

Power of Eight Tree Shape Statistics to Detect Nonrandom Diversification: A Comparison by Simulation of Two Models of Cladogenesis

PAUL-MICHAEL AGAPOW¹ AND ANDY PURVIS²

¹*School of Applied Statistics, University of Reading, Reading RG6 6FN, U.K.*

²*Department of Biological Sciences, Imperial College, Silwood Park, Ascot SL5 7PY, U.K.*

Abstract.—We used simulations to compare the relative power of eight statistical tests to detect imbalance in phylogenies that is too great to be ascribed to an equal-rates Markov null model. Three of these tests have never had their power assessed before. Our simulations are the first to assess performance under scenarios in which the speciation rates of various lineages can evolve independently. In one of the scenarios explored, rates depend upon the value of an evolving trait, whereas in the other the probability that a species will speciate declines with the time since it last did so. The results indicate that the relative performance of the methods depends upon how the imbalance is generated. Different types of processes lead to different imbalance signatures, i.e., different patterns of imbalance at different depths in the phylogeny, and the measures of tree shape differ in the depth of phylogeny at which they are most sensitive. Relative performance is also affected by tree size but does not appear to depend greatly upon the degree of speciation rate variation among lineages. Two of the indices (Colless's index I_c and Shao and Sokal's \bar{N}) show reasonable performance throughout, but another (Shao and Sokal's B_2) is never indicated to be a preferred method. Two tests that do not require completely resolved phylogenies, mean I' and mean I'_{10} , have reasonable power. [Imbalance; imbalance signatures; macroevolution; MeSA; phylogeny; power; simulation.]

Phylogenies are valuable troves of information concerning the trajectory of evolution, and their shapes carry the fingerprints of the processes that have formed them (Harvey et al., 1996; Purvis, 1996; Mooers and Heard, 1997). A striking instance of this fingerprinting is the presence of asymmetry within phylogenies, where some groups are markedly more speciose than their sister clades. It is commonly held that these diverse groups have intrinsic features that account for their enhanced diversification and evolutionary success (e.g., Lydeard, 1993; Sanderson and Donoghue, 1996).

There are many possible measures for assessing tree topology and asymmetry. These measures often have been used to assess whether individual phylogenies or sets of phylogenies have shapes that are compatible with a simple null model in which probabilities per unit time of diversification are constant among contemporaneous species (see Mooers and Heard, 1997, for a review); this model is commonly known as the equal-rates Markov (ERM) model. The various measures of tree shape are strongly intercorrelated but capture slightly different aspects of tree shape, which raises two issues. First, some measures may permit more powerful tests of the null model than others measures. Second, which tests are more powerful may depend

upon the way in which non-ERM trees have grown and on how large the trees are. Together, these considerations make it difficult to compare directly the results of different studies that have used different tests and to know how best to test ERM models using a given phylogeny or set of phylogenies.

In the only previous study to compare power among tests based on several different measures (Kirkpatrick and Slatkin, 1993), the power of five tree shape statistics (\bar{N} , σ_N^2 , I_c , B_1 , and B_2 ; see Table 1 for formulations and references) at three tree sizes (i.e., numbers of terminals: 10, 20, and 40 species) and four strengths of non-ERM process was measured. However, Kirkpatrick and Slatkin considered only one kind of process, generating unbalanced phylogenies by making the instantaneous rates of speciation of every pair of sister lineages differ by a constant factor. Kirkpatrick and Slatkin found B_1 to be always the most powerful of the five tests listed above; the behaviour of σ_N^2 and I_c approached this power, and \bar{N} also performed well except for the smallest trees. Kirkpatrick and Slatkin cautioned, however, that the relative performance of the tests they considered might well be different under other processes for generating imbalance. An obvious limitation of the process they simulated is that the speciation rates of lineages

TABLE 1. Imbalance metrics. \bar{N} is the mean pathlength from root to tip, where n is the number of tips and N_i is the number of internal nodes between tip i and the root of the tree. σ_N^2 measures the variance of N_i . I_c examines the $n - 1$ internal nodes, partitioning the tips that descend from them into groups of sizes r and s , where $r \geq s$. B_1 examines every subtree of size >3 , where M is the maximum number of nodes between the tips and root of this subtree. I' gives imbalance at a single node, and B is the size of the larger subtree. M and m are the maximum and minimum values possible for B . I' is corrected by $(n - 1)/n$ if n is even. $\Sigma I'$ is the sum of I' over all nodes where it is possible to calculate. I'_{10} is I' calculated for the 10 oldest nodes in a tree.

Measure	Formula	Reference
\bar{N}	$\frac{1}{n} \sum_{i=1}^n N_i$	Shao and Sokal, 1990
σ_N^2	$\frac{1}{n} \sum_{i=1}^n (N_i - \bar{N})^2$	Shao and Sokal, 1990
I_c	$\frac{2}{(n-1)(n-2)} \sum_{i=1}^{n-1} (r_i - s_i)$	Colless, 1982; corrected by Heard, 1992
B_1	$\sum_{i=1}^{n-2} \frac{1}{M_i}$	Shao and Sokal, 1990
B_2	$\sum_{i=1}^n \frac{N_i}{2^{N_i}}$	Shao and Sokal, 1990
I'	$\frac{(B-m)}{(M-m)}$	Fusco and Cronk, 1995; corrected by Purvis et al., 2002
$\Sigma I'$		this paper
Mean I'		Purvis et al., 2002
Mean I'_{10}		this paper

do not change independently but are tightly coordinated. A doubling of the speciation rate of one lineage must be accompanied by a halving in its sister's speciation rate. More recent simulations (e.g., Heard and Mooers, 2000) vary rates more plausibly depending upon characters that evolve along lineages but have not been used to compare the power of various tree shape statistics. Here, we consider two processes in which rate of speciation in a lineage evolves independently of the rates in other lineages. In the first model, rates depend upon the value of an evolving trait, and in the second model a lineage's rate declines with the time since the last speciation event it experienced. These models are obviously still caricatures of whatever processes truly generate phylogenies, but they are a step in the direction of greater biological plausibility. We also consider three tests based on a recently proposed (and hitherto unsimulated) measure of phylogeny shape (Purvis et al., 2002), two of which can be used both on incompletely resolved phylogenies and on the fully bifurcating trees required for the use of the older measures.

COMPARING TESTS OF IMBALANCE

Phylogenies were generated and their imbalance was measured by MeSA (version 1.3.2), a framework for analyzing and simulating the growth of phylogenies (Agapow and Purvis, in prep., available from <http://www.agapow.net/software/mesa/>).

We used two non-ERM processes to evolve unbalanced phylogenies. In one, speciation rates are a decreasing function of a lineage's time since last speciation (age), and in the other speciation rates depend upon values of an evolving trait. For each process, we ran simulations using two values of one of the parameters to vary the degree of imbalance produced. Parameter values were chosen to yield trees having a level of imbalance broadly similar to those observed in a recent compilation of phylogenies (Purvis and Agapow, 2002).

For age-based rates, speciation rate of each lineage is a function of its time since last speciation event, t , with both products of a speciation event being considered newborn. Specifically, we set $\lambda = At^{-0.5} + 0.5$, where A was set to 0.3 and 0.6 for weaker and stronger dependency of rates on age, respectively. MeSA uses discrete-time approximation to represent continuous processes. In this case, ages and hence rates of species were reassessed at least every 0.001 time units (they were also assessed at every speciation event). By inspection, this step size was sufficient to prevent the pathologies that can arise using discrete-time approximations. This model differs from previous ones (Losos and Adler, 1995; Chan and Moore, 1999) in which rates were age dependent in that rate increases with time after a speciation.

For trait-based rates, speciation rate of each lineage is a function of its value of a continuous character, X . The root value of X

was set to 100; X then evolved according to speciation Brownian evolution, with both lineages changing only at speciation events and changes being drawn from a normal distribution of $\mu = 0$. The strength of clade selection was varied by adjusting the rate of Brownian motion evolution. The trait-based model is largely similar to that used by Heard (1996), but here the trait evolves by Brownian rather than log-Brownian motion. This difference is not purely semantic, because both models have rates of speciation depending upon absolute trait values. Thus, our model is not merely a log-transformed version of Heard's. Although there are good reasons to expect many traits to evolve in a log-Brownian rather than Brownian fashion (Heard, 1996), it is unclear whether rates of speciation would be more likely to depend upon raw or log-transformed trait values.

Tests were carried out on fully bifurcating ERM trees of each of four clade sizes (8, 16, 32, and 64 species) spanning the size range for which most of these tests are likely to be used. (Larger trees would be more informative, but most of the tests can be used only with fully bifurcating trees, which, because each branch in the true phylogeny has a nonzero chance of being reconstructed as having zero length, tend to be small [Heard and Mooers, 1997].) We generated 5,000 eight-species trees and 1,000 trees of each of the other sizes. We measured \bar{N} , $\sigma_{\bar{N}}^2$, B_1 , B_2 , I_c , and $\Sigma I'$ (see Table 1) for each ERM tree, thus building up null distributions against which other trees could be compared. Given that real phylogenies tend to be too unbalanced for ERM rather than too balanced (e.g., Mooers and Heard, 1997), we used one-tailed tests of ERM, with imbalance as the alternative hypothesis. Observed values of measures for which high values indicate imbalance (\bar{N} , $\sigma_{\bar{N}}^2$, I_c , and $\Sigma I'$) were therefore compared with the 95th percentile of their null distributions, whereas observed values of measures where high values indicate balance (B_1 and B_2) were tested against the 5th percentile of their null distributions.

We also used a randomization test to assess whether mean I' exceeded 0.5, the expectation of a node's I' under ERM. For each tree, 4,000 randomizations were performed in which each I' value had an independent 50% probability of being replaced by $1 - I'$.

The P value for the null hypothesis that the tree was not unbalanced was then the proportion of means of the randomized data that at least equalled the observed mean I' . The randomization test does not have elevated type I error rates when applied to ERM trees (results not shown). The reason for using a randomization test rather than comparing observed trees with the distribution from ERM trees is that this test can be applied even when phylogenies are not fully resolved. If we were instead to compare observed trees with the null distribution, we should also need to know the number of nodes from which a nodal I' score could be computed. The distinction is clearest where, for the largest tree size, we also assessed the power of the mean I' test applied to only the earliest 10 nodes in each tree (mean I'_{10}). This assessment is intended to roughly indicate the power that might be expected using mean I' when the tree is incompletely resolved and the other tests can therefore not be applied.

MeSA 1.3.2 was used to simulate all trees and data sets and to compute all tree shape measures. All trees were saved onto CD-ROM (topology, branch lengths, and where applicable values of the evolved trait on which speciation probabilities depended), facilitating comparisons with any future measures. All statistical testing was performed using R 1.3.1 for Macintosh (Ihaka and Gentleman, 1996).

Table 2 shows the mean and critical values of \bar{N} , $\sigma_{\bar{N}}^2$, B_1 , B_2 , I_c , and $\Sigma I'$ under ERM for each tree size. Similar tables have appeared elsewhere for all but $\Sigma I'$, although they have been designed for use with two-tailed tests of ERM (e.g., Heard, 1992; Kirkpatrick and Slatkin, 1993). The other tables show the mean values of these measures, mean I' , and for 64-species trees the mean I'_{10} and the power of all the tests to reject ERM under age-based (Tables 3, 4) and trait-based (Tables 5, 6) scenarios.

A large number of the power values are very low, notably for eight-species trees generated under trait-based speciation rates: the power of some tests is $<5\%$, the expected type I error rate. This is unsurprising because, for a tree to give a significant result, it must be more unbalanced than 95% of ERM trees of its size. For very small trees (e.g., five species), even a totally unbalanced

TABLE 2. Null distributions. Means are rounded to three significant figures. The last figure in the critical values (CV) is rounded down for B_1 and B_2 (for which values lower than the critical value are significant) and rounded up for the other measures. These are the critical values for a one-tailed test of ERM, with imbalance being the specified alternative hypothesis.

Test	8 tips		16 tips		32 tips		64 tips	
	Mean	CV	Mean	CV	Mean	CV	Mean	CV
\bar{N}	3.44	4.00	4.76	5.56	6.13	7.13	7.49	8.58
σ_N^2	1.25	2.75	2.25	4.66	3.43	6.78	4.71	8.66
I_c	0.416	0.762	0.262	0.420	0.165	0.254	0.100	0.146
B_1	3.98	3.28	8.34	7.18	16.9	15.2	34.0	31.6
B_2	2.59	2.10	3.32	2.61	4.02	3.13	4.73	3.76
$\Sigma I'$	1.50	3.17	3.50	5.91	7.52	10.9	15.7	20.3

TABLE 3. Age-dependent rates with the weaker of the two sets of parameter values. Power is the percentage of 1,000 trees that rejected ERM at the $\alpha = 0.05$ level.

Test	8 tips		16 tips		32 tips		64 tips	
	Mean	Power	Mean	Power	Mean	Power	Mean	Power
\bar{N}	3.58	10.1	5.14	21.2	6.80	27.2	8.53	42.0
σ_N^2	1.69	10.1	3.44	20.6	5.82	30.4	8.66	38.9
I_c	0.505	10.1	0.343	23.5	0.225	28.3	0.143	42.4
B_1	3.76	6.2	7.82	18.4	15.7	33.8	31.6	47.9
B_2	2.48	10.1	3.04	20.9	3.60	22.4	4.11	27.2
$\Sigma I'$	1.95	7.5	4.53	19.1	9.94	35.2	20.4	48.6
Mean I'	0.552	6.3	0.580	17.2	0.599	32.3	0.602	49.1
Mean I'_{10}							0.618	29.4

TABLE 4. Age-dependent rates with the stronger of the two sets of parameter values. Power is the percentage of 1,000 trees that rejected ERM at the $\alpha = 0.05$ level.

Test	8 tips		16 tips		32 tips		64 tips	
	Mean	Power	Mean	Power	Mean	Power	Mean	Power
\bar{N}	3.69	16.9	5.39	34.8	7.15	43.8	9.03	60.5
σ_N^2	1.95	16.9	4.41	37.7	7.24	44.0	10.5	56.6
I_c	0.559	16.9	0.395	38.2	0.255	44.5	0.162	60.5
B_1	3.64	9.9	7.48	33.0	15.2	47.5	30.5	72.6
B_2	2.41	16.9	2.91	30.7	3.43	31.4	3.86	41.5
$\Sigma I'$	2.18	11.3	5.21	32.8	10.8	47.7	22.4	73.9
Mean I'	0.594	10.0	0.632	28.8	0.631	44.7	0.639	73.0
Mean I'_{10}							0.658	39.7

TABLE 5. Trait-dependent rates with the slower rate of trait evolution (and hence weaker selection). Power is the percentage of 1,000 trees that rejected ERM at the $\alpha = 0.05$ level.

Test	8 tips		16 tips		32 tips		64 tips	
	Mean	Power	Mean	Power	Mean	Power	Mean	Power
\bar{N}	3.54	8.9	5.08	17.9	6.83	28.6	8.59	43.4
σ_N^2	1.55	8.9	3.23	19.5	5.64	26.0	8.32	35.8
I_c	0.478	8.9	0.329	21.2	0.226	28.5	0.144	43.1
B_1	3.86	3.3	8.02	12.2	16.3	15.9	32.9	18.7
B_2	2.50	8.9	3.08	16.7	3.45	25.0	4.04	30.6
$\Sigma I'$	1.78	4.7	4.20	13.3	9.00	18.3	18.2	22.4
Mean I'	0.517	3.3	0.556	12.5	0.567	19.9	0.563	25.7
Mean I'_{10}							0.618	29.7

TABLE 6. Trait-dependent rates with the faster rate of trait evolution (and hence stronger selection). Power is the percentage of 1,000 trees that rejected ERM at the $\alpha = 0.05$ level.

Test	8 tips		16 tips		32 tips		64 tips	
	Mean	Power	Mean	Power	Mean	Power	Mean	Power
\bar{N}	3.622	12.5	5.215	25.6	6.956	36.5	8.843	55.9
σ_N^2	1.779	12.5	3.682	27.3	6.098	32.0	9.225	48.9
I_c	0.528	12.5	0.356	29.3	0.237	36.2	0.134	55.6
B_1	3.739	5.6	7.922	16.5	16.148	19.3	32.868	19.3
B_2	2.437	12.5	2.998	21.4	3.478	29.0	3.909	38.8
$\Sigma I'$	2.006	7.5	4.446	16.0	9.307	21.4	18.465	26.1
Mean I'	0.561	5.6	0.577	18.1	0.580	21.2	0.571	31.8
Mean I'_{10}							0.636	31.6

topology will not be more unbalanced than 95% of ERM trees because the most unbalanced topology is found in >5% of ERM trees. Likewise, for eight-species trees, <5% of ERM trees are actually more unbalanced than 95% of ERM trees. The low power of single trees to reject ERM, also found in earlier simulations (e.g., Kirkpatrick and Slatkin, 1993), led previous workers to analyze either collections of trees (Guyer and Slowinski, 1991; Heard, 1992; Mooers, 1995; Mooers et al., 1995; Purvis and Agapow, 2002; Stam, 2002) or single nodes from deeper within phylogenies (Guyer and Slowinski, 1993). In the first case, collection of a sufficient number of trees is eased when these trees need not be completely resolved. Of the tests considered here, only mean I' and mean I'_{10} can be used with partially resolved phylogenies. In the second case, our results support the solution first suggested by Fusco and Cronk (1995): a set of basal nodes from a single phylogeny, which does not need to be fully resolved to the level of species, can often contain enough signal of imbalance to reject ERM.

The rank ordering of the different tests in terms of power varies with tree size and, more markedly, with the process used to generate imbalance. When rates are based on the ages of species (Table 3), B_1 and the I' measures are least powerful for small trees but most powerful for larger trees. B_2 is as powerful as any measure for small trees but is easily the least powerful for the largest trees considered. The remaining measures always perform quite well. When rates depend upon trait values, B_1 and the two I' measures are the least powerful measures, with \bar{N} and I_c always among the most powerful. The difference arises because the two scenarios we simulated leave different imbalance signatures, and the various measures are most sensitive

to imbalance in different regions of the tree. When rates are based on age, imbalance is spread fairly evenly between basal and distal nodes in the tree. When rates are based on trait values, however, imbalance is concentrated toward the root of the tree. This effect is seen most clearly by comparing the values of mean I' and mean I'_{10} for the trees with 64 tips. The basal nodes are markedly more unbalanced than average nodes when speciation rates depend upon trait values (with mean I'_{10} being about as powerful as mean I'), whereas the difference is much smaller when rates depend upon ages. Mean I' and B_1 do not distinguish between basal and distal imbalance, whereas B_2 , I_c , \bar{N} , and σ_N^2 are more influenced by basal imbalance (see also Shao and Sokal, 1990). Mooers and Heard (1997) discussed this feature with respect to I_c and proposed a variant of I_c in which all nodes have equal influence, but the variant has to our knowledge never been used. Our results and the above interpretation are consistent with the findings of Kirkpatrick and Slatkin's (1993) simulations. Their process also spread imbalance evenly throughout the tree (every sister pair of lineages differed in rate by the same factor), and they too found B_1 to outperform \bar{N} , σ_N^2 , and I_c , with B_2 the least powerful. Although the relative performances of the imbalance measures depend upon tree size and the process used to generate the trees, performance is essentially unaffected by the strength of the unbalancing process, at least across the range of parameter values we have explored.

Given that our two scenarios led to different measures being most powerful, how should a given actual phylogeny be analyzed? If the aim were simply to test ERM, one approach would be to use a test that has reasonable power under both of our

scenarios and under that simulated by Kirkpatrick and Slatkin (1993). \bar{N} and I_c are the best general choices, whereas B_1 and to a lesser extent $\Sigma I'$ do poorly under one of our scenarios, even though they are the most powerful measures under the other. Another approach (Kirkpatrick and Slatkin, 1993) would be to use two tests that are most sensitive to imbalance in different parts of the tree, in which case either \bar{N} or I_c might be used in conjunction with either B_1 or $\Sigma I'$, with suitable adjustment of critical values to account for multiple tests. This sort of approach will be particularly important if real processes give qualitatively different imbalance signatures from either process simulated here (e.g., greater imbalance at the tips than near the root). Mean I' has reasonable power and is the only test that is easy to apply to nonbinary trees, for which it is therefore by default the current method of choice. Depending upon the scenario, analysis of only the oldest few nodes may give power at least comparable to that to be had from analyzing the full set. These two tests might usefully be used in conjunction. B_2 never performs well and should not be used.

The importance of imbalance location can also be seen from real phylogenies. Kirkpatrick and Slatkin (1993) used the measures they assessed on two species-level phylogenies. One of them (*Phyllobrotica*; Farrell and Mitter, 1990) is very unbalanced near the root and much more symmetric distally; only B_1 failed to reject ERM. The other phylogeny (*Ophraella*; Futuyma and McCafferty, 1990) has one quite balanced node near the root and is otherwise very unbalanced; here, B_1 was the only test to reject ERM. Mean I' shows the same qualitative pattern as B_1 (*Phyllobrotica* $P = 0.15$; *Ophraella* $P = 0.04$).

Capturing tree shape in multiple ways opens up the possibility of analyzing the imbalance signature for insights into the process generating the phylogeny, a possibility we are currently investigating. For instance, the pattern of significant results for *Phyllobrotica* seems to be more consistent with trait-based variation in diversification rates, whereas the pattern for *Ophraella* seems more consistent with age-based variation. Testing trees against models with greater biological realism than ERM may help gain understanding of some of the "rules" that govern phylogenesis.

ACKNOWLEDGMENTS

We are grateful to Nick Isaac, Rich Grenyer, David Orme, Kai Chan, Mike Charleston, Steve Heard, and an anonymous reviewer. This work was supported by the Natural Environment Research Council (U.K.) through grant GR3/11526.

REFERENCES

- CHAN, K. M. A., AND B. R. MOORE. 1999. Accounting for mode of speciation increases power and realism of tests of phylogenetic asymmetry. *Am. Nat.* 153:332–346.
- COLLESS, D. H. 1982. *Review of Phylogenetics: The Theory and Practice of Phylogenetic Systematics*, by E. O. Wiley. *Syst. Zool.* 31:100–104.
- FARRELL, B., AND C. MITTER. 1990. Phylogenesis of insect/plant interactions: Have *Phyllobrotica* leaf beetles (Chrysomelidae) and the Lamiales diversified in parallel? *Evolution* 44:1389–1403.
- FUSCO, G., AND Q. C. CRONK. 1995. A new method for evaluating the shape of large phylogenies. *J. Theor. Biol.* 175:235–243.
- FUTUYMA, D. J., AND S. S. MCCAFFERTY. 1990. Phylogeny and the evolution of host plant associations in the leaf beetle genus *Ophraella* (Coleoptera, Chrysomelidae). *Evolution* 44:1885–1913.
- GUYER, C., AND J. B. SLOWINSKI. 1991. Comparison of observed phylogenetic topologies with null expectations among three monophyletic lineages. *Evolution* 45:340–350.
- GUYER, C., AND J. B. SLOWINSKI. 1993. Adaptive radiation and the topology of large phylogenies. *Evolution* 47:253–263.
- HARVEY, P. H., A. J. LEIGH BROWN, J. MAYNARD SMITH, AND S. NEE. 1996. *New uses for new phylogenies*. Oxford Univ. Press, Oxford, U.K.
- 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., AND A. Ø. MOOERS. 2000. Phylogenetically patterned speciation rates and extinction risks change the loss of evolutionary history during extinctions. *Proc. R. Soc. Lond. B Biol. Sci.* 267:613–620.
- IHAKA, R., AND R. GENTLEMAN. 1996. "R": A language for data analysis and graphics. *J. Comput. Graph. Stat.* 5:299–314.
- KIRKPATRICK, M., AND M. SLATKIN. 1993. Searching for evolutionary pattern in the shape of a phylogenetic tree. *Evolution* 47:1171–1181.
- LOSOS, J. B., AND F. R. ADLER. 1995. Stumped by trees? A generalized null model for patterns of organismal diversity. *Am. Nat.* 145:329–342.
- LYDEARD, C. 1993. Phylogenetic analysis of species richness: Has viviparity increased the diversification of actinopterygian fishes? *Copeia* 1993:514–518.
- MOOERS, A. Ø. 1995. Tree balance and tree completeness. *Evolution* 49:379–384.
- MOOERS, A. Ø., AND S. B. HEARD. 1997. Inferring evolutionary process from phylogenetic tree shape. *Q. Rev. Biol.* 72:31–54.
- MOOERS, A. Ø., R. D. M. PAGE, A. PURVIS, AND P. H. HARVEY. 1995. Phylogenetic noise leads to

- unbalanced cladistic tree reconstructions. *Syst. Biol.* 44:332–342.
- PURVIS, A. 1996. Using interspecific phylogenies to test macroevolutionary hypotheses. Pages 153–168 *in* New uses for new phylogenies (P. H. Harvey, A. J. Leigh Brown, J. Maynard Smith, and S. Nee, eds.). Oxford Univ. Press, Oxford, U.K.
- PURVIS, A., AND P.-M. AGAPOW. 2002. Phylogeny imbalance: Taxonomic level matters. *Syst. Biol.* 51:844–854.
- PURVIS, A., A. KATZOURAKIS, AND P.-M. AGAPOW. 2002. Evaluating phylogenetic tree shape: Two modifications to Fusco and Cronk's method. *J. Theor. Biol.* 214:99–103.
- SANDERSON, M. J., AND M. J. DONOGHUE. 1996. Reconstructing shifts in diversification rates on phylogenetic trees. *Trends Ecol. Evol.* 11:15–20.
- SHAO, K., AND R. R. SOKAL. 1990. Tree balance. *Syst. Zool.* 39:266–276.
- STAM, E. 2002. Does imbalance in phylogenies reflect only bias? *Evolution* 56:1292–1295.

*First submitted 14 December 2001; revised manuscript
returned 16 May 2002; final acceptance 15 August 2002
Associate Editor: Steve Heard*