

## STUMPED BY TREES? A GENERALIZED NULL MODEL FOR PATTERNS OF ORGANISMAL DIVERSITY

JONATHAN B. LOSOS<sup>1,\*</sup> AND FREDERICK R. ADLER<sup>2,†</sup>

<sup>1</sup>Department of Biology, Campus Box 1137, Washington University, St. Louis, Missouri 63130;

<sup>2</sup>Center for Population Biology, University of California, Davis, California 95616

*Submitted May 4, 1993; Revised April 20, 1994; Accepted April 22, 1994*

*Abstract.*—Evolutionary biologists increasingly have become interested in the factors determining the structure of phylogenetic trees. For example, highly asymmetric trees seem to suggest that the probability of extinction and/or speciation differs among lineages. Before looking for the cause of such differences, one must establish that the structure of the trees differs significantly from results of an appropriate null model. A commonly invoked null model assumes that speciation is equally probable along all branches of a diversifying tree and predicts that highly asymmetric trees are not unexpected. However, this model of diversification assumes more than equality of rates among lineages. We demonstrate theoretically and via simulations that relaxation of the hidden assumptions that speciation is essentially instantaneous and occurs independently in separate lineages leads to qualitatively different expectations concerning the expected phylogenetic topology. In particular, highly asymmetric trees are considerably less likely to arise by chance when the speciation process takes up a significant fraction of the expected time between initiation of speciation events or when lineages tend to speciate simultaneously. As with most null models in ecology and evolution, the assumptions encoded in the model play an important role in determining the null expectation.

Recent years have seen a resurgence in efforts to explicitly incorporate phylogenetic information into studies of evolutionary phenomena (see, e.g., Brooks and McLennan 1991; Harvey and Pagel 1991 and references therein). One application of this approach is to questions of evolutionary diversification. Biologists have long noted that some lineages are more diverse than others and have proposed causal explanations for such observations. Before asking such a question, however, one must determine whether a group is more (or less) diverse than would be expected by chance. To appropriately investigate this question, one might compare a group with its sister taxon and ask what is the probability that a clade should be composed of two subclades, one with more than  $x$  species and the other with fewer than  $n - x$  species (Slowinski and Guyer 1989).

Such questions require a null model of evolutionary diversification. Null models have proven to be useful tools for investigating whether ecological, evolutionary, or biogeographical patterns differ from random expectations (Harvey et al. 1983).

\* E-mail: losos@biodec.wustl.edu.

† Current address: Department of Mathematics and Department of Biology, University of Utah, Salt Lake City, Utah 84112; E-mail: adler@math.utah.edu.

Nonetheless, null models themselves are based on particular assumptions. Consequently, it is important to critically examine what processes are subsumed within the null model so as to avoid using a null model that is biased (Colwell and Winkler 1984).

One possible null model for evolutionary diversification assumes that as a clade diversifies from one to  $n$  species, speciation is equally probable along all branches in the growing clade (Yule 1924; Harding 1971). This model (hereafter referred to as the Markovian model) is Markovian with respect to the number of species because transition probabilities depend only on the current number of species (Karlin and Taylor 1975); it has been used as the basis for a variety of tests of evolutionary and biogeographical phenomena (Simberloff et al. 1981; Savage 1983; Simberloff 1987; Slowinski and Guyer 1989, 1993; Guyer and Slowinski 1991, 1993; Maddison and Slatkin 1991; Heard 1992; Hey 1992; Harvey and Nee 1993; Kirkpatrick and Slatkin 1993; Brown 1994). The Markovian expectation provides a null model for the likelihood of different phylogenetic topologies when speciation is equally probable in all lineages. The alternative hypothesis might seem to be that speciation (and/or extinction) rates differ among clades, which leads some clades to diversify to a greater or lesser extent than others. Our goal in this article is to show that the Markovian model incorporates a number of other assumptions as well as the assumption that speciation is equally probable on all branches. We construct an alternative model of the speciation process that, without assuming different speciation rates on different branches, produces a distribution of phylogenetic topologies different from those produced by the Markovian model.

Slowinski and Guyer's (1989) examination of the Markovian model is the most detailed to date. Their most surprising conclusion is that asymmetric or unbalanced topologies (i.e., those in which some lineages are considerably more diverse than their sister lineages) are considerably more likely than one might expect. For example, one might think that if probability of speciation were constant among lineages, then it would be unlikely for an eight-species phylogeny to arise in which one species is the sister taxon to a clade containing the other seven (note that we are interested in the topology of unlabeled trees and are not concerned with the particular arrangement of taxa on that topology). Such a result might suggest that probabilities of speciation and/or extinction differ considerably between the two clades. However, the null probability of such a topology with the Markovian model is  $2/7$  (Slowinski and Guyer 1989). Even a tree in which one species is the sister taxon to a clade of 39 species is not significantly nonrandom at the  $P = .05$  level ( $P > .051$ ). More generally, the probability of generating a  $1 + (n - 1)$  tree is  $2/(n - 1)$ , which is equal to the probability of any division of  $n$  species into lineages of unequal size (Slowinski and Guyer 1989); the probability is  $1/(n - 1)$  when the species are evenly divided.

#### ASSUMPTIONS OF THE MARKOVIAN MODEL

The Markovian model assumes that speciation in each lineage is initiated at a rate independent of the history of that lineage. This requires instantaneous specia-

tion, because any duration to the process introduces memory or lag. Noninstantaneous speciation would violate the assumptions of the Markovian model because a population in the process of speciating would not be capable of initiating another speciation event. Consequently, for the duration of the speciation process, the probability of the next speciation event would not be equal on all branches of the tree.

It is not hard to imagine scenarios in which the speciation process must be completed before another speciation event can be initiated. For example, suppose that speciation in a clade occurs primarily via genetic revolutions occurring in small and isolated populations located on the periphery of the species' range (peripatric speciation; Mayr 1963). Species that arose as peripheral isolates are probably not likely to be involved in subsequent speciation events until after genetic restructuring has stabilized and the species has expanded in geographical range and numbers to the extent that it can produce its own peripheral isolates. Similarly, daughter species in a vicariance speciation model will have smaller geographical ranges than their ancestors and thus should be less likely to experience a subsequent vicariant event until their range expands.

Processes that promote speciation in all lineages, such as major vicariant events, may also violate the assumptions of the Markovian model. Imagine, for example, a piece of land breaking off from a continent. If all members of a clade occur both on the continent and on the new island, then each species is likely to split into two species. Such an occurrence would be inconsistent with the Markovian model because speciation events would be more evenly distributed among lineages relative to the distribution that would have resulted if each subsequent speciation event had occurred independently (Kirkpatrick and Slatkin 1993).

We suggest two conditions under which a Markovian model of tree topology generation may be inappropriate. The first occurs when the time between initiation and termination of a speciation event (i.e., the length of time during which daughter species are themselves incapable of speciating, which we term the "refractory period") is long relative to the expected time between initiation of a subsequent speciation event. The second occurs when lineages tend to speciate simultaneously, as when vicariant events promote speciation in all lineages. Modes of speciation vary across taxa, time, and geography, so it is not possible to come to conclusions about the generality of these caveats. Perhaps more importantly, even now so little is known about speciation that we do not have the data to evaluate how important these circumstances may be. We do know, however, that except for speciation via polyploidy and hopeful monsters, speciation is generally not instantaneous. Further, it is clear that in at least some cases, speciation may occur quite rapidly (see, e.g., Mayr 1963; Stanley 1979), but in others it probably takes a considerable period of time. For example, the length of the speciation process averages 1.5–3 million yr in *Drosophila* (Coyne and Orr 1989) and 2 million yr in some salamanders (Larson and Chippindale 1993).

#### A SIMULATION MODEL

Given the paucity of relevant data, we developed a simulation to assess the extent to which violation of the Markovian assumptions leads to alteration in the

distribution of tree topologies. By changing the parameters affecting speciation refractory period and rate, one can determine the effect of the concerns just discussed. To the extent that the simulations produce topologies differing from expectations under a Markovian hypothesis, the Markovian model is not a robust null hypothesis. At the outset, however, we must stress that these simulations were conducted for heuristic purposes. Before simulations such as these can be compared to real phylogenetic topologies, information on rates, refractory period, and mode of speciation must be obtained. Here we investigate the use of Markovian models in addressing the question of whether the basal subdivision of a clade into two subclades is more unbalanced (or balanced) than expected by chance (Slowinski and Guyer 1989). However, the results should apply analogously to other uses of a Markovian model to investigate tree topology (see, e.g., Heard 1992; Hey 1992; Kirkpatrick and Slatkin 1993).

As in the Markovian model, our simulation is based on a speciation rate  $\lambda$ , meaning that a species has a probability of  $\lambda\Delta t$  of initiating speciation in any short interval of time  $\Delta t$  (app. A lists all parameters and variables used in the model). Unlike the Markovian model, when a lineage in our simulation speciates, one of the descendants cannot speciate for an interval of time  $T$ . This model is akin to peripatric or stasipatric models of speciation (Mayr 1963; White 1977) in that one descendant population undergoes marked changes whereas the ancestral population remains unchanged. A scenario in which both descendant populations evolve (and are not capable of speciating again for a period) produces distributions of tree topologies that agree even less with the Markovian model. A more precise model might have the probability of speciation in a lineage be a function of time since the previous speciation event. Although we did not explore this possibility, we suspect that it would yield qualitatively similar results.

To simulate the model, we discretized time into short intervals of equal length and picked a small probability of speciation of 0.001 in each interval. The length  $T$  of the speciation process, the refractory period, is expressed in units of time intervals. We tested  $T$ 's of 10, 50, 100, 500, and 1,000. The simulations begin with two species and proceed until the clade reaches a predetermined number of species. Clades exceeding the predetermined size (because of multiple speciation events within a single time interval) were discarded. We used these simulations to investigate the effect of  $T$  on the likelihood of obtaining various tree topologies. For each parameter value, we created 1,000 trees of 5, 7, 10, 20, and 40 species. For the small five- and seven-species trees, there are only three and 11 possible topologies, respectively, and the relative frequency of each was scored. For the larger trees, we scored only the proportion of trials in which one species was the sister taxon to a clade composed of the remaining  $n - 1$  species, denoting this fraction by  $P(1, n - 1)$ .

In addition, we modified the simulations to assess the effect of infrequent intervals of increased probability of speciation. We set the speciation refractory period to be 10 intervals and the background probability of speciation to be 0.001/interval, a combination that yields results consistent with the Markovian model. How-

ever, in each interval a chance existed that the probability of speciation would be increased to a higher value in every lineage. We tested the effects of varying both the chance of increased speciation probability per interval and the value to which the probability increased. Higher probabilities of speciation were maintained for 10 intervals before returning to the background level. We present results for clades of 20 species.

#### THE SIMULATION APPROXIMATED

Simulations, particularly to produce large trees, can be very slow. We therefore developed a method to approximate  $P(1, n - 1)$ . This fraction can be compared with the results from the Markovian model to give an estimate of the deviation between the two models. The approximation proceeds in two stages. We first find the asymptotic rate of growth of a clade speciating at a rate  $\lambda$  with a speciation refractory period of  $T$ . This rate, denoted by  $\mu$ , is less than  $\lambda$ . We then compute the probability that a single species will not speciate in the time it takes an average species to diversify into  $n - 1$  species.

The asymptotic rate of growth can be found using results of continuous-time branching processes (Karlin and Taylor 1975). In appendix B, we show that the asymptotic growth rate  $\mu$  satisfies

$$\lambda = \mu e^{\mu T}. \quad (1)$$

We find that  $\lambda = \mu$  if  $T = 0$ , as expected, and that  $\mu$  becomes smaller as  $T$  becomes larger. Assuming that one lineage increases in species number at this asymptotic rate of  $\mu$  and reaches a size of  $n - 1$ , we can compute the probability that another lineage will not divide at all. We do this by calculating the density function for the time it takes a lineage to produce  $n - 1$  species and then averaging over the probability  $\exp(-\lambda t)$  that a single species has failed to speciate. This procedure is done explicitly in appendix C, where we find that

$$P(1, n - 1) = 2 \cdot (n - 2) B\left(\frac{\lambda}{\mu} + 1, n - 2\right), \quad (2)$$

where  $B$  is the beta function (Abramowitz and Stegun 1965). For example, if  $T = 0$ , we find that  $P(1, n - 1) = 2/(n - 1)$ , which is identical to the exact result.

Note that equation (1) can be solved for  $\mu T$  given only the value of  $\lambda T$  and thus that  $P(1, n - 1)$  depends on a single parameter. Because the mean time between speciation events is  $1/\lambda$ , the parameter  $\lambda T$  is the ratio of the refractory period to the mean time between speciation events. We term it the "relative refractory period." The Markovian model has a relative refractory period of zero. When the relative refractory period is small, the process is similar to the Markovian model but deviates more and more as the relative refractory period becomes larger. Increasing the rate of the whole process by simultaneously increasing  $\lambda$  and decreasing  $T$  while keeping the relative refractory period  $\lambda T$  constant does

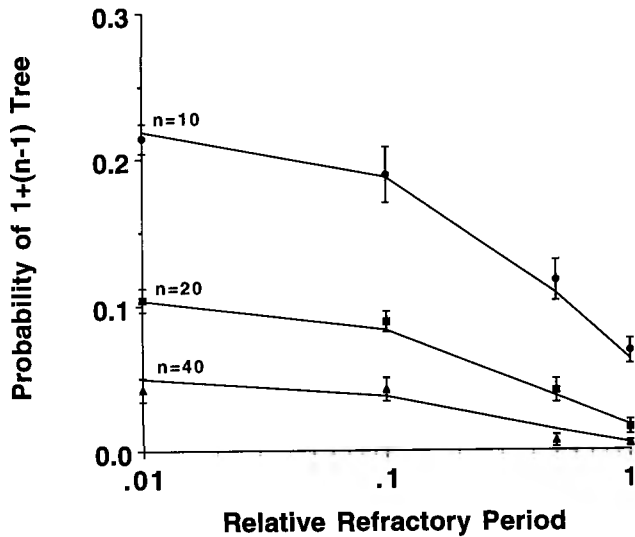


FIG. 1.—Comparison of probabilities of exhibiting a  $1 + (n - 1)$  topology ( $P[1, n - 1]$ ) for the simulation and the approximation for three values of  $n$  (10, 20, and 40). The error bars show one standard deviation from the mean of 500 simulations. The connected lines show the results of the approximation.

not alter the distribution of topologies but only the time it takes to generate them.

#### RESULTS AND DISCUSSION

Figure 1 compares the fraction of  $1 + (n - 1)$  trees produced by the simulation with the fraction predicted by the approximation. Clearly, the approximation does very well over a wide range of parameter values.

As predicted, as the length of the speciation process increases, the distribution of tree topologies diverges from the predictions of the Markovian model. In all the simulations, topology proportions are essentially Markovian when the relative refractory period  $\lambda T$  is less than 0.01, but they become increasingly non-Markovian with an increased frequency of more symmetric trees as the relative refractory period increases. For small trees of size five and seven (figs. 2 and 3, respectively), we illustrate the distribution of all possible topologies. When the speciation refractory period is long, the difference from the Markovian prediction can be large, but noticeable deviations appear even with shorter speciation refractory periods, particularly in the seven-species simulation.

Figure 4 illustrates the effect of the relative refractory period on the topology of trees containing 10–100 species. In this figure, we measure the deviation from the Markovian predictions as

$$\epsilon = 1 - \frac{P(1, n - 1)}{P_M(1, n - 1)}, \quad (3)$$

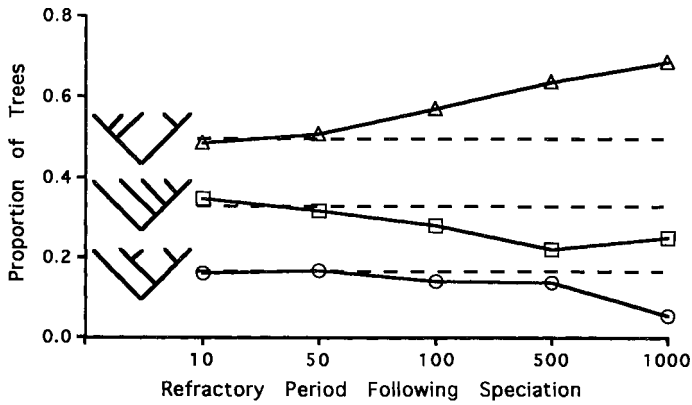


FIG. 2.—Proportion of five-species trees exhibiting each of the three possible topologies. Probability of speciation in all simulations is 0.001 per interval. The dashed line is the Markovian expectation.

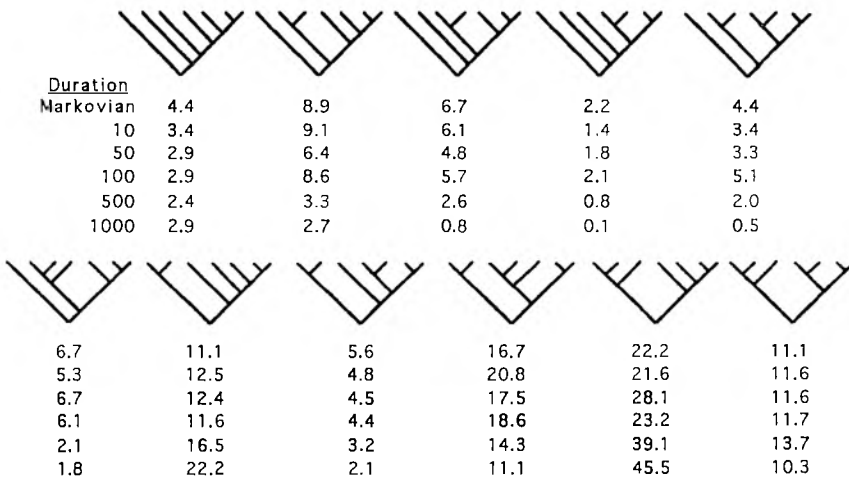


FIG. 3.—Proportion of trees in each of the 11 topologies possible in a seven-species tree. Speciation refractory period varies from 0 (the Markovian case) to 1,000 time intervals.

where  $P_M(1, n - 1)$  represents the Markovian expectation of  $2/(n - 1)$ . Recall that for  $T = 0$  we found that  $P(1, n - 1) = P_M(1, n - 1)$ , so that the error is equal to zero. We show the values that produce errors of 10%, 30%, and 50% for trees of size ranging from 10 to 100.

As with the simulation, deviation from the Markovian model increases with the relative refractory period. Note, too, that larger trees produce the same relative error with a smaller relative refractory period. Consequently, all else being equal, the larger the tree, the less likely it is that Markovian expectations will provide a suitable null model for clade diversification.

The simulations with occasional increased probabilities of speciation indicate

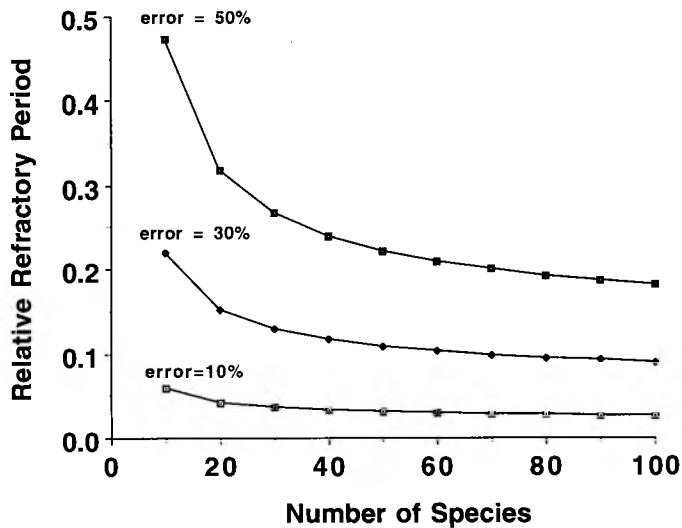


FIG. 4.—The value of the relative refractory period ( $\lambda T$ ) producing errors (fractional deviation from the Markovian model) of three given sizes for a range of clade sizes (i.e., number of species,  $n$ ).

that such rare events can have a large effect on tree topologies (fig. 5). Tree topologies become increasingly non-Markovian with increases in either the probability of such intervals occurring or the degree to which speciation probability is increased. The effect of intervals of increased speciation probability is first noticeable when the probability of such intervals occurring is the same as the background probability of speciation. When such intervals occur at higher probabilities than the background speciation rate, the impact on resulting tree topologies can be quite large.

These results clearly indicate that predictions based on a Markovian branching process may significantly overestimate the likelihood of asymmetric trees when speciation takes an appreciable length of time relative to the rate at which speciation occurs or when circumstances, such as vicariant events, increase probabilities of speciation for all species. This finding would be irrelevant if the rate and length of speciation in nature fell within the range of applicability of the Markovian model. Unfortunately, we know of no data sufficient to estimate all of the relevant parameters. Further, we reemphasize that these results are unique to a particular mode of speciation and set of parameters. We need information on probabilities of speciation, whether (or to what extent) a speciation event must be completed before a lineage can begin to split again, and how likely are various modes of speciation. In addition, we need to know whether most speciation events are concentrated in relatively brief intervals of time when events lead to greatly enhanced probabilities of speciation, as, for example, vicariant biogeographers might suggest (see also Vrba 1985). To the extent that such periods are relatively important, Markovian models are particularly likely to be inaccurate.

Thus, we do not conclude from these analyses that a Markovian model is an inappropriate means of estimating a null hypothesis for lineage diversification.



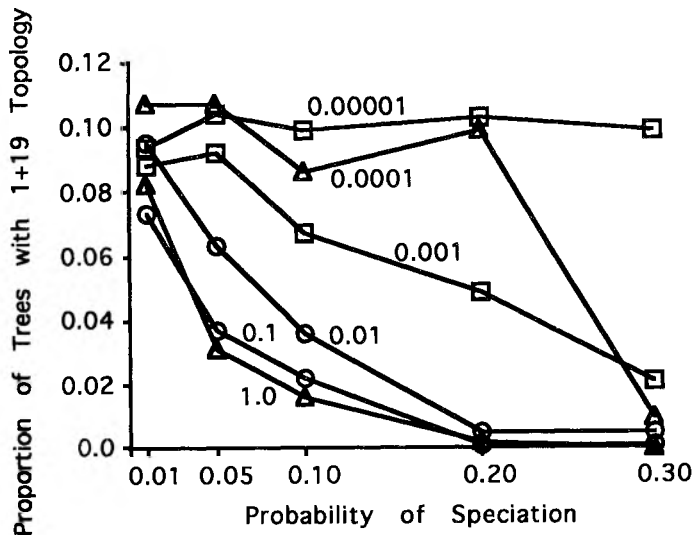


FIG. 5.—Proportion of trees exhibiting a 1 + 19 topology. The background speciation rate is 0.001 with a speciation refractory period of 10 intervals. In some intervals, however, speciation probability was increased. The X-axis indicates the level to which the speciation probability was increased, whereas the numbers in the figure indicate the probability that the speciation rate would be increased from the background rate in any given interval (once increased, the speciation rate did not revert to background levels for 10 generations).

Rather, our analyses are useful in two ways. First, they focus attention on what are the important parameters that need to be known before the applicability of the Markovian model can be evaluated. Given that this information is available for few, if any, lineages, the acceptance of a single robust null hypothesis may be unlikely. Alternatively, our approach can serve as a sensitivity analysis to determine how robust a particular result is to alteration in various parameters. If the likelihood of generating a particular observed tree topology is less than 0.05 or greater than 0.50 (or some other large value) regardless of what parameters are used, then it probably does not matter which parameters are actually correct.

A second approach to evaluate the applicability of the Markovian model is possible when the age and size of a lineage are known. In such cases, we can identify the range of speciation refractory periods that are consistent with the Markovian model. We show that in some cases differentiation could have followed a Markovian model only if speciation refractory periods were extremely short, a finding that might imply that diversification has not been Markovian.

Suppose a radiation has produced  $n$  species in time  $t$ . Assume a speciation refractory period of  $T$ . The realized rate of increase  $\mu$  of the population then satisfies

$$e^{\mu t} = n$$

or

$$\mu = \frac{\log(n)}{t}.$$

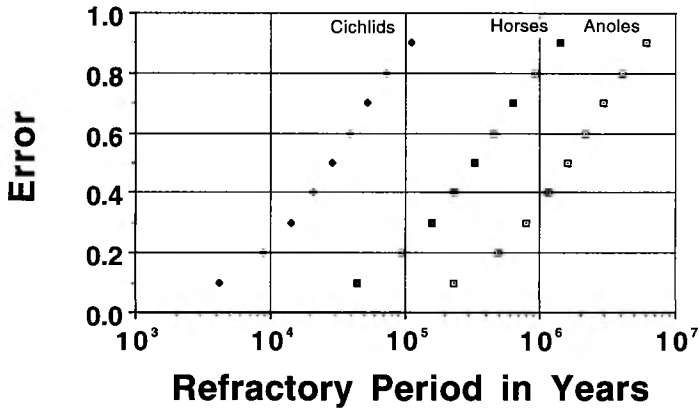


FIG. 6.—The error (deviation from the Markovian expectation) as a function of speciation refractory period for the three clades discussed in the text.

The deviation from a Markovian process is a function of the ratio of  $\lambda$  to  $\mu$  (see eq. [2]), which is equal to  $\exp(\mu T)$  from equation (1). Therefore,

$$\frac{\lambda}{\mu} = e^{\log(n)T/t} \quad (4)$$

This is sufficient information to compute the error  $\epsilon$  as a function of the assumed speciation time  $T$  using equation (3). As before, if  $T = 0$ , we have  $\epsilon = 0$  and no deviation from the Markovian process.

Figure 6 shows the results of this calculation based on three well-studied clades: Lake Victoria cichlids, Miocene grazing horses, and anoline lizards. In Lake Victoria, a monophyletic clade of cichlid fishes has produced more than 200 species in less than 1 million yr (Meyer et al. 1990). For the results of the speciation process to be within 10% of the Markovian prediction, the speciation refractory period  $T$  would have to be less than 4,000 yr. To even be within 50% of the Markovian prediction, the speciation refractory period would have to be less than 29,000 yr. Thus, unless speciation is completed rapidly in these fishes, the high rate of speciation in the Lake Victoria cichlid radiation is probably inconsistent with a Markovian model of diversification.

By contrast, for the radiation of anoline lizards over the past 60+ million yr (250+ species, following figures in Slowinski and Guyer 1989), speciation refractory periods need only be less than 230,000 yr for results to be within 10% of the Markovian prediction. To the extent that this is a reasonable speciation time, the Markovian model may be an appropriate null model. The radiation of Miocene grazing horses (19 species in 3 million yr; MacFadden and Hulbert 1988) is intermediate between the other two examples.

More generally, figure 7 shows various levels of deviation from the Markovian model as a function of the number of species in a clade and the speciation time relative to the age of the clade. When the speciation refractory period is even a

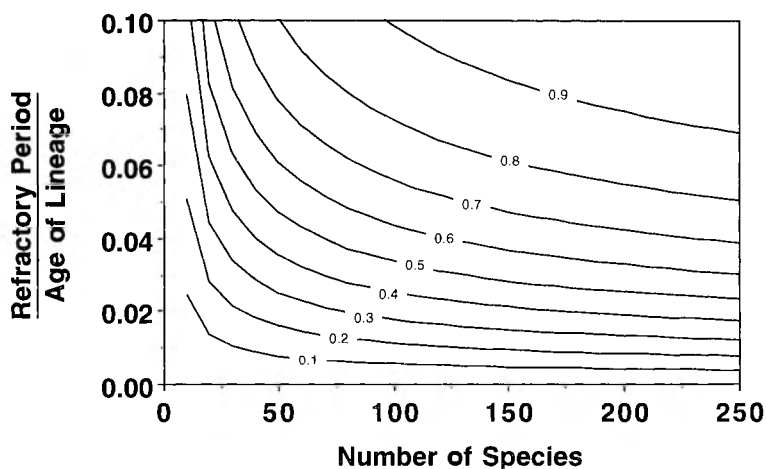


FIG. 7.—A contour plot of error as a function of the size of a clade and the refractory period of a single speciation event ( $T$ ) relative to the age of the clade ( $t$ ).

few percentage points of the age of a moderately sized clade, the Markovian model will be inaccurate.

The calculations for these cases may overestimate the applicability of the Markovian model because they consider neither extinction nor the possibility of periods of increased rates of speciation. Periods of increased probability of speciation could be consonant with the Markovian model only if speciation refractory periods decreased simultaneously. In addition, because we cannot account for extinct species, the number of speciation events may have been considerably greater than inferred based solely on knowledge of extant taxa, and thus the probability of speciation estimated from extant taxa might be an underestimate. Claims that extinction would not alter the results of Markovian null models (Simberloff et al. 1981; Slowinski and Guyer 1989) are correct only when both speciation and extinction occur in a Markovian manner.

As with most null models in ecology and evolution, the Markovian model of branching diversification encodes many biological assumptions. When these assumptions are met, the Markovian model provides an easy way to estimate the null probability of generating a given tree topology. To the extent that these assumptions are not met, however, a Markovian model will not accurately indicate the likelihood of obtaining a particular topology when speciation rates are equal across all lineages and thus will not be an appropriate null model to contrast with deterministic explanations of species diversity within a clade. Our analyses are a first step in evaluating under what circumstances the Markovian model is likely to prove robust as a null hypothesis.

#### ACKNOWLEDGMENTS

This work was begun while J.B.L. was a postdoctoral fellow at the Center for Population Biology, University of California, Davis, and F.R.A. held a post-

doctoral position with M. Mangel at the University of California, Davis. We thank C. Guyer, M. Mangel, W. Morris, J. Slowinski, A. Templeton, and an anonymous reviewer for helpful comments on previous versions of this work and M. Kirkpatrick and J. Slowinski for providing manuscript copies prior to publication.

## APPENDIX A

TABLE A1

VARIABLES AND THEIR MEANINGS

Symbol	Description
$n$	Number of species in lineage
$\lambda$	Speciation rate
$T$	Refractory period
$\mu$	Asymptotic growth rate of lineage
$P(1, n - 1)$	Probability of a $1 + (n - 1)$ tree topology
$P_M(1, n - 1)$	Probability of a $1 + (n - 1)$ tree topology (Markovian case)
$\epsilon$	Fractional deviation from Markovian prediction
$t$	Lineage age

## APPENDIX B

## EFFECTS OF SPECIATION REFRACTORY PERIOD ON CLADE SIZE

In this appendix, we compute the asymptotic growth rate of a clade speciating at intrinsic rate  $\lambda$  with speciation refractory period  $T$ . The density function describing the time at which a new type initiates speciation after creation is denoted by  $f(\tau)$  and satisfies

$$f(\tau) = \begin{cases} \frac{\lambda}{2} e^{-\lambda(\tau-T)} + \frac{\lambda}{2} e^{-\lambda\tau} & \text{if } \tau > T \\ \frac{\lambda}{2} e^{-\lambda\tau} & \text{if } \tau \leq T. \end{cases}$$

This is an average of the time to initiate speciation by the ancestral and descendent populations. Defining

$$F(t) = \int_0^t f(\tau) d\tau,$$

we have that the expected number of species at time  $t$ ,  $m(t)$  satisfies (Karlin and Taylor 1975, p. 435)

$$\begin{aligned} m(t) &= 2 \int_0^t m(t - \tau) f(\tau) d\tau + 1 - F(t) \\ &= \int_0^t \lambda m(t - \tau) e^{-\lambda\tau} d\tau + \int_T^t \lambda m(t - \tau) e^{-\lambda(\tau-T)} d\tau + \frac{1 + e^{\lambda T}}{2} e^{-\lambda t}. \end{aligned}$$

Multiplying both sides by  $e^{\lambda t}$ , we find that

$$\begin{aligned} m(t) e^{\lambda t} &= \int_0^t \lambda m(t - \tau) e^{\lambda(t-\tau)} d\tau + e^{\lambda t} \int_T^t \lambda m(t - \tau) e^{\lambda(t-\tau)} d\tau + \frac{1 + e^{\lambda T}}{2} \\ &= \int_0^t \lambda m(u) e^{\lambda u} du + e^{\lambda T} \int_0^{t-T} \lambda m(u) e^{\lambda u} du + \frac{1 + e^{\lambda T}}{2}. \end{aligned}$$

Differentiating gives

$$e^{\lambda t}[\lambda m(t) + m'(t)] = \lambda m(t)e^{\lambda t} + e^{\lambda T}\lambda m(t - T)e^{\lambda(t-T)},$$

where  $m'$  denotes the derivative of  $m$ . We then have that

$$m'(t) = \lambda m(t - T).$$

If we assume that

$$m(t) = e^{\mu t},$$

the solution must satisfy

$$\mu e^{\mu t} = \lambda e^{\mu(t-T)}.$$

Rearranging terms, we find

$$\mu e^{\mu T} = \lambda,$$

which is precisely equation (1).

### APPENDIX C

#### APPROXIMATE EFFECTS OF SPECIATION REFRACTORY PERIOD ON THE PROBABILITY OF HIGHLY UNBALANCED TREES

In this appendix, we compute the probability that a single species, with the potential to speciate at rate  $\lambda$ , fails to do so by the time that a clade increasing in size at rate  $\mu$  reaches  $n - 1$  species. We pretend that the growing clade follows exactly a pure birth process, so that the probability of exactly  $k$  species at time  $t$  satisfies (Karlin and Taylor 1975, p. 122)

$$p_k(t) = e^{-\mu t}(1 - e^{-\mu t})^{k-1}.$$

The density function for the time to reach exactly  $k$ ,  $s_k(t)$  is then

$$s_k(t) = (k - 1)\mu p_{k-1}(t) = (k - 1)\mu e^{-\mu t}(1 - e^{-\mu t})^{k-2}.$$

The probability that the singleton has not speciated in time  $t$  is  $e^{-\lambda t}$ . Thus, the probability of a  $1 + (n - 1)$  tree,  $\Pi(1, n - 1)$ , is

$$\begin{aligned} \Pi(1, n - 1) &= \int_0^\infty e^{-\lambda t} s_{n-1}(t) dt \\ &= \int_0^\infty (n - 2)\mu e^{-\lambda t} e^{-\mu t}(1 - e^{-\mu t})^{n-3} dt \\ &= \int_0^1 (n - 2)u^{\lambda/\mu}(1 - u)^{n-3} du \\ &= (n - 2)B(\lambda/\mu + 1, n - 2), \end{aligned}$$

where we substituted  $u$  for  $e^{-\mu t}$ , and  $B$  denotes the standard beta function (Abramowitz and Stegun 1965). As used in the text, the probability  $P(1, n - 1)$  of a  $1 + (n - 1)$  tree includes both the chance of a  $1 + (n - 1)$  tree as computed above and the identical probability of an  $(n - 1) + 1$  tree and is thus twice  $\Pi(1, n - 1)$ . This gives equation (2).

#### LITERATURE CITED

- Abramowitz, M., and I. A. Stegun. 1965. Handbook of mathematical functions. Dover, New York.  
 Brooks, D. R., and D. A. McLennan. 1991. Phylogeny, ecology, and behavior: a research program in comparative biology. University of Chicago Press, Chicago.  
 Brown, J. K. M. 1994. Probabilities of evolutionary trees. Systematic Zoology 43:78-91.

- Colwell, R. K., and D. W. Winkler. 1984. A null model for null models in biogeography. Pages 344–359 in D. R. Strong, Jr., D. Simberloff, L. G. Abele, and A. B. Thistle, eds. *Ecological communities: conceptual issues and the evidence*. Princeton University Press, Princeton, N.J.
- Coyne, J. A., and H. A. Orr. 1989. Patterns of speciation in *Drosophila*. *Evolution* 43:362–381.
- Guyer, C., and J. B. Slowinski. 1991. Comparisons of observed phylogenetic topologies with null expectations among three monophyletic lineages. *Evolution* 45:340–350.
- . 1993. Adaptive radiation and the topology of large phylogenies. *Evolution* 47:253–263.
- Harding, E. F. 1971. The probabilities of rooted tree-shapes generated by random bifurcation. *Advances in Applied Probability* 3:44–77.
- Harvey, P. H., and S. Nee. 1993. New uses for new phylogenies. *European Review* 1:11–19.
- Harvey, P. H., and M. D. Pagel. 1991. *The comparative method in evolutionary biology*. Oxford University Press, Oxford.
- Harvey, P. H., R. K. Colwell, J. V. Silvertown, and R. M. May. 1983. Null models in ecology. *Annual Review of Ecology and Systematics* 14:189–211.
- Heard, S. B. 1992. Patterns in tree balance among cladistic, phenetic, and randomly generated phylogenetic trees. *Evolution* 46:1818–1826.
- Hey, J. 1992. Using phylogenetic trees to study speciation and extinction. *Evolution* 46:627–640.
- Karlin, S., and H. M. Taylor. 1975. *A first course in stochastic processes*. 2d ed. Academic Press, New York.
- Kirkpatrick, M., and M. Slatkin. 1993. Searching for evolutionary patterns in the shape of a phylogenetic tree. *Evolution* 47:1171–1181.
- Larson, A., and P. Chippindale. 1993. Molecular approaches to the evolutionary biology of plethodontid salamanders. *Herpetologica* 49:204–215.
- MacFadden, B. J., and R. C. Hulbert. 1988. Explosive speciation at the base of the adaptive radiation of Miocene grazing horses. *Nature (London)* 336:466–468.
- Maddison, W. P., and M. Slatkin. 1991. Null models for the number of evolutionary steps in a character on a phylogenetic tree. *Evolution* 45:1184–1197.
- Mayr, E. 1963. *Animal species and evolution*. Belknap, Cambridge, Mass.
- Meyer, A., T. D. Kocher, P. Basasibwaki, and A. C. Wilson. 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature (London)* 347:550–553.
- Savage, H. M. 1983. The shape of evolution: systematic tree topology. *Biological Journal of the Linnean Society* 20:225–244.
- Simberloff, D. S. 1987. Calculating probabilities that cladograms match: a method of biogeographic inference. *Systematic Zoology* 36:175–195.
- Simberloff, D. S., K. L. Hecht, E. D. McCoy, and E. F. Connor. 1981. There have been no statistical tests of cladistic biogeographical hypotheses. Pages 40–63 in G. Nelson and D. E. Rosen, eds. *Vicariance biogeography: a critique*. Columbia University Press, New York.
- Slowinski, J. B., and C. Guyer. 1989. Testing the stochasticity of patterns of organismal diversity: an improved null model. *American Naturalist* 134:907–921.
- . 1993. Testing whether certain traits have caused amplified diversification: an improved method based on a model of random speciation and extinction. *American Naturalist* 142:1019–1024.
- Stanley, S. M. 1979. *Macroevolution: pattern and process*. W. H. Freeman, San Francisco.
- Vrba, E. S. 1985. Environment and evolution: alternative causes of the temporal distribution of evolutionary events. *South African Journal of Science* 81:229–236.
- White, M. J. D. 1977. *Modes of speciation*. W. H. Freeman, San Francisco.
- Yule, G. U. 1924. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis. F.R.S. *Philosophical Transactions of the Royal Society of London A, Mathematical and Physical Sciences* 213:21–87.