Raup DM. 1978. Cohort analysis of generic survivorship. *Paleobiology* 4:1–15
- Raup DM. 1985. Mathematical models of cladogenesis. *Paleobiology* 11:42–52
- Raup DM, Gould SJ, Schopf TJM, Simberloff DS. 1973. Stochastic models of phylogeny and the evolution of diversity. *J. Geol.* 81:525–42
- Raup DM, Valentine JW. 1983. Multiple origins of life. *Proc. Natl. Acad. Sci. USA* 80:2981–84
- Ribera I, Barraclough T, Vogler AP. 2001. The effect of habitat type on speciation rates and range movements in aquatic beetles: inferences from species-level phylogenies. *Mol. Ecol.* 10:721–35
- Sanderson MJ. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* 14:1218–31
- Sepkoski JJ Jr. 1978. A kinetic model of Phanerozoic taxonomic diversity I. Analysis of marine orders. *Paleobiology* 4:223–51

- Sepkoski JJ Jr. 1982. A compendium of fossil marine animal families. *Milwaukee Public Mus. Contrib. Biol. Geol.* 51:1–125
- Sepkoski JJ Jr, Kendrick DC. 1993. Numerical experiments with model monophyletic and paraphyletic taxa. *Paleobiology* 19:168–84
- Slatkin M, Hudson RR. 1991. Pairwise comparison of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129:555–62
- Smith AB. 1994. *Systematics and the Fossil Record*. Oxford, UK: Blackwell Sci.
- Smith AB, Patterson C. 1988. The influence of taxonomic method on the perception of patterns of evolution. *Evol. Biol.* 23:127–216
- Stanley SM. 1975. A theory of evolution above the species level. *Proc. Natl. Acad. Sci. USA* 72:646–50
- Stanley SM, Newman WA. 1980. Competitive exclusion in evolutionary time: the case of the acorn barnacles. *Paleobiology* 6:173–83
- Stanley SM, Signor PW III, Lidgard S, Karr AF. 1981. Natural clades differ from “random” clades: simulations and analyses. *Paleobiology* 7:115–27
- Strathman RR, Slatkin M. 1983. The improbability of animal phyla with few species. *Paleobiology* 9:97–106
- Turgeon J, Stoks R, Thum RA, Brown JM, McPeek MA. 2005. Simultaneous Quaternary radiations of three damselfly clades across the Holarctic. *Am. Nat.* 165:E78–107
- Uhen MD. 1996. An evaluation of clade-shape statistics using simulations and extinct families of mammals. *Paleobiology* 22:8–22
- Van Valen L. 1973. A new evolutionary law. *Evol. Theory* 1:1–30
- Yule GU. 1924. A mathematical theory of evolution based on the conclusions of Dr. J.C. Willis, FRS. *Philos. Trans. R. Soc. London Ser. B* 213:21–87
- Zink RM, Slowinski JB. 1995. Evidence from molecular systematics for decreased avian diversification in the Pleistocene epoch. *Proc. Natl. Acad. Sci. USA* 92:5832–35