INFERRING EVOLUTIONARY PROCESS FROM PHYLOGENETIC TREE SHAPE

ARNE Ø. MOOERS

Department of Zoology, University of British Columbia
6270 University Boulevard, Vancouver, V6T 1Z4 Canada

STEPHEN B. HEARD

Department of Biological Sciences, University of Iowa
138 Biology Building, Iowa City, Iowa 52242–1324 USA

ABSTRACT

Inferences about macroevolutionary processes have traditionally depended solely on the fossil record, but such inferences can be strengthened by also considering the shapes of the phylogenetic trees that link extant taxa. The realization that phylogenies reflect macroevolutionary processes has led to a growing literature of theoretical and comparative studies of tree shape. Two aspects of tree shape are particularly important: tree balance and the distribution of branch lengths. We examine and evaluate recent developments in and connections between these two aspects, and suggest directions for future research.

Studies of tree shape promise useful and powerful tests of macroevolutionary hypotheses. With appropriate further research, tree shape may help us detect mass extinctions and adaptive radiations, measure continuous variation in speciation and extinction rates, and associate changes in these rates with ecological or biogeographical causes.

The usefulness of tree shape extends well beyond the study of macroevolution. We discuss applications to other areas of biology, including coevolution, phylogenetic inference, population biology, and developmental biology.

INTRODUCTION

THERE IS probably no greater nor more fundamental problem in ecology and evolutionary biology than explaining the diversity of life on Earth. A comprehensive explanation of this diversity should account for not only the regulation of diversity in extant ecological communities (Ricklefs and Schluter 1993), but also provide an understanding of what factors have determined historical patterns of diversification among evolutionary groups. This latter question is a familiar one:

Why are there so many species of passerine birds (e.g., Raikow 1986; Kochmer and Wagner 1988) or beetles (e.g., Evans 1975)? J B S Haldane's famous quip about God's inordinate fondness for beetles can only be part of the story.

Questions about the relative diversity of evolutionary groups are often posed in terms of taxonomies (see literature reviewed by Anderson 1974). Ultimately, however, these are questions about variation in speciation and extinction rates, and their signatures in the shapes of phylogenetic trees. The recent explosion in the availability of phylogenetic data

(fueled in part by the molecular revolution) has given us a new and powerful approach for studying tempo and mode in evolution: phylogenetic trees as histories of the diversification of clades. The availability of these data has in turn led to dramatic growth in the study of phylogenetic tree shape, which provides a possible new approach to the study of evolution and diversification (there are other applications of tree shape, but we touch on them only briefly). The time is ripe for an appraisal of the power and prospects of the study of tree shape. Here we outline the history behind the use of tree shape, provide some necessary technical background, and consider current descriptions of and proposed explanations for the shapes we observe. We then conclude with suggestions for future study.

HISTORICAL SKETCH

Branching diagrams were considered fitting representations of organismal affinities long before Darwin (Panchen 1992). However, Darwin is usually recognized as the first to consider trees explicitly as representations of actual genealogies of species, and his theory of natural selection gave him definite ideas as to their shape (Darwin 1859:110). Faced with limited data, however, Darwin turned to taxonomies instead of phylogenies for evidence to support his hypotheses (Darwin 1859:110). The use of taxonomies as reflections of phylogenetic pattern continues up to the present.

Willis (1922) noted the concave shape of many frequency distributions of taxonomic units containing various numbers of subunits (e.g., species per genus), when units are ordered from those with the greatest number to those with fewest subunits, and called this the "hollow curve" distribution. For example, Figure 1 shows the sizes of chrysomelid beetle genera (from Williams 1964; based on data collected by Willis). It plots the number of chrysomelid beetle genera with varying numbers of species. The hollow curve phenomenon is familiar to anyone who has noticed the conspicuous abundance of, for example, beetles among insects, passerines among birds, or teleosts among fish.

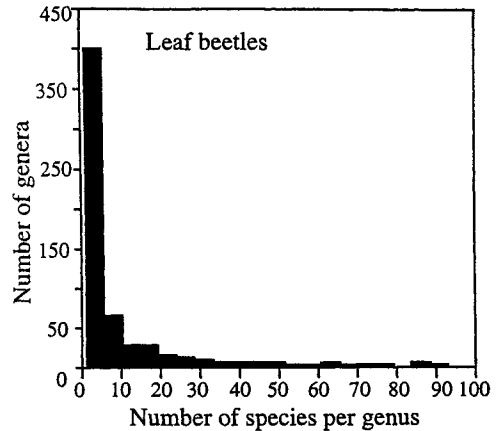Early authors described these hollow curves as hyperbolic (Chamberlin 1924), logarithmic

FIGURE 1. A HOLLOW CURVE DISTRIBUTION OF THE NUMBER OF GENERA CONTAINING VARIOUS NUMBERS OF SPECIES OF LEAF BEETLE (CHRYSOMELIDAE). Data from Williams (1964, Table 50).

(Fisher et al. 1943), or log-normal series (Williams 1964), and saw in them the stamp of deterministic processes (Williams 1964; Anderson 1974). Walters (1961; see also Cronk 1989) considered them merely psychohistorical artifacts. Yule (1924) and Wright (1941) pointed to a stochastic or probabilistic explanation for the hollow curve pattern. Ecological theory—e.g., MacArthur's (1957) "broken-stick" model of ecological diversity, which is also probabilistic—was adapted to explain hollow curves, most prominently by Anderson (1974). Recent investigators have even described the hollow curves in terms of fractal geometry (Burlando 1990; Minelli et al. 1991; see also Nee et al. 1992). Information-rich, rigorous phylogenies have now largely replaced the use of taxonomic lists in the study of macroevolution (but see exceptions in Dial and Marzluff 1989; Ricklefs and Renner 1994; Tiffney and Mazer 1995).

The modern study of phylogenetic tree shape and macroevolution can be traced to a set of meetings of paleontologists and population biologists at Woods Hole in the early 1970s (Raup et al. 1973; Gould et al. 1977; Schopf 1979), and to a biogeographic study (Rosen 1978; Simberloff et al. 1981; Simberloff 1987). Both sets of investigators were seeking falsifiable hypotheses of phylogenetic tree

shape. The "Woods Hole group" (coined by Slowinski and Guyer 1989) was concerned that paleontologists took an overly descriptive and deterministic approach when considering the patterns of fossil diversity. The group felt that some deductive methodology was needed, with a view to formulating general evolutionary laws (Raup et al. 1973). Taking their cue from ecology (Simberloff 1972), they created a simple stochastic model of macroevolution (with lineages splitting and becoming extinct) and used it to simulate phylogenies evolving through geological time. Because the distribution of shapes of real clades from the literature resembled the simulated random clades, Raup et al. (1973; Gould et al. 1977) held that deterministic explanations for clade shape were unnecessary. However, Stanley et al. (1981) found that with more realistic parameter values the histories of clades from the fossil record could not be reproduced with simple stochastic models. The Woods Hole studies are part of a rich literature concerning the course of diversity in the fossil record, but we will not review it here. Current studies of phylogenetic tree shape are generally intellectual descendants of the Woods Hole group and its use of a simulation-driven, stochastic model of tree production.

The biogeographic approach owes much to Hennig (1966), who postulated a series of biogeographic "rules" (summarized by Ashlock 1974) by which cladograms could be used to infer past biogeographic events. In order to give these rules a quantitative basis, Rosen (1978; see also Platnick and Nelson 1978) considered the probability that two area cladograms (i.e., two cladograms of different groups with the common areas of occurrence substituted for species) were congruent by chance alone. For instance, Rosen considered perfectly congruent area cladograms for two groups of five species of fish, and pointed out that, given the 105 possible labeled cladograms having five tips, the chance that two five-tipped trees would match perfectly owing to chance would be $(1/105) = 0.0095$. This was unlikely enough to justify appeals to similar biogeographic trajectories for the two groups. Simberloff and coworkers (Simberloff et al. 1981; Simberloff 1987) pointed out that Rosen's statistics were based on the particular

and unsupported null expectation that all labeled cladograms were equally likely, and presented two other null distributions (see the section below on null models for tree balance, p 35). Simberloff et al. (1981) highlighted the necessity of using clear null models when studying congruence among phylogenies, and showed how certain shapes of trees might be expected to arise in nature. Together, the Woods Hole group's agenda of a quantitative approach to paleontological studies of tempo and mode, and the response to the use of null models in tests of biogeography, set the stage for the modern use of phylogenetic tree shape in the study of macroevolution. The two approaches also highlight two different but closely related aspects of tree shape: changes in diversification rates through time, and differences in diversification patterns among contemporary clades. These might be thought of as two different axes of variation. We consider the latter aspect first, beginning with a brief summary of terminology and the statistical approaches used.

TERMINOLOGY FOR PHYLOGENETIC TREE SHAPE

Most biological taxa have arisen by a branching evolutionary process of descent with modification (we are still a long way from a proper account of reticulation in phylogenetic theory). While taxonomies generally reflect (to various degrees) historical relationships between taxa, phylogenetic trees are explicit statements of those relationships. That is, a phylogenetic tree is an estimate of the evolutionary history of the group in question. We will use the term *estimated tree* to denote a particular phylogenetic tree inferred from real data; this contrasts with the unknowable *true tree* we are trying to retrieve. The estimation may have involved any of a number of techniques, either phenetic or cladistic; despite differences among techniques, in current practice all are normally intended to retrieve evolutionary relationships.

Some further definitions will be useful. Any phylogenetic tree (Figure 2A) is made up of *tips, nodes,* and *branches.* A *node* is a point where two or more *branches* meet, and each node ultimately represents a speciation event whose age may or may not be estimated; the oldest
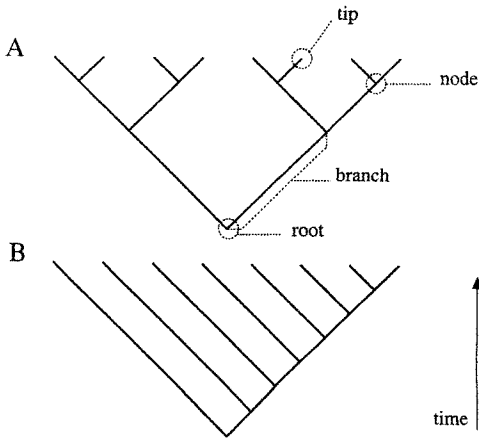
FIGURE 2.   TWO HYPOTHETICAL 8-TIPPED
PHYLOGENETIC TREES.

These trees illustrate the range of possibilities
for tree balance and some basic terminology for
elements of trees. Tree (A) is completely balanced;
tree (B) is completely imbalanced.

node is the *root*. *Tips* are the smallest taxo-
nomic units shown in a tree. A *branch* connects
a tip or node to the next nearest node, and its
length may reflect inferred amounts of time,
genetic change, or morphological change be-
tween nodes or between nodes and tips. *Binary
trees* have no more than three branches meet-
ing at any node: an ancestral branch and two
daughter branches. A tree is labeled if each
tip is uniquely identified (normally as a species
or other taxonomic group); it is unlabeled
otherwise. A tree is *complete* if it includes all
extant taxa in a given group (for instance, all
seven recent species in the horse family).
Many trees are incomplete because species are
unknown, or either unavailable to or omitted
by the systematist.

Tree topology refers to the branching pattern
of a tree, but ignores branch length informa-
tion and labels on the tips. Two unlabeled
trees have the same topology if they can be
made identical (ignoring branch lengths) by
swapping the positions of the right-hand and
left-hand branches at any set of nodes. *Balance*,
or the extent to which nodes define subgroups
of equal sizes (Figure 2B), is the most widely
considered aspect of topology. Balanced trees
have also been called symmetrical, while im-

balanced trees have been called pectinate,
chained, or comblike ("unbalanced" is also
used, and is identical in meaning). We use *tree
shape* as a general term that includes both bal-
ance and the distribution of branch lengths in
the tree.

TREE BALANCE

MEASURES OF TREE BALANCE

Although it is possible to examine patterns
in the frequencies of all possible tree topolo-
gies (Savage 1983; Brown 1994), this quickly
becomes unwieldy with trees larger than a few
species. An alternative is to use a summary in-
dex of overall tree balance. Many different in-
dices have been proposed; they differ in calcu-
lation and to some minor extent in behavior
(Shao and Sokal 1990; Kirkpatrick and Slatkin
1993). Most have in common the fact that
each summarizes, in a single number, the de-
gree to which a tree is balanced or imbalanced
(but see Sackin 1972; Fusco and Cronk 1995;
and the use of only a single node by Slowinski
and Guyer 1989, 1993; Guyer and Slowinski
1993). Balance indices ignore any informa-
tion about branch lengths (Figure 2) and con-
sider only the distribution of tips across nodes.
We assume for the moment that the trees un-
der consideration are binary.

The most widely used index is Colless'
(1982) index of imbalance I_c (C in Kirkpatrick
and Slatkin 1993). I_c sums, over all $(n - 1)$
nodes in a tree with n tips, the difference in
the numbers of tips subtended by the right-
hand and left-hand branches arising at each
node, and then normalizes by dividing by the
largest possible score. That is,

$$I_c = \frac{\sum_{\text{(all interior nodes)}} | T_R - T_L |}{\frac{(n - 1)(n - 2)}{2}}, \quad (1)$$

where at each node the right and left branches
subtend T_R and T_L tips. $I_c = 0$ for a completely
balanced tree and $I_c = 1$ for a completely im-
balanced one. Kirkpatrick and Slatkin (1993)
point out that I_c gives more weight to nodes
close to the root (for which $| T_R - T_L |$ can be
largest). This is reasonable because those
nodes are the most informative, as the sub-
clades they define are older and therefore
sample longer periods of evolutionary time. It

is possible to construct an alternative measure that weights all nodes equally:

$$I_2 = \frac{\displaystyle\sum_{(j-\text{nodes},\,j>2)} \frac{\mid T_R - T_L \mid}{j-2}}{(n-2)}, \qquad (2)$$

where j is the number of tips subtended by each interior node. The cost of this alternative is that I_2 is a less familiar index and its behavior has yet to be characterized.

Five other indices have been examined in some detail by Kirkpatrick and Slatkin (1993): $\overline{N}$, σ_N^2, B_1, B_2, and R. $\overline{N}$ and σ_N^2 derive from Sackin's (1972) vector measure of tree balance: If N_i is the number of nodes between tip i and the root, then the mean and variance of N over all i (i.e., $\overline{N}$ and σ_N^2) are two slightly different measures of tree shape [Shao and Sokal (1990) used ΣN instead of $\overline{N}$]. Both increase with increasing imbalance. A third related measure is $B_2 = \Sigma(N_i/2^{N_i})$, which is related to the Shannon-Wiener information function [note that B_2 in Kirkpatrick and Slatkin (1993) is equal to B_2 divided by $\log(2)$ in Shao and Sokal (1990)]. B_2 is larger for more balanced trees. B_1 is more complicated: For each of the interior nodes except the root, determine M_i, the maximum number of other nodes between that node and a tip it subtends. Then $B_1 = \Sigma\ 1/M_i$ (Shao and Sokal 1990); B_1 is larger for more balanced trees. Finally, Page (1993b) developed an alternative measure, which requires an ordering of all possible trees of size n by Furnas' (1984) "left-light rooting" rules. These rules place a perfectly imbalanced tree first, and a perfectly balanced one last, and the balance index R for a particular tree is just the position of the tree in that order.

There seems to be no end to the number of balance indices that can be derived. Rogers (1996) has suggested counting the number of nodes in a tree that are imbalanced (J). Other variations on the indices discussed here appear in Shao and Sokal (1990). Furthermore, tree topology is of interest in other fields, and analogous approaches can be found (e.g., in hydrology, Smart 1972; in cell morphology, Van Pelt et al. 1989).

Trees that include polytomies (i.e., which are not binary) pose problems for the measurement of balance. Shao and Sokal (1990) discuss the application of several indices to trees including polytomies, but for such trees expected values and moments of the indices are much less easy to derive. In many or most cases, trees with polytomies reflect only our incomplete knowledge; for instance, a wide-ranging species, budding new species from peripheral isolates but not itself changing phenotype, could produce a clade we would estimate under maximum parsimony as a polytomy (Wagner and Erwin 1995). The balance of such unresolved trees may have little to tell us. However, "hard" (true) polytomies can result from simultaneous speciation events in a single lineage (Hoelzer and Melnick 1994); such cases are difficult to accommodate with current methods. The usual practice is to consider only binary trees.

How are we to choose among the alternative measures of balance? Kirkpatrick and Slatkin (1993) compared the power of six statistics for discriminating between two simple models of evolution. $\overline{N}$, R, and B_2 showed relatively low power, but I_c, σ_N^2, and B_1 all performed fairly well. Of the latter three, I_c is computationally the simplest, and is the only one whose expectation and variance (under a simple null model; see below) are known analytically (Heard 1992; Rogers 1994). For these reasons we expect that I_c will be, for most purposes, the most useful measure of balance, despite a slight edge in power reported for B_1 (Kirkpatrick and Slatkin 1993).

NULL MODELS FOR TREE BALANCE

Given that we can calculate a measure of balance for a phylogenetic tree, what do we do with this information? Balance was first discussed with reference to the information content of a taxonomy (e.g., Mayr 1969; Sackin 1972), but the recent surge of interest in balance reflects instead the realization that it can be used to study the evolutionary process. Such study involves seeking biological (evolutionary or other) explanations for patterns in balance. A useful baseline for this sort of enterprise is provided by *null models* for tree balance: that is, simple models for which we can predict the distribution of tree topologies, and against which we can evaluate samples of estimated trees or the results of more sophisticated evolutionary models.

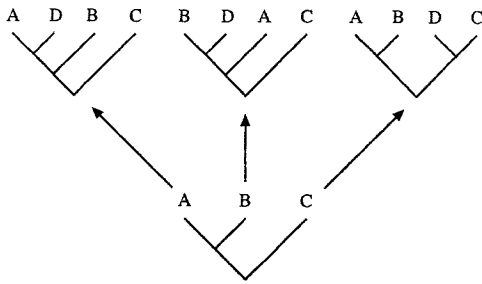Three rather different null models have

FIGURE 3.   THREE POSSIBLE WAYS A 3-TIPPED TREE CAN GIVE RISE TO A 4-TIPPED TREE THROUGH SPECIATION OF ONE LINEAGE.

Note that there are only two different topologies; the two trees on the left are identical if unlabeled, but different if labeled.

been proposed for tree balance: the equal-rates-Markov model, the proportional-to-distinguishable-arrangements model, and the equiprobable-types model. No null model is completely free of assumptions and different (even if minimal) sets of assumptions yield different null models.

The equal-rates-Markov (often just "Markov") model is attributed to Harding (1971) and to Cavalli-Sforza and Edwards (1967), but is based on models of the diversification process that date at least to Yule (1924). Under this model, trees are generated by an evolutionary process under which, at any time t, all extant lineages have equal probabilities (denoted λ) of speciating in the interval $(t, t + \Delta t)$. "Equal rates" here refers to the equality of speciation probabilities across lineages; speciation rates are often considered constant through time as well, but this is not necessary provided that when they vary, they vary together in all lineages. To see how the Markov model works, consider a four-tipped tree; it must be the result of three speciation events (of which only the last can produce more than one topology). This last speciation event can happen with equal probability to three species (see Figure 3); two of these events give an imbalanced topology and one a balanced one, so the probabilities of the imbalanced and balanced topologies are $\frac{2}{3}$ and $\frac{1}{3}$, respectively. Because the Markov model explicitly considers a simple model of evolutionary diversification,

it is an appropriate null model for studies in which one is interested in balance for evolutionary reasons. Analytical treatments of the imbalance index I_c under the Markov model are available (Heard 1992; Rogers 1994).

Notice that the Markov model does not explicitly include extinction. This is normally justified by considering the modeled (or calculated) speciation rate as the "net rate of diversification," the actual rate of speciation minus the rate of extinction. If extinction is random across lineages it does not affect tree balance (Slowinski and Guyer 1989), and so a model with net diversification is sufficient. For this reason we and others often use "diversification" and "speciation" as if they were equivalent. We caution, however, that this equivalency does not hold if branch length information is taken into account (see the section on estimating branch length, p 41). Also, because the two rates are considered together, their effects are conflated; while patterns in extinction rates could play as important a role in shaping tree balance as patterns in speciation rates, we cannot normally separate the two processes using balance alone.

Under the proportional-to-distinguishable-arrangements (PDA) model (Rosen 1978; originally described for area cladograms), there is no explicit model of growing trees. Instead, each possible arrangement of n taxa into a tree—that is, each possible estimate of the phylogeny of the n taxa—is equally likely. Within this larger set, the frequency of each topology is proportional to the number of distinguishable trees sharing that topology [Rogers (1994) gives recursion equations for the number of topologies for any tree size n]. For four-tipped trees, there are 12 different imbalanced trees and 3 different balanced ones (shuffle the labels on the tips in Figure 3, remembering that any clade can be rotated at the node that subtends it without producing a new distinguishable tree). The imbalanced and balanced topologies therefore have frequencies 0.8 and 0.2 respectively. The PDA model produces trees that are more imbalanced than the equal-rates Markov, and mimics the action of systematists who choose at random among all possible phylogenetic trees (all possible labeled trees; Slowinski 1990). As such, it is not of much interest from an evolu-

tionary point of view, but whenever tree estimation is imperfect we can consider the likelihood of a bias in tree balance towards the PDA expectation (Rogers 1994; Cunningham 1995).

Finally, the equiprobable-types (EPT) model (Simberloff et al. 1981) holds that each different topology for a given tree is equiprobable. For instance, there are two different four-tipped topologies (Figure 3), which under the EPT model should each occur in a sample of trees with frequency 0.5. This model produces the most balanced trees of the three. However, no credible model either of the evolutionary process or of the action of systematists leads to the EPT balance distribution. We therefore concur with Rogers (1993) that neither agreement nor disagreement with the EPT null model is interesting, and it can safely be disregarded.

PATTERNS IN TREE BALANCE

Patterns in tree balance may arise from several sources. First, all reasonable models of evolution include stochasticity; any observed pattern in balance can therefore reflect purely stochastic effects. Second, because estimated trees are not always identical to the underlying true trees, patterns in balance might reflect properties of the estimation algorithms or decisions of systematists in the definition and inclusion of taxa. Third, and most interesting, patterns in balance reflect the macroevolutionary processes through which the diversity of life on Earth has been generated. The task of balance studies is to determine to what extent the signal (the third source of balance patterns) comes through the noise and bias (the first and second sources, respectively), and to draw conclusions from data appropriately filtered or statistically controlled to leave only the evolutionary signal. We discuss first patterns in the available data and possible biological explanations, deferring consideration of possible biases until the section on nonevolutionary explanations for imbalance (p 38).

Balance in Estimated Trees and Simulations

Two complementary approaches to the study of patterns in tree balance have been taken: mining the literature for estimated trees, and computer simulations. Several authors have examined the balance of estimated trees, either in large samples (e.g., Savage 1983; Heard 1992) or one tree at a time (e.g., Guyer and Slowinski 1993), and compared them with the expectations based on null models. Others have produced simulated trees under a variety of simple evolutionary models. The two approaches can then be combined in hopes of understanding which evolutionary models are, and are not, consistent with data from the literature.

Imbalance has been examined for five major compilations of estimated trees. Savage (1983) collected over 1,000 estimated trees of sizes 4 to 7 (although many smaller trees were subsets of the larger ones) and found that these largely fit the Markov model (trees with 4, 6, and 7 tips, but not those with 5 tips). Savage interpreted this result as indicating that among lower taxa "lineage diversification . . . within a group . . . is largely stochastic" (Savage 1983:236), although he did not expect his result to hold for higher taxa (roughly, above the family level).

In contrast, the consensus of four more recent studies is that even small estimated trees are consistently and significantly more imbalanced than expected under the Markov null model (Figure 4): Guyer and Slowinski (1991) examined 120 complete trees of 5 species; Guyer and Slowinski (1993) examined 30 trees of 100 to 20,000 species; Heard (1992) examined 196 trees of 4 to 14 tips; and Mooers (1995) examined 39 complete trees of 8 to 14 tips. Each study rejected the Markov model, although none could account for the disagreement with Savage (1983). The rejection of the Markov model was very robust: it held for trees of species or of higher taxa, for phenetic or cladistic trees, and for complete and incomplete trees (further discussion of tree completeness below). If this "excess imbalance" can be taken at face value, it documents the existence of substantial variation in speciation and/or extinction rates among lineages within even quite small phylogenetic trees.

Rejection of the Markov model tells us only that lineages vary in their speciation/extinction rates, but we would obviously like to know how much and what kind of variation would be needed to produce trees like those in the literature. Computer simulations have recently been applied to this problem, although the
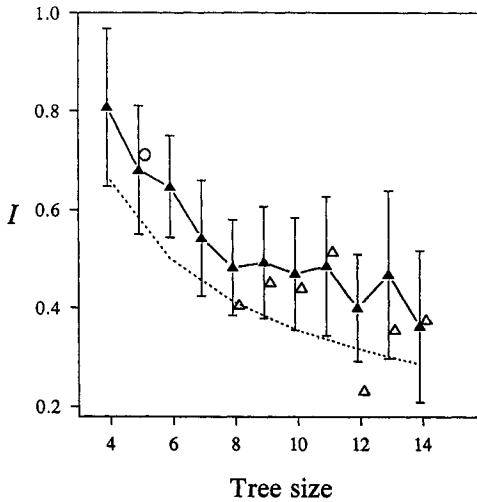
FIGURE 4.   A SUMMARY OF DATA ON THE BALANCE
OF ESTIMATED PHYLOGENIES.

Solid triangles: Heard 1992 (error bars are ±
two standard errors); open triangles: Mooers 1995;
open circle: Guyer and Slowinski 1991; dashed line
is the Markov expectation. The figure includes all
available compilations measuring balance for sys-
tematic samples of literature trees, except Savage
(1983), which we excluded because his small trees
are subsets of larger trees, and therefore the tree
sizes are not independent samples.

evolutionary models underlying them are not
yet very sophisticated. Heard (1996) consid-
ered a model where diversification rates de-
pend on the value of a trait (such as body size)
evolving in a random walk, and are also subject
to random variation. Rate variation related to
heritable traits ("heritable" at the level of an-
cestor to daughter species), and some forms of
rate variation unrelated to anything heritable,
produced trees with increased imbalance.
However, extremely large amounts of rate
variation were required to produce trees as im-
balanced as the estimated trees discussed
above. Kirkpatrick and Slatkin (1993) reached
a similar conclusion from a related model of
evolving speciation rates. Heard (1996) also
found that gradual and punctuated models of
macroevolution produce different patterns in
tree balance. Specifically, under gradualism
there appears to be an upper limit to imbal-
ance no matter how fast the traits on which
diversification rates are based evolve; no such

limit exists for punctuated evolution. This
contrast could potentially help to illuminate
the relative roles of these two major modes of
macroevolution, but more detailed and so-
phisticated evolutionary models will have to
be explored first.

Losos and Adler (1995) modeled a very dif-
ferent form of speciation-rate variation; one
that arises because a newly formed species passes
through a "refractory period" during which it
cannot speciate further—perhaps as it ex-
pands in density and range from a small
founder population. Short refractory periods
tend to produce trees more balanced than the
Markov expectation, because they dampen
the activity of the most actively speciating lin-
eages. However, Rogers (1996) has recently
shown that very long (but perhaps unreason-
ably long) refractory periods can produce
strongly imbalanced trees.

Nonevolutionary Explanations for Imbalance

We should not interpret patterns in tree bal-
ance solely in terms of evolutionary process.
We must also consider the possibility that the
trees we make give biased or inaccurate esti-
mates of the balance of the true trees underly-
ing them (this is not an issue for patterns in
computer simulations). Several possible sources
of such bias have been identified. Three in
particular have seen considerable discussion:
bias in poorly supported trees, bias in system-
atic methodology, and bias in incomplete
trees. Recent work indicates that at least the
first two are actually closely related.

Poorly Supported Trees. Several studies have
examined the effects of tree accuracy. Esti-
mated trees are not always identical to the
(usually unknown) true trees. Several studies
have also examined how estimation errors
might affect tree balance. Work with simu-
lated data suggests that different methods
make different kinds of estimation errors, with
different consequences for balance.

Mooers et al. (1995) found by simulation
(for trees with 10 tips) that incorrect maxi-
mum parsimony estimates tended to be more
imbalanced than the underlying true trees,
and Zarkikh and Li (1993) documented con-
ditions under which balanced 5-tipped trees
were incorrectly estimated as imbalanced, us-
ing maximum parsimony. Heijerman (1992,

1993) undertook a large simulation study involving trees with 20 tips, and found that generally, under a Markov model of tree production, balanced estimates were closer to the underlying true tree than imbalanced estimates. The parameters that Heijerman chose, however, made the true trees very hard to recover, so his results are hard to interpret. Interestingly, Heijerman found no relationship between tree balance and accuracy of reconstruction for an equilibrium model of tree production; this may serve as a caution that different underlying models may lead to different biases.

Related work by Huelsenbeck and Kirkpatrick (1996) shows that as data deteriorates (as sequence data accumulates very large numbers of site changes and corresponding tree estimates contain more errors), maximum parsimony phylogenies become increasingly imbalanced; simulations by Colless (1995) using artificially evolved data with very few characters per taxon show similar results. In the extreme, when data are entirely random (so there is no true tree to estimate), cladistic methods produce random guesses from the set of all possible trees (Heard and Mooers 1996). The expectation under the PDA null model (defined in the section on null models for tree balance, p 35) is strongly imbalanced compared to both the Markov model and collections of estimated trees. For maximum parsimony methods, therefore, increasing amounts of error in tree estimation produce increasingly imbalanced trees until, with random data, imbalance reaches that of the PDA null model. Combined models with Markov and PDA components can be developed in an attempt to accommodate contributions from both models, although a difficulty arises in distinguishing imbalance generated by macroevolution from imbalance generated by systematic error.

Phenetic methods treat poor data differently. Huelsenbeck and Kirkpatrick (1996) found that as data deteriorated, UPGMA (unweighted pair-group method using arithmetic averages) estimates did not change much in balance, despite containing roughly as many errors as cladistic estimates. Colless (1995) found quite balanced phenetic trees with random data (and also with very unreliable real data sets; Heard and Mooers 1996). It appears that UPGMA estimates with random data may be biased (Heard and Mooers 1996; see also Colless 1996).

Documenting these effects with trees estimated from real data is more difficult, as there is no good method for measuring how well an estimated tree reflects the (unknown) true tree. There are many indices of how well data agree within a single tree: for instance, consistency and retention indices (Farris 1989), homoplasy-slope ratios (Meier et al. 1991), decisiveness ratios (Goloboff 1991), skew in the distribution of tree lengths (Le Quesne 1989), and permutation tests (tree support values; Källersjö et al. 1992) or resampling measures (bootstrap values; Felsenstein 1985a) of agreement. We might consider these as surrogate measures of reliability, although they are often poorly intercorrelated (Goloboff 1991; Mooers et al. 1995; Sanderson and Donoghue 1989) and so we warn that recognition of reliable and unreliable trees may often be difficult.

There have been a few studies of the balance effects of estimation errors in real data sets. Guyer and Slowinski (1991) found that maximum parsimony trees with more characters defining interior nodes were more balanced than trees with fewer characters defining nodes, and Mooers et al. (1995) found a negative correlation between measures of character agreement and tree balance. Colless (1995) examined a set of unreliable (Heard and Mooers 1996) data matrices and found that cladistic estimated trees were highly imbalanced while phenetic estimates were quite balanced. The empirical results therefore support the expectations from simulation studies, despite being hampered by problems in measuring tree accuracy.

Systematic Methodology. The question of bias in tree shape arising from systematic methodology has been contentious. Recent discussion has centered on possible differences (in balance) between phenetic (almost always UPGMA) and maximum parsimony trees. Colless (1982) suggested that published cladograms were more imbalanced than phenograms, but produced only anecdotal evidence. Savage (1983) suggested that phenograms were more imbalanced than cladograms, but recognized that the differences were of uncertain statistical significance.

More recently, Colless (1995) claimed that cladograms and phenograms produced from the same data differ dramatically in balance (the cladograms being more imbalanced). In contrast, Heard (1992) and Mooers (1995) analysed large and systematic compilations of literature trees, and found no differences in balance between cladistic and phenetic trees after controlling for tree size and the inclusion of outgroups. This apparent contradiction can be resolved (Heard and Mooers 1996) by considering tree reliability: nearly all data sets used by Colless (1995) produced poorly supported trees, for which we have seen that phenetic and cladistic methods make different errors. For robust trees, simulations have shown little disagreement between methods (see Huelsenbeck and Kirkpatrick 1996), which lines up well with Heard's (1992) and Mooers' (1995) results. However, Huelsenbeck and Kirkpatrick (1996; their Figure 2) found that even with very good data, both methods (and other phylogenetic methods such as maximum likelihood) seem to produce estimated trees which are slightly biased towards imbalance.

Incomplete Trees. Mooers (1995) has examined the issue of tree completeness. The problem arises because we are interested in the balance of a real tree, but often we actually measure the balance of a subset of this tree: the sample of taxa available, or of interest, to the systematist. While balance is unaffected by random omission (or extinction) of taxa from a tree (Slowinski and Guyer 1989), this need not be true for nonrandom omission of taxa. Brown's (1994) argument that the frequencies of tree shapes are "independent of the extinction process and even of the process by which a scientist selects the species to be studied" (p 79), appears in fact to apply only to random extinction or selection on equal-rates Markov trees.

One can imagine many ways in which the taxa under study could be a nonrandom subset of all possible taxa. Some extant taxa may be unknown or some specimens unavailable to a systematist. A particularly common case is when a systematist chooses to include only representative taxa (for instance, one from each species group, genus or family). Two studies of large sets of trees (Heard 1992;

Mooers 1995) have found that trees that are incomplete because they have genera or families as tips do not differ in balance from trees of species. Mooers (1995) has shown by simulation that some, but not all, other plausible kinds of sampling (for instance, selection or exclusion of taxa that have similar values for heritable traits) do affect balance, often leading to incomplete trees that are more imbalanced than their corresponding complete trees.

Mooers (1995) also found a significant difference in balance between a set of complete trees assembled from the literature (hollow triangles in Figure 4), and a sample of incomplete trees from Heard's (1992) compilation: the complete trees were more balanced, though still not as balanced (as a group) as the Markov expectation. The opposite effect is also possible: in some large phylogenies, smaller subclades may be omitted, and this can produce a more balanced incomplete tree (Fusco and Cronk 1995).

It is possible that the difference between Mooers' (1995) and Heard's (1992) results are actually driven by tree reliability, not completeness. In making his compilation, Mooers included only datasets giving unique (or at most two) most parsimonious trees without ad hoc weighting. In contrast, Heard made no effort to assess tree quality, and so that dataset may have included some less robust trees, which we might expect to be more imbalanced. In support of this interpretation is Guyer and Slowinski's (1991) compilation of 120 complete 5-tipped trees (hollow circle in Figure 4). Guyer and Slowinski, like Heard, did not attempt to assess tree quality, and they found imbalance slightly (but not significantly) higher than Heard's compilation. It remains unclear, then, whether the differences between Mooers' and Heard's compilations (Figure 4) can be laid to completeness, tree quality, or some combination of the two.

We emphasize that the conclusions presented here are based on a distribution of true trees that can be mimicked with an equal-rates Markov model of diversification. We do not know how noise in character data sets, tree reconstruction methods, or sampling biases affect tree balances when the true trees being estimated deviate strongly from the Markov

expectation (except for the trivial cases where the true trees are all perfectly balanced or all perfectly imbalanced: error then could only be in one direction). The equal-rates Markov trees are, on average, more balanced than the PDA expectation and less balanced than the EPT expectation, and so would seem a good baseline for studying bias, but they may not properly represent the universe of true trees we ultimately want to study.

In summary, imbalance indices for estimated trees are subject to some bias. Unreliable cladistic trees tend to be more imbalanced than reliable ones, all estimated trees may err slightly towards imbalance, and incomplete trees can be more imbalanced than full trees. We do not yet know to what extent these biases account for patterns in sets of literature trees (e.g., Heard 1992; Mooers 1995; Mooers et al. 1995; see also the section on biases in tree shape, p 48), but the best evidence suggests that they are usually mild. Therefore, we do not believe that the biases are responsible for all (or even most) of the difference between estimated trees and the Markov null model.

TREE SHAPE AND BRANCH LENGTH DISTRIBUTIONS

ESTIMATING BRANCH LENGTH

Until recently, most phylogenetic trees were presented without estimates of branch lengths (see Figure 2); this gives only half the story of evolution. A time axis allows us to ask more precise questions about the variation in diversification rates among extant lineages (i.e., looking at variation within a horizontal slice through the tree in Figure 2), and also to ask about variation in rates as a whole clade diversifies (i.e., variation along the vertical axis of the tree in Figure 2). In either case, by comparing observed patterns with expectations from null models, we may gain insight into the processes driving diversity patterns.

Inferring branch lengths on a tree requires either some knowledge of (or guess at) the rates of change in the characters used to construct the tree, or recourse to fossil evidence. Cladistic trees inferred from morphological data and parsimony algorithms contain limited information about distances between successive nodes. Character changes can be "mapped" onto a topology, such that each branch is given a length equal to the number of changes inferred to occur on it. However, there need be no unique way of mapping characters onto most parsimonious trees (Swofford and Maddison 1987), and so other information is needed to choose among alternative mappings. With few models for rates of change in different morphological characters available, inferred morphological change is rarely used to estimate branch lengths on phylogenetic trees (Ross 1974; Felsenstein 1985b), and this extends to phenograms made from morphological data. However, if nodal dates can be inferred from geological data, morphology-based cladograms and phenograms can be used to make inferences about macroevolution. Cloutier (1991) offers an interesting example based on a cladogram of the coelacanth and related fossil Actinistian fishes.

Branch length data from molecular phylogenies are routinely based on some model of molecular evolution that assumes clock-like behavior (for early evidence of rate constancy in molecular evolution, see Zuckercandl and Pauling 1965; for a review of the theoretical basis for clocks, see Kimura 1983; Chao and Carr 1993; see also Gillespie 1991). These algorithms include the common UPGMA method (Sokal and Michener 1958), and the KITCH and the DNAMLK algorithms in PHYLIP (Felsenstein 1993; who also offers a concise explanation of these methods). The assumption of a molecular clock means that inferred branch lengths can be read as time intervals; i.e., a time axis can be placed on the tree (see Figure 2). If the branch length data are reliable, various comparisons can be made between these inferred trees and expectations based on null models [see, however, Felsenstein (1988:542), Nei et al. (1985), Hasegawa et al. (1993), and the section below on biases in tree shape (p 48) for discussions of the errors associated with inferred branch lengths, which can often be substantial].

TREE SHAPE MODELS WHICH INCLUDE TIME

The equal-rates-Markov null model of tree shape is also the starting point for tests that incorporate branch-length information. However, the actual values of the speciation rate [λ, or birth rate, estimated by b, the probability that a lineage will split between time t and

$(t + \Delta t)$] and the extinction rate [μ, or death rate, estimated by d, the probability that a lineage will go extinct between time t and $(t + \Delta t)$] now play prominent roles. The simplest case is one which sets μ to zero, and models λ as unchanging through time (Yule 1924), which sets a probability distribution for the time to speciation for any extant lineage as (Hey 1992):

$$P(t) = \lambda e^{-\lambda t} . \tag{3}$$

For n extant lineages, the probability distribution for the time until the next speciation event will be the same with parameter $n\lambda$:

$$P(t)_n = n\lambda e^{-n\lambda t} . \tag{4}$$

The waiting times between speciation events in the tree decrease as n increases, because there are more lineages that are free to speciate. Whereas the equal-rates-Markov model (see the earlier section on null models for tree balance, p 35) stipulates that at any one time, all extant lineages have the same λ, this "constant-rate Markov" model goes one step further, specifying that λ does not change through time. In fact, the speciation parameter λ is specified in many simulation studies of tree topology (e.g., Kirkpatrick and Slatkin 1993; Heard 1992, 1996; Losos and Adler 1995; Mooers 1995; Rogers 1996), but the resulting branch lengths are not analysed further. Below, we consider some applications.

The addition of an extinction parameter (μ) expands the versatility of the models, allowing two classes of processes to be considered. With λ greater than μ, the model is one of nonequilibrium growth. An equilibrium model, on the other hand, begins with a tree of a certain size, perhaps produced by way of a nonequilibrium model, and then regulates λ and μ so that the tree is maintained at this size (except perhaps for stochastic fluctuations). More complex models can specify different relationships between speciation and extinction that produce different expectations of tree shape—for instance, they may allow the tree to first grow and then shrink through time.

In order to compare the shapes of phylogenetic trees with expectations from the various models, the phylogenies are often represented as lineages-through-time plots (Nee et al. 1992; Figure 5). These are semilogarithmic plots of the number of lineages in existence as a function of time since the first branching of the tree at its root. Figure 5 highlights the various parameters that are depicted in such a plot. First, under the simplest Markov model with a constant and positive net rate of diversification [$(\lambda - \mu) > 0$; for simplicity, we refer to this also as a constant-rate Markov model, and rate is understood to be net], there will be exponential increase in the actual number of lineages, resulting in a straight line, with stochastic wiggles, on a semilog plot (Harvey et al. 1994; top line in Figure 5). Second, if there has been extinction ($\mu > 0$), there will be a difference between the actual number of lineages present at any one time and the number that are estimated (the bottom line in Figure 5): the phylogeny derived from extant taxa can only recover lineages that left descendants to the present day. At the end-point (the present) the two lines converge on the actual number of lineages in existence. Over most of the history of an expanding clade, the slope of the lineages-over-time plot is an estimate of ($\lambda - \mu$) (Nee et al. 1992). [If there is no error in the branch length data, the slope is in fact the realized rate of diversification ($b - d$) (Purvis 1996)]. If there is no extinction, then the actual and reconstructed lines will be coincident, and the slope estimates λ. Computer packages are available, which allow estimation of these various parameters (Harvey et al. 1995b; Rambaut and Harvey 1996).

The lineages-through-time plots make explicit several other features of the constant-rate Markov model. First, the model produces a steep slope of actual lineage production early on in the diversification of a clade. This "push of the past" (Nee et al. 1994b) is due to the fact that we confine ourselves to clades that have left at least one living descendant (in fact, we study only clades that have left many descendants): even when λ is higher than μ, some entire groups will become extinct before we sample the tree, owing to chance alone. This means we tend to observe the nonrandom sample of clades that got off to a flying start; such clades should show an increased number of fossil species early on in their history. Second, because the actual and reconstructed numbers of species must coincide at
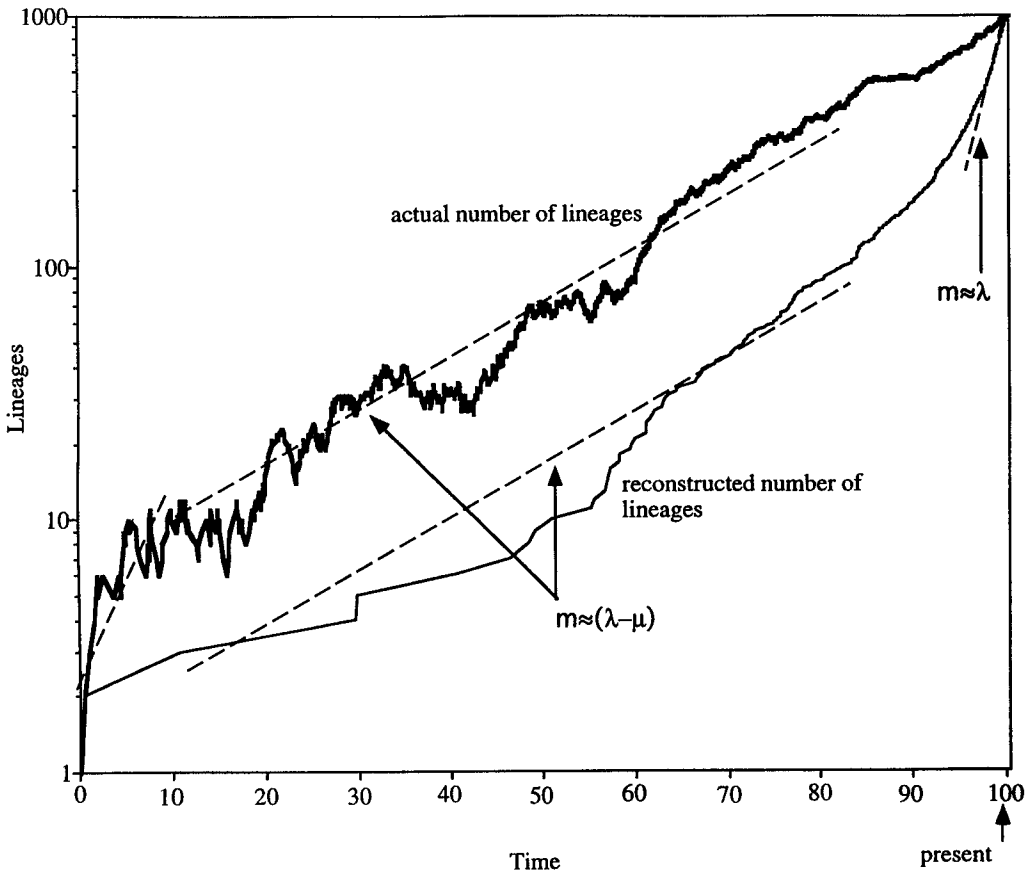
FIGURE 5. LINEAGES-THROUGH-TIME PLOTS FOR A SIMULATED PHYLOGENY.

A constant-rate Markov process has given rise to a phylogeny, depicted here as a semilogarithmic plot of the number of extant lineages as function of time since the first split in the tree at its root. The top dashed line is the actual number of lineages, and the bottom dashed line is the reconstructed number (the top line minus extinct lineages). The slope of the two lines in the mid time range estimates $\lambda - \mu$. Near the present, the bottom line steepens, and estimates λ. Time units are arbitrary. The simulation was done using the Bi-De simulation package (Rambaut and Harvey 1996) with $\lambda = 0.3$ and $\mu = 0.25$.

the present day and the reconstructed line must lie below the actual line when $d > 0$, the reconstructed curve will show an upswing at the present, and the slope of the reconstructed curve near the present (its tangent at the present) will be b. This is owing to the inevitable time lag between the birth of a lineage and its possible death; close to the present we sample many lineages that will become extinct, but haven't yet done so. Under the constant-rate Markov model, we can therefore measure both $(b - d)$, the slope of the middle part of the plot, and of b, which means that

d, the extinction rate, can also be estimated. Maximum likelihood techniques can be applied for any given tree (Nee et al. 1994a,b, 1995a).

APPROACHES USING THE NULL MODELS

Hey (1992) presented two new approaches to the study of phylogenetic tree shape using branch lengths. The first is a time-based version of the balance tests considered in the section on tree balance (p 34) for the Markov model, and is referred to as a Links test. It combines the probabilities that successive specia-

tion events occur along the same lineage over all the speciation events in a given tree, and is sensitive to both an excess of speciation events along one lineage (leading to imbalanced topologies) and the sort of repulsion of successive events envisioned by Losos and Adler (1995). Hey found that the test was much more sensitive to the former pattern than to the latter. Harvey et al. (1991) presented another version of this test that considered only the differences in splitting probabilities between nested sets of sister lineages. Mooers et al. (1994; following a suggestion made by A Burt) carried this test further, and asked whether, over a whole tree, the propensity to split was heritable among lineages: Did lineages that split quickly produce daughter lineages that also split quickly? All these tests use the time axis to ask questions about the variation in diversification rates among extant lineages, as do the topology tests presented in the section on tree balance (p 34).

Hey's (1992) second approach compares the time intervals between successive branching events (the distribution of t's in Equations 3 and 4) with the expectation based on two simple theoretical models. The first is the constant-rate Markov model with no extinction. The maximum likelihood parameter for the speciation rate (the value of b which maximizes the fit between the observed times t on a tree and those generated by Equation 4) and the actual likelihood of the data arising under the simple one-parameter process can be calculated directly. Hey suggests that this likelihood be compared with that in which the speciation rate is allowed to vary freely for each of the $(n-2)$ internodes. The likelihood ratio for this comparison is distributed as a chisquare, allowing quick assessment of the probability of a particular tree having arisen under the simple model versus the more complex scenario.

The second model considered by Hey is a special case of the equilibrium model, where clades remain at a constant size through time. A monophyletic group of n species first arises under an unspecified process. Then, for every speciation event, governed by some diversification parameter, there is a simultaneous extinction event somewhere in the tree, such that the phylogeny is maintained with n spe-

cies. This might model the situation where niche space is completely saturated by a clade. The expected internode times are less easily derived than those from Equation 4, but both the most likely value for the replacement parameter governing the process and the likelihood of observing a given tree if the most likely value is fit can be calculated directly. Again, we can compare the fit with that from a model where a new value is calculated for each internode in a given tree, using the chisquare distribution.

Note that the formal rejection of these simple models does not on its own imply the rejection of a stochastic process, but simply a rejection of those processes that include only one speciation or replacement parameter when compared with a more general model. Hey (1992) found that neither model could be rejected for any of eight trees taken from the literature, suggesting that simple one-parameter models were sufficient. For each of the eight cases, the simple growth model offered a better fit to the data than did the second, constant replacement model. Hey interpreted this as evidence against extinction in the eight clades.

Harvey et al. (1991; see also Harvey and Nee 1994a) also compared time intervals between successive branching events on a phylogenetic tree, presenting a lineages-through-time plot of the early portion of the infamous tree of the birds (Sibley and Ahlquist 1990). Rather than a straight line, as expected under the equal-rates Markov model with constant $(\lambda - \mu)$ (Figure 5), Harvey et al. (1991) reported a distinct downward curvature, evidence of a deficit of new lineages as one moved towards the tips. Nee et al. (1992) were able to statistically reject the simplest constant-rate model, but not a two-parameter model with declining diversification rates. Nee et al. (1992) interpreted their model as consistent with a scenario (outlined in a paleontological context by Valentine 1973; and modeled by Sepkoski 1978) of initial rapid diversification of birds into empty adaptive zones (aerial carnivores, aerial granivores, shorebirds), followed by a more stately process of speciation and extinction within these zones. The same pattern was found for much more recently diverged groups of birds by Zink and Slowinski (1995). They found that

9 of 10 phylogenies of recent North American bird genera also showed a downward curvature in lineages-through-time plots. Zink and Slowinski (1995) suggested that the deficit of new lineages was owing to increased extinction rates during the Pleistocene. Purvis et al. (1995) tested various scenarios with a tree of the primates, and found that the shape of the lineages-through-time plot of the tree was consistent with a constant-rate Markov model.

Lineages-through-time plots can also supply information about mass extinction events. Harvey and Nee (1993) reported that certain forms of severe mass extinctions might leave a signature in the form of an S-shaped curve, and this was formalized by Harvey et al. (1994). Kubo and Iwasa (1995) point out that while mass extinction events will be very hard to detect in real data sets (a point made by Harvey and Nee 1993), sudden jumps in speciation rates will leave a measurable footprint on the lineages-through-time plot. Kubo and Iwasa reasoned that the two processes might be linked (a mass extinction allowing a subsequent radiation), and so evidence for the latter may suggest looking for independent evidence for the former.

INTERPRETATION OF BRANCH-LENGTH DATA

Care must be taken in interpreting lineages-through-time plots. Nee et al. (1994b) highlighted how a model of a smooth decline in global splitting rate over time (Strathmann and Slatkin 1983), whereby a clade first grows and then shrinks, could be interpreted as a sporadic history of boom and stasis when converted to a lineages-through-time plot of extant lineages. Similarly, a tree with a large number of nodes near the present, which would correspond to a steep upswing in the curve near the present, appears to suggest an increased rate of diversification; however, a constant rate of net diversification with a high extinction rate is another simple explanation [if $b \gg (b - d)$, then d must be large; Nee et al. 1994a]. Nee et al. (1994b) also pointed out that when only a random subset of extant lineages is sampled under a nonequilibrium model, the resulting plot could mirror the decreasing per-lineage diversification rate scenario attributed to the birds. Extant lineages not present in a sample cannot contribute to
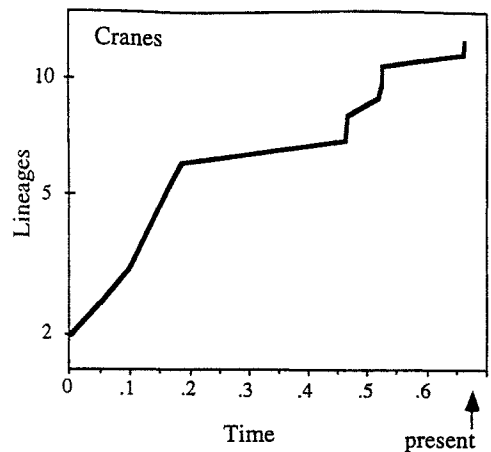

FIGURE 6. RECONSTRUCTED LINEAGES-THROUGH-TIME PLOT FOR THE CRANES.

The curve is S-shaped, suggesting a decreasing rate of diversification through time. Data from Krajewski (1990).

nodes in the tree near the present, causing a deficit of new nodes and a flattening of the curve near the present. A tree of HIV-1 lineages (Harvey et al. 1995a) showed just this sampling effect, while a tree of poliovirus 1, also made from a subsample of extant lineages, showed an S-shaped curve, consistent with a mass-extinction event.

Figures 6 and 7 offer examples of a related problem. Recall that Hey (1992) had no reason to reject a simple growth model (having a single parameter b), in favor of his general model for any of eight clades. However, Hey's general model [with $(n - 2)$ parameters] is much more complex than a scenario of a single b and a single d (e.g., Harvey et al. 1994); the many extra degrees of freedom used by Hey for his general model demanded that it offer a much better fit to the data before it could be preferred over the simpler scenario. Fairly simple models that include extinction might offer better fits to these trees than that supplied by Equations 3 and 4. The plots in Figures 6 and 7 are of two of the clades considered by Hey. The Crane phylogeny (Figure 6) seems compatible with various models of density-dependent cladogenesis (i.e., a changing rate of diversification), while the *Plethodon* phylogeny (Figure 7) is consistent with a scenario
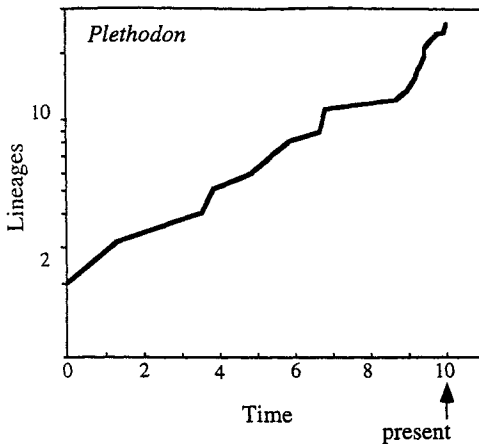
FIGURE 7.   RECONSTRUCTED LINEAGES-THROUGH-
            TIME PLOT FOR *PLETHODON*
            SALAMANDERS.

The steep rise near the present can be interpreted as evidence for a high rate of extinction in this group. Time is measured in arbitrary units (adapted from Nee et al. 1994a).

of a constant but high extinction rate and was presented as an example of such a process (Nee et al. 1994a). The plot is also superficially similar to the expectation from a more sophisticated equilibrium model of diversification [equivalent to Hey's (1992) replacement model], adapted from coalescent theory to describe endemic population processes (Nee et al. 1995a,b). However, such a model for replacement of species might be hard to justify; there is ample evidence in nature that niche space may be either undersaturated (e.g., Cornell and Lawton 1992; Ricklefs and Schluter 1993) or oversaturated (e.g., Sale 1978). These examples serve to emphasize that the model to be tested should be based on knowledge of the clade at hand; a population of endemic virus strains and a group of salamander species may yield similar phylogenetic patterns for very different reasons. Many models can be rejected by (or fit) the same data.

Kubo and Iwasa (1995) expand on this theme. They point out that many functions of b and d may produce the same phylogenetic pattern under a simple Markov model. As well, they show that while b may accurately estimate the true λ, d will be a poor measure of the

true extinction rate, even when the presumed model is true, owing to the stochasticity in the realized diversification process (see also Nee et al. 1994a).

Purvis (1996) has even suggested using the fossil record to check on the plausibility of various models. Lineages-through-time plots allow the total number of extinct species to be estimated: if more extinct species are known than predicted, we would do well to reconsider our model.

In summary, lineages-over-time plots offer a powerful approach to testing null models of macroevolution, having been successfully applied to both large trees (e.g., of the primates and of the birds) as well as small trees, such as that of *Plethodon* salamanders. As molecular data continue to accumulate, so will the opportunities for tests of various models. However, the fit for any one model should be seen as the starting point for further research, not as an explanation of the data.

TREE SHAPE AND
KEY INNOVATION HYPOTHESES

We turn now to a specific application of the study of tree shape: the search for the causes of radiations. One approach, that of testing individual phylogenies for agreement with a simple Markov model, can be extended to particular clades within a phylogeny. For instance, under the Markov model, at any single node, all possible partitions of n species into subgroups of size r and $(r - n)$ (all vectors of progeny numbers) are equally likely (Farris 1976). This property means that sister clades of the same size are no more likely than sister clades of greatly different sizes. More importantly, individual clades can be identified that are bigger than expected under the null model (i.e., a split of 41 species into sister clades of 1 and 40 species would only occur 5% of the time under the Markov model). Unfortunately, the flat probability density function for splits means that tests for single nodes are of very low power.

Nee et al. (1992) presented a generalized case for an arbitrary number of ancestor lineages when a time axis can be added to the tree. If we draw a horizontal slice through a tree at a single point in time, thereby defining n ancestral lineages, we can model the distri-

bution of the sizes of the clades that result from these ancestral lineages at any time later in the tree by randomly and repeatedly breaking a stick with the same length as the number of descendent species into n fragments. This allows us to identify subclades that are significantly larger than we would expect, a possible first step when looking for evidence of key innovations. Another, related, approach is to estimate speciation and extinction rates separately for different clades, and then to ask whether the estimates differ among clades. Purvis et al. (1995) applied both of these tests to the tree of the primates having 203 species (Purvis 1995) and were able to identify clades with higher than average diversification parameters (e.g., the Old World monkeys), as well as groups with very different ecologies but similar diversification parameters (e.g., the hominoids and the New World monkeys). This result is particularly interesting, given that there was no evidence of a change in the average diversification rate across the entire tree over time (Purvis et al. 1995), and serves to show how the two axes of variation that we have been considering (the one through time and the other across lineages) may be independent.

Sanderson and Donoghue (1994) have extended the search for differences in diversification rates among contemporary lineages to take into account uncertainty in the timings of branching events. They found a change in the rate of diversification near the base of the angiosperm tree, but most likely occurring *after* the origination of angiosperms, suggesting that characters that unite all angiosperms may not be sufficient to explain the clade's taxonomic success. This approach— used by Hodges and Arnold (1995) to associate nectar spurs with plant diversification— may prove useful when dealing with unique radiations, where statistical tests for correlates of diversification are inappropriate (Purvis 1996).

Slowinski and Guyer (1993) present a method for using the property of the flat distribution of sister-clade sizes to study correlates of radiations across many taxa. Their test combines the individually weak tests of differences in sister-clade sizes across many taxa using Fisher's combined probability test, each prob-

ability coming from a comparison in clade size between a group exhibiting a putative key innovation and its sister-group (for an example, see Hodges and Arnold 1995). Sanderson and Donoghue (1996) survey a range of techniques available for pin-pointing changes in diversification rates. Barraclough et al. (1996) presented a novel analysis, which correlates higher rates of angiosperm diversification with higher rates of genetic change early in that diversification. This analysis was based in part on the equal-rates Markov null model. However, if it turns out that the Markov null model is generally a poor predictor of actual tree shape, as it seems to be (see the section on patterns in tree balance, p 37), then tests that do not compare trees with the equal-rate Markov null model might be preferred (see the section on sophisticated evolutionary models, p 49). Nee et al. (1996; see also Barraclough et al. 1995; Mooers and Møller 1996) offer an alternative test that does not assume a Markov null expectation.

OTHER USES FOR TREE SHAPE

The shapes of phylogenetic trees are a rich resource for many areas of study not fully explored in this review. One such area is in technical aspects of tree estimation: the shape of a true tree appears to affect the possibility of estimating it correctly. This is the inverse of the problem of bias discussed in the section on patterns in tree balance (p 00). This issue has been explored through simulations by Rohlf et al. (1990) and Heijerman (1992, 1993), and a biological investigation of the relationship between shape and ease of estimation has been made by Arnold (1990). All these studies suggest that the shape of the true tree can strongly affect the probability of recovering it with different estimation techniques.

Another area is the comparison of particular trees with one another. The connection between the purely topological null models presented by Simberloff et al. (1981) and the time-based Markov null model of diversification (Page 1991; Brown 1994) allows branch length information to be added to data from topologies in comparisons of shapes of particular trees. Page (1990) has pointed out that both studies of biogeography and studies of host-parasite cospeciation ask whether trees of

different taxa are similar, and both make use of the same null models of diversification. This insight has led to a common approach to these questions, drawing directly on concepts of tree shape (Page 1993a,c, 1996; Paterson et al. 1993).

Tree shape can also be applied to the study of population processes. Lineages-through-time plots can in theory be used as an aid to reconstructing the population history of a species. This approach was pioneered by Nee et al. (1994b, 1995b), and has been applied to data from virus populations (Harvey et al. 1995a; Holmes et al. 1995; Zanotto et al. 1996; see also Holmes and Garnett 1994; Harvey and Nee 1994b), and to data from whales (Nee et al. 1995b). Clough et al. (1996) used tree shape to discriminate between possible replication mechanisms in "populations" of retrotransposons within mammalian genomes.

Finally, cell lineage diagrams are close relatives of phylogenetic trees, and their shapes may be of considerable interest. Such trees show the history of cells dividing and dying, and the parallel to speciation and extinction of lineages is obvious. Recent progress in developmental biology means that cell lineage data are rapidly improving; a complete cell lineage tree is available for *Caenorhabditis elegans* (Sulston et al. 1988), and smaller trees (showing descendants of particular cells) are being worked out in organisms such as leeches (e.g., Shankland and Martindale 1992), zebrafish (e.g., Eisen 1992), and *Drosophila* (e.g., Goodman and Doe 1993). Here tree shape relates to variation in the numbers of descendent cells traceable to different mother cells (e.g., Sternberg et al. 1992). The balance measures $\overline{N}$ and σ_N^2 have important and straightforward biological implications for these trees, because they represent mean and variance in numbers of cell divisions per "adult" cell. Balanced cell lineage trees minimize $\overline{N}$ (and σ_N^2) and perhaps, therefore, the total exposure of the genome to replication error; this may be a good developmental strategy in tissues where replication error is particularly damaging (e.g., in the germ line). In contrast, imbalanced trees might be a good developmental strategy where change through such error is advantageous (perhaps where antibodies are produced, or in a parasite where surface antigens are presented to the host), or where error has more serious consequences for certain cells within the cell lineage. We are not aware of any serious exploration of the use of tree shape in this context, although it would certainly seem worth pursuing.

WHERE NEXT FOR TREE SHAPE?

Although the past decade has seen a flurry of activity in studies of phylogenetic tree shape, we are not close to our ultimate goal: being able to account for at least some features of the diversity of life on Earth. This lack of answers does not, however, mean a lack of progress. We are beginning to have a much better idea of what the interesting questions are, and how one might look for convincing data to address them. We here consider some potentially fruitful avenues for future work on tree shape.

BIASES IN TREE SHAPE

Despite recent advances (see the section on patterns in tree balance, p 37), we do not have a thorough enough understanding of the various biases that affect the shape of estimated trees. This is important, because we need to know how much of the deviation of estimated trees from the Markov model needs macroevolutionary explanation and how much can be laid to methodological problems. In a nutshell, it seems that for well-supported trees the choice of estimation algorithms is unimportant, but unreliable maximum parsimony trees and possibly incomplete trees will be more imbalanced than the underlying true and complete trees. But how much, for what levels of reliability, and what kinds of incompleteness? Existing compilations of complete trees (Mooers 1995; Guyer and Slowinski 1991) are few, and the assessment of tree reliability is currently very difficult. So although in theory we know how to produce a perfect data set for the study of tree balance (use only complete and well-supported trees), no such data set has yet been compiled. It is also possible that reliable and unreliable trees are being sampled from different underlying distributions of true trees; the results to date assume that all the underlying trees can be modeled as equal-rate Markov trees.

We also do not know how robust estimates

of diversification parameters (see the section on tree shape models which include time, p 41) are to error in branch-length estimation. The entire endeavor rests on the reality of a molecular clock, which may not be very real (Gillespie 1991). More molecular data, more realistic models of molecular change (Felsenstein 1993; Yang 1994), and more calibrations with fossil data are needed in order to increase our confidence in the trees we use. Random error in molecular data, even when the model is correct and a clock is in effect, must surely affect distributions of branch lengths as it does topologies (cf. Huelsenbeck and Kirkpatrick 1996). Incorporating branch-length error estimates (Nei et al. 1985; Hasegawa et al. 1993) into the tests of various macroevolutionary models is one possible avenue (Nee et al. 1995b). We do not yet know whether branch length errors will only make it more difficult to choose among alternative models (since many more may fit the poorer data) or will, more seriously, produce biased parameter estimates.

SOPHISTICATED EVOLUTIONARY MODELS

The Markov model has been well studied, but we are only just beginning to understand patterns in tree shape expected from more sophisticated evolutionary models (Heard 1996; Kirkpatrick and Slatkin 1993; Losos and Adler 1995; Rogers 1996). Such an understanding is essential if we are to say more about patterns in estimated trees than merely that they do not fit the null model. Realistic shape behavior from a particular detailed model could lead to hypotheses (but not conclusions) about factors affecting speciation and extinction rates. Such models should consider both ecological (Heard and Hauser 1995) and biogeographical (Jablonski et al. 1983; Page 1991) influences on diversification. For instance, the effects of natural selection, ecological opportunity, or mass extinction have not yet been incorporated into models of tree balance, and the study of their effects on branch lengths is still in its infancy (Harvey et al. 1994). As the evolutionary models we use improve, studies of tree shape may have important things to tell us about macroevolution; for instance, patterns in tree balance may help distinguish punctuated equilibria from gradualism (Heard 1996).

Furthermore, the continued poor fit of the equal-rates-Markov model to data sets of estimated trees suggests that the model should not be used indiscriminately. We have alluded to this with reference to the studies of bias, as well as to the study of key innovation hypotheses. This caution may also be extended to all other areas of comparative biology that use the Markov null model to generate null distributions of tree shapes (e.g., Losos 1994; Martins 1996; Sanderson and Donoghue 1994).

TREE SHAPE, ECOLOGY, AND BIOGEOGRAPHY

We should soon be in a position to use patterns in sets of estimated trees to test hypotheses about ecology, biogeography, and speciation. For instance, predictions about tree balance could be drawn from Jablonski et al.'s (1983) onshore-offshore hypothesis. This hypothesis holds that for marine invertebrates, origination rates are highest in deep-water environments. If deep-water members of a clade speciate more often than shallow-water ones, imbalanced phylogenetic trees should result. Are phylogenies of marine invertebrates more imbalanced than those of groups lacking this regional bias in speciation? We do not know, but intriguingly Burlando (1990) reported that *taxonomies* of marine organisms were more imbalanced than those of terrestrial ones [but see Minelli et al. (1991:93); furthermore, we caution that some terrestrial groups may show patterns analogous to the onshore-offshore hypothesis (DiMichele and Aronson 1992)].

Other biogeographical contrasts are possible. For instance, consider a lowland species whose range is sequentially divided by the uplifting of multiple, parallel mountain ranges (or a species expanding along a linear island chain). Allopatric speciation would then be expected to produce an imbalanced tree. Extending this argument, recolonization after glaciation on continents with north/south versus east/west mountain chains should produce different tree balance patterns. For a marine ancestor with populations isolated by postglacial isostatic rebound in similar-aged lakes along a coastline [e.g., the threespine stickleback *Gasterosteus aculeatus* complex (Schluter and McPhail 1992)], we might have yet another expectation: if the ancestral popu-

lation gave rise to many descendants at the same time, the history of the separate lineages is not one of consecutive bifurcations, and the PDA model may be the best representation. The same prediction could be made for areas of high relief that are affected by cooling periods during Milankovitch cycles: high levels of allopatry owing to range contraction should produce diversification among taxa, and the resulting trees may be indistinguishable from the PDA expectation (since no tree-like process of diversification actually occurred). Vrba (1992) has presented a series of such hypotheses concerning the effect of climate change on the tempo and mode of diversification, and several of these lead to tree shape predictions: her turnover-pulse hypothesis, for example, predicts that diversification should be rapid, and hence reconstructed branches will be many and short during periods of climate change. Comparisons of tree balance among groups with different biogeographical histories can be used to test various models of biogeographically-driven allopatric speciation.

CONCLUSION

Nothing is more fundamental to biology than the search for the processes responsible for the diversity of life on Earth. Historically, the fossil record offered the main evidence of the workings of these processes; Darwin decried the niggardliness with which it offered up that account. However, the processes that generated the diversity of life on Earth have etched their signal not only in the fossil strata but also in the shapes of phylogenetic trees of extant organisms. If the estimated trees that we produce give a clear view of this signal, we would do well to look that way. More complete and more accurate trees are clearly needed, as is theoretical exploration of more sophisticated evolutionary models.

REFERENCES

Anderson S. 1974. Patterns of faunal evolution. *Quarterly Review of Biology* 49:311–332.

Arnold E N. 1990. Why do morphological phylogenies vary in quality? An investigation based on the comparative history of lizard clades. *Proceedings of the Royal Society of London, B* 240:135–172.

Ashlock P D. 1974. The uses of cladistics. *Annual Review of Ecology and Systematics* 5:81–99.

Barraclough T G, Harvey P, Nee S. 1995. Sexual selection and taxonomic diversity in passerine birds. *Proceedings of the Royal Society of London, B* 259:211–215.

Barraclough T G, Harvey P H, Nee S. 1996. Rate of rbcL gene sequence evolution and species diversity in flowing plants (angiosperms). *Proceedings of the Royal Society of London, B* 263:589–591.

Brown J K M. 1994. Probabilities of evolutionary trees. *Systematic Biology* 43:78–91.

Burlando B. 1990. The fractal dimension of taxonomic systems. *Journal of Theoretical Biology* 146:99–114.

Cavalli-Sforza L L, Edwards A W F. 1967. Phylogenetic analysis: models and estimation procedures. *Evolution* 21:550–570.

Chamberlin J C. 1924. The hollow curve of distribution. *American Naturalist* 58:350–374.

Chao L, Carr D E. 1993. The molecular clock and the relationship between population size and generation time. *Evolution* 47:688–690.

Clough J E, Foster J A, Barnett M, Wichman H A. 1996. Computer simulation of transposable element evolution: random template and strict master models. *Journal of Molecular Evolution* 42:52–58.

Cloutier R. 1991. Patterns, trends, and rates of evolution within the Actinistia. *Environmental Biology of Fishes* 32:23–58.

Colless D H. 1982. Review of: *Phylogenetics: The Theory and Practice of Phylogenetic Systematics. Systematic Zoology* 31:100–104.

Colless D H. 1995. Relative symmetry of cladograms and phenograms: an experimental study. *Systematic Biology* 44:102–108.

Colless D H. 1996. A further note on symmetry of taxonomic trees. *Systematic Biology* 45:385–390.

Cornell H V, Lawton J H. 1992. Species interactions, local and regional processes, and limits to the richness of ecological communities: a theoretical perspective. *Journal of Animal Ecology* 61:1–12.

Cronk Q C B. 1989. Measurement of biological and historical influences on plant classifications. *Taxon* 38:357–370.

Cunningham S A. 1995. Problems with null models in the study of phylogenetic radiation. *Evolution* 49:1292–1294.

Darwin C. 1859. *The Origin of Species.* London: Penguin Books. [Reprinted 1968].

Dial K P, Marzluff J M. 1989. Nonrandom diversification within taxonomic assemblages. *Systematic Zoology* 38:26–37.

DiMichele W A, Aronson R B. 1992. The Pennsylvanian-Permian vegetational transition: a terrestrial analogue to the onshore-offshore hypothesis. *Evolution* 46:807–824.

Eisen J S. 1992. Development of motoneural identity in the zebrafish. In M Shankland, E R Macagno, editors. *Determinants of Neuronal Identity.* San Diego (CA): Academic Press. pp 469–496.

Evans G E. 1975. *The Life of Beetles.* New York: Hafner Press.

Farris J S. 1976. Expected asymmetry of phylogenetic trees. *Systematic Zoology* 25:196–198.

Farris J S. 1989. The retention index and the rescaled consistency index. *Cladistics* 5:417–419.

Felsenstein J. 1985a. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791.

Felsenstein J. 1985b. Phylogenies and the comparative method. *American Naturalist* 125:1–15.

Felsenstein J. 1988. Phylogenies from molecular sequences: inference and reliability. *Annual Review of Genetics* 22:521–565.

Felsenstein J. 1993. PHYLIP (phylogeny inference package). Version 3.5c. Distributed by the author.

Fisher R A, Corbet A S, Williams C B. 1943. The relationship between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology* 12:42–58.

Furnas G W. 1984. The generation of random, binary, unordered trees. *Journal of Classification* 1:187–233.

Fusco G, Cronk Q C B. 1995. A new method for evaluating the shape of large phylogenies. *Journal of Theoretical Biology* 175:235–243.

Gillespie J H. 1991. *The Causes of Molecular Evolution.* Oxford: Oxford University Press.

Goloboff P A. 1991. Homoplasy and the choice among cladograms. *Cladistics* 7:215–232.

Goodman C S, Doe C Q. 1993. Embryonic development of the *Drosophila* central nervous system. In M Bate, A M Arias, editors. *The Development of Drosophila.* Volume 2. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. pp 1131–1206.

Gould S J, Raup D M, Sepkowski J J, Schopf T J M, Simberloff D S. 1977. The shape of evolution: a comparison of real and random clades. *Paleobiology* 3:23–40.

Guyer C, Slowinski J B. 1991. Comparisons of observed phylogenetic topologies with null expectations among three monophyletic lineages. *Evolution* 45:340–350.

Guyer C, Slowinski J B. 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, Holmes E C, Mooers A Ø, Nee S. 1995a. Inferring evolutionary process from molecular phylogenies. In R W Scotland et al., editors. *Systematics Association Special Volume 52: Models in Phylogeny Reconstruction.* London: Clarendon Press. pp 312–333.

Harvey P H, May R M, Nee S. 1994. Phylogenies without fossils. *Evolution* 48:523–529.

Harvey P H, Nee S. 1993. New uses for new phylogenies. *European Review* 1:11–19.

Harvey P H, Nee S. 1994a. Comparing real with expected patterns from molecular phylogenies. In P Eggleton, R I Vane-Wright, editors. *Phylogenetics and Ecology.* London: Academic Press. pp 219–231.

Harvey P H, Nee S. 1994b. Phylogenetic epidemiology lives. *Trends in Ecology and Evolution* 9:361–363.

Harvey P H, Nee S, Mooers A Ø, Partridge L. 1991. These hierarchical views of life: phylogenies and metapopulations. In R J Berry et al., editors. *Genes in Ecology.* Oxford: Blackwell Scientific. pp 123–137.

Harvey P H, Rambaut A, Nee S. 1995b. New computer packages for analysing phylogenetic tree structure. In M E Hochberg et al., editors. *Genesis and Maintenance of Biodiversity.* Oxford: Oxford University Press. pp 60–68.

Hasegawa M, Di Rienzo A, Kocher T D, Wilson A C. 1993. Towards a more accurate time scale for the human mitochondrial tree. *Journal of Molecular Evolution* 37:347–354.

Heard S B. 1992. Patterns in tree balance among cladistic, phenetic, and randomly generated phylogenetic trees. *Evolution* 46:1818–1826.

Heard S B. 1996. Patterns in phylogenetic tree balance with variable and evolving speciation rates. *Evolution* 50:2141–2148.

Heard S B, Hauser D L. 1995. Key evolutionary innovations and their ecological mechanisms. *Historical Biology* 10:151–173.

Heard S B, Mooers A Ø. 1996. Imperfect information and the balance of cladograms and phenograms. *Systematic Biology* 45:115–118.

Heijerman T. 1992. Adequacy of numerical taxonomic methods: a comparative study based on simulation experiments. *Zeitschrift für Zoologische und Systematische Evolutionsforschungen* 30:1–20.

Heijerman T. 1993. Adequacy of numerical taxonomic methods: further experiments using simulated data. *Zeitschrift für Zoologische und Systematische Evolutionsforschungen* 31:81–97.

Hennig W. 1966. *Phylogenetic Systematics.* Urbana (IL): University of Illinois Press.

Hey J. 1992. Using phylogenetic trees to study speciation and extinction. *Evolution* 46:627–640.

Hodges S A, Arnold M L. 1995. Spurring plant diversification: are floral nectar spurs a key innovation? *Proceedings of the Royal Society of London, B* 262:343–348.

Hoelzer G A, Melnick D J. 1994. Patterns of speciation and limits to phylogenetic resolution. *Trends in Ecology and Evolution* 9:104–107.

Holmes E C, Garnett G P. 1994. Genes, trees, and infections: molecular evidence in epidemiology. *Trends in Ecology and Evolution* 9:256–260.

Holmes E C, Nee S, Rambaut A, Garnett G P, Harvey P H. 1995. Revealing the history of infectious disease epidemics through phylogenetic trees. *Philosophical Transactions of the Royal Society of London, B* 349:33–40.

Huelsenbeck J P, Kirkpatrick M. 1996. Do phylogenetic methods produce trees with biased shapes? *Evolution* 50:1418–1424.

Jablonski D, Sepkoski J J, Jr, Bottjer D J, Sheehan P M. 1983. Onshore-offshore patterns in the evolution of Phanerozoic shelf communities. *Science* 222:1123–1125.

Kimura M. 1983. *The Neutral Theory of Molecular Evolution.* Cambridge: Cambridge University Press.

Kirkpatrick M, Slatkin M. 1993. Searching for evolutionary pattern in the shape of a phylogenetic tree. *Evolution* 47:1171–1181.

Kochmer J P, Wagner R H. 1988. Why are there so many kinds of passerine birds? Because they are small. A reply to Raikow. *Systematic Zoology* 37:68–69.

Krajewski C. 1990. Relative rates of single-copy DNA evolution in Cranes. *Molecular Biology and Evolution* 7:65–73.

Kubo T, Iwasa Y. 1995. Inferring the rates of branching and extinction from molecular phylogenies. *Evolution* 49:694–704.

Källersjö M, Farris J S, Kluge A G, Bult C. 1992. Skewness and permutation. *Cladistics* 8:275–287.

LeQuesne W J. 1989. Frequency distributions of lengths of possible networks from a data matrix. *Cladistics* 5:395–407.

Losos J B. 1994. An approach to the analysis of comparative data when a phylogeny is unavailable or incomplete. *Systematic Biology* 43:117–123.

Losos J B, Adler F R. 1995. Stumped by trees? A generalized null model for patterns of organismal diversity. *American Naturalist* 145:329–342.

MacArthur R H. 1957. On the relative abundances of birds. *Proceedings of the National Academy of Sciences, USA* 43:293–295.

Maddison W P. 1990. A method for testing the correlated evolution of two binary characters: are gains or losses concentrated on certain branches of an evolutionary tree? *Evolution* 44:539–557.

Martins E P. 1996. Conducting phylogenetic comparative analyses when the phylogeny is not known. *Evolution* 50:12–22.

Mayr E. 1969. *Principles of Systematic Zoology.* New York: McGraw-Hill.

Meier R, Kores P, Darwin S. 1991. Homoplasy slope ratio: a better measurement of observed homoplasy in cladistic analyses. *Systematic Zoology* 40:74–88.

Minelli A, Fusco G, Sartori S. 1991. Self-similarity in biological classifications. *BioSystems* 26:89–97.

Mooers A Ø. 1995. Tree balance and tree completeness. *Evolution* 49:379–384.

Mooers A Ø, Møller A P. 1996. Colonial breeding and speciation in birds. *Evolutionary Ecology* 10:375–385.

Mooers A Ø, Nee S, Harvey P H. 1994. Biological and algorithmic correlates of phenetic tree pattern. In P Eggleton, R I Vane-Wright, editors. *Phylogenetics and Ecology.* London: Academic Press. pp 232–251.

Mooers A Ø, Page R D M, Purvis A, Harvey P H. 1995. Phylogenetic noise leads to unbalanced cladistic tree reconstructions. *Systematic Biology* 44:332–342.

Nee S, Barraclough T G, Harvey P H. 1996. Temporal changes in biodiversity: detecting patterns and identifying causes. In K J Gaston, editor. *Biodiversity: A Biology of Numbers and Difference.* Oxford: Oxford University Press. pp 230–252.

Nee S, Holmes E C, May R M, Harvey P H. 1994a. Extinction rates can be estimated from molecular phylogenies. *Philosophical Transactions of the Royal Society of London, B* 344:77–82.

Nee S, Holmes E C, May R M, Harvey P H. 1995a. Estimating extinction from molecular phylogenies. In J L Lawton, R M May, editors. *Estimating Extinction Rates.* Oxford: Oxford University Press. pp 164–182.

Nee S, Holmes E C, Rambaut A, Harvey P H. 1995b. Inferring population history from molecular phylogenies. *Philosophical Transactions of the Royal Society of London, B* 349:25–31.

Nee S, May R M, Harvey P H. 1994b. The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society of London, B* 344:305–311.

Nee S, Mooers A Ø, Harvey P H. 1992. Tempo and mode of evolution revealed from molecular phylogenies. *Proceedings of the National Academy of Sciences, USA* 89:8322–8326.

Nei M, Stephens J C, Saitou N. 1985. Methods for computing the standard errors of branching points in an evolutionary tree, and their application to molecular data from humans and apes. *Molecular Biology and Evolution* 2:66–85.

Page R D M. 1990. Component analysis: a valiant failure? *Cladistics* 6:119–136.

Page R D M. 1991. Random dendrograms and null hypotheses in cladistic biogeography. *Systematic Zoology* 40:54–62.

Page R D M. 1993a. Genes, organisms, and areas: the problem of multiple lineages. *Systematic Biology* 42:77–84.

Page R D M. 1993b. On describing the shape of rooted and unrooted trees. *Cladistics* 9:93–99.

Page R D M. 1993c. Parasites, phylogeny and co-speciation. *International Journal of Parasitology* 23:499–506.

Page R D M. 1996. Temporal congruence revisited: comparison of mitochondrial DNA sequence divergence in cospeciating pocket gophers and their chewing lice. *Systematic Biology* 45:151–167.

Panchen A L. 1992. *Evolution, Classification and the Nature of Biology.* Cambridge: Cambridge University Press.

Paterson A M, Gray R D, Wallis G P. 1993. Parasites, petrels, and penguins: does louse prevalence reflect phylogeny? *International Journal of Parasitology* 23:515–526.

Platnick N I, Nelson G. 1978. A method of analysis for historical biogeography. *Systematic Zoology* 27:1–16.

Purvis A. 1995. A composite estimate of primate phylogeny. *Philosophical Transactions of the Royal Society of London, B* 348:405–421.

Purvis A. 1996. Using interspecies phylogenies to test macroevolutionary hypotheses. In P H Harvey et al., editors. *New Uses for New Phylogenies.* Oxford: Oxford University Press. pp 153–168.

Purvis A, Nee S, Harvey P H. 1995. Macroevolutionary inferences from primate phylogeny. *Proceedings of the Royal Society of London, B* 260:329–333.

Raikow R J. 1986. Why are there so many kinds of passerine birds? *Systematic Zoology* 35:255–259.

Rambaut A, Grassly N C, Nee S, Harvey P H. 1996. Bi-De: an application for simulating phylogenetic processes. *Computer Applications in the Biosciences* 12:469–471.

Raup D M, Gould S J, Schopf T J M, Simberloff D S. 1973. Stochastic models of phylogeny and the evolution of diversity. *Journal of Geology* 81:525–542.

Ricklefs R E, Renner S S. 1994. Species richness within families of flowering plants. *Evolution* 48:1619–1636.

Ricklefs R E, Schluter D, editors. 1993. *Species Diversity in Ecological Communities: Historical and Geographical Perspectives.* Chicago (IL): University of Chicago Press.

Rogers J S. 1993. Response of Colless's tree imbalance to number of terminal taxa. *Systematic Biology* 42:102–105.

Rogers J S. 1994. Central moments and probability distribution of Colless's coefficient of tree imbalance. *Evolution* 48:2026–2036.

Rogers J S. 1996. Central moments and probability distributions of three measures of phylogenetic tree imbalance. *Systematic Biology* 45:99–110.

Rohlf F J, Chang W S, Sokal R R, Kim J. 1990. Accuracy of estimated phylogenies: effects of tree topology and evolutionary model. *Evolution* 44:1671–1684.

Rosen D E. 1978. Vicariant patterns and historical explanation in biogeography. *Systematic Zoology* 27:159–188.

Ross H R. 1974. *Biological Systematics.* Reading (MA): Addison-Wesley.

Sackin M J. 1972. "Good" and "bad" phenograms. *Systematic Zoology* 21:225–226.

Sale P F. 1978. Coexistence of coral reef fishes—a lottery for living space. *Environmental Biology of Fishes* 3:85–102.

Sanderson M J, Donoghue M J. 1989. Patterns of variation in levels of homoplasy. *Evolution* 43:1781–1795.

Sanderson M J, Donoghue M J. 1994. Shifts in diversification rate with the origin of the angiosperms. *Science* 264:1590–1593.

Sanderson M J, Donoghue M J. 1996. Reconstructing shifts in diversification rates on phylogenetic trees. *Trends in Ecology and Evolution* 11:15–20.

Savage H M. 1983. The shape of evolution: systematic tree topology. *Biological Journal of the Linnean Society* 20:225–244.

Schluter D, Mcphail J D. 1992. Ecological character displacement and speciation in sticklebacks. *American Naturalist* 140:85–108.

Schopf T J M. 1979. Evolving paleontological views on deterministic and stochastic approaches. *Paleobiology* 5:337–352.

Sepkoski J J, Jr. 1978. A kinetic model of Phanerozoic taxonomic diversity. I. Analysis of marine orders. *Paleobiology* 4:223–251.

Shankland M, Martindale M Q. 1992. Segmental differentiation of lineally homologous neurons in the central nervous system of the leech. In M Shankland, E R Macagno, editors. *Determinants of Neuronal Identity.* San Diego (CA): Academic Press. pp 45–72.

Shao K, Sokal R R. 1990. Tree balance. *Systematic Zoology* 39:266–276.

Sibley C G, Ahlquist J A. 1990. *Phylogeny and Classification of Birds.* New Haven (CT): Yale University Press.

Simberloff D. 1972. Models in biogeography. In T J M Schopf, editor. *Models in Paleobiology.* San Francisco (CA): Freeman, Cooper. pp 160–191.

Simberloff D. 1987. Calculating probabilities that cladograms match: a method of biogeographic inference. *Systematic Zoology* 36:175–195.

Simberloff D, Heck K L, McCoy E D, Connor E F. 1981. There have been no statistical tests of cladistic biogeographical hypotheses. In G Nelson, D E Rosen, editors. *Vicariance Biogeography: A Critique.* New York: Columbia University Press. pp 40–63.

Slowinski J B. 1990. Probabilities of n-trees under two models: a demonstration that asymmetrical interior nodes are not improbable. *Systematic Zoology* 39:89–94.

Slowinski J B, Guyer C. 1989. Testing the stochasticity of patterns of organismal diversity: an improved null model. *American Naturalist* 134:907–921.

Slowinski J B, Guyer C. 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.

Smart J S. 1972. Channel networks. *Advances in Hydroscience* 8:305–346.

Sokal R R, Michener C D. 1958. A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin* 38:1409–1438.

Stanley S M, Signor P W, III, Lidgard S, Karr A F. 1981. Natural clades differ from "random" clades: simulations and analyses. *Paleobiology* 7:115–127.

Sternberg P W, Liu K, Chamberlin H M. 1992. Specification of neuronal identity in Caenorhabditis elegans. In M Shankland, E R Macagno, editors. *Determinants of Neuronal Identity.* San Diego (CA): Academic Press. pp 1–43.

Strathmann R R, M Slatkin. 1983. The improbability of animal phyla with few species. *Paleobiology* 9:97–106.

Sulston J, Horvitz H R, Kimble J. 1988. Appendix III: cell lineage. In W B Wood, editor. *The Nematode Caenorhabditis Elegans.* Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. pp 457–489.

Swofford D L, Maddison W P. 1987. Reconstructing ancestral character states under Wagner parsimony. *Mathematical Biosciences* 87:199–230.

Tiffney B H, Mazer S J. 1995. Angiosperm growth habit, dispersal and diversification reconsidered. *Evolutionary Ecology* 9:93–117.

Valentine J W. 1973. Phanerozoic taxonomic diversity: a test of alternate models. *Science* 180:1078–1079.

Van Pelt J, Verwer R W H, Uylings H B M. 1989. Centrifugal-order distributions in binary topological trees. *Bulletin of Mathematical Biology* 51:511–536.

Vrba E S. 1992. Mammals as a key to evolutionary theory. *Journal of Mammalogy* 73:1–28.

Wagner P J, Erwin D H. 1995. Phylogenetic patterns as tests of speciation models. In D H Erwin, R Anstey, editors. *New Approaches to Studying Speciation in the Fossil Record.* New York: Columbia University Press. pp 87–122.

Walters S M. 1961. The shaping of Angiosperm taxonomy. *New Phytologist* 60:74–84.

Williams C B. 1964. *Patterns in the Balance of Nature.* London: Academic Press.

Willis J C. 1922. *Age and Area.* Cambridge: Cambridge University Press.

Wright S. 1941. Review of: *The "Age and Area" Concept Extended. Ecology* 22:345–347.

Yang Z. 1994. Comparison of models for nucleotide substitution used in maximum likelihood phylogeny estimation. *Molecular Biology and Evolution* 11:316–324.

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* 213:21–87.

Zanotto P M A, Gould E A, Fu G G, Harvey P H, Holmes E C. 1996. Population dynamics of flaviviruses revealed by molecular phylogenies. *Proceedings of the National Academy of Sciences, USA* 93:548–553.

Zharkikh A, Li W -H. 1993. Inconsistency of the maximum-parsimony method: the case of five taxa with a molecular clock. *Systematic Biology* 42:113–125.

Zink R M, Slowinski J B. 1995. Evidence from molecular systematics for decreased avian diversification in the Pleistocene. *Proceedings of the National Academy of Sciences, USA* 92:5832–5835.

Zuckercandl E, Pauling L. 1965. Evolutionary divergence and convergence in proteins. In V Bryson, H J Vogel, editors. *Evolving Genes and Proteins.* New York: Academic Press. pp 97–165.