



COMMENTARIES

Uncertainty in the reconstruction of ancestral character states and limitations on the use of phylogenetic comparative methods

JONATHAN B. LOSOS

Department of Biology, Washington University, St Louis

(Received 14 January 1999; initial acceptance 24 May 1999;
final acceptance 19 July 1999; MS. number: AS-1236)

Phylogenetic approaches to the study of comparative data have a long history in the field of animal behaviour (e.g. Lorenz 1941; Greene & Burghardt 1978; Gittleman 1981). However, since the publication in 1991 of Brooks & McLennan's *Phylogeny, Ecology and Behavior* and Harvey & Pagel's *The Comparative Method in Evolutionary Biology*, there has been a huge increase in the use of phylogenetic methods in comparative studies. Indeed, a comparative study would be unpublishable in most journals today if it did not incorporate phylogenetic methods of analysis, or at least a discussion of phylogenetic relationships.

Welcome as this development is, the time has come for a critical reappraisal of phylogenetic methods for the analysis of comparative data (Ricklefs & Starck 1996; Price 1997; Cunningham et al. 1998). A variety of methods have now been published and one might wonder how much an analysis is affected by the choice of method. More generally, given that the use of phylogenetic methods now appears to be standard practice, one might also wonder whether situations exist in which phylogenetic methods are unnecessary or even inappropriate.

Reconstructing Ancestral Character States

One common use of phylogenetic methods is to reconstruct ancestral character states. These reconstructions can then be used to test a wide variety of hypotheses. Although ancestral reconstructions are now widely used, a number of methods are available for reconstructing ancestral attributes. These methods differ greatly in their assumptions and can lead to markedly different reconstructions (Swofford & Maddison 1992; Losos & Miles 1994; Martins & Hansen 1996; Omland 1997). Moreover,

when using ancestral reconstructions to test evolutionary hypotheses, one might want to have an idea of how much confidence should be placed in these estimates. Only recently have methods to do so been developed (Martins & Hansen 1997; Schluter et al. 1997).

In a pioneering study, Ryan & Rand (1999) considered the extent to which different methods of estimating ancestral states yielded different reconstructions. They then investigated experimentally whether differences in reconstructions altered the conclusions of a previous study based on only one method of reconstructing ancestral states (Ryan & Rand 1995). In this commentary, I focus on the first half of their paper: the evaluation of differences in reconstructions provided by different methods (I use Ryan & Rand's study, which is exemplary and pathbreaking in many respects, only for illustrative purposes; the problems I discuss apply to many, perhaps most, papers that reconstruct quantitative ancestral traits, including my own previous work). Ryan & Rand used two methods, squared-change and local squared-change parsimony (for discussion, see Martins and Hansen 1996, pp. 45–46), to reconstruct ancestral states of eight different parameters of frog calls (each parameter was reconstructed separately) at seven ancestral nodes (Fig. 1). In addition, for both of these methods, they used two modes of character evolution, gradual and punctuated (also termed speciation; Garland et al. 1992). Ryan & Rand also considered three other sets of reconstructions based on different phylogenies, but I only consider the reconstructions based on the preferred hypothesis of phylogenetic relationships. Comparisons of the reconstructed values produced by these different methods and assumptions revealed that, in many instances, the estimates produced for the same node differed by greater than 10%. However, as Ryan & Rand noted, the two methods they employed were fairly similar in their underlying assumptions; other methods might provide substantially more divergent reconstructions. To examine this possibility, I

Correspondence: J. B. Losos, Department of Biology, Box 1137, Washington University, St Louis, MO 63130, U.S.A. (email: losos@biology.wustl.edu).

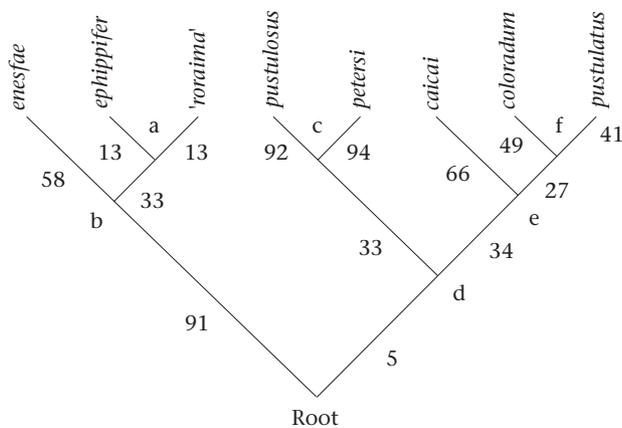


Figure 1. Phylogeny of eight species of *Physalaemus* (from Ryan & Rand 1999). Branch lengths were calculated from molecular studies (Ryan & Rand 1995). *Physalaemus 'roraima'* is an undescribed species.

reconstructed ancestral states with another method, linear parsimony, which is probably the oldest and one of the most widely used methods (Swofford & Maddison 1987; Maddison & Maddison 1992). In addition, for reconstructions produced by squared-change parsimony, I calculated confidence intervals to investigate how confident one can be in particular reconstructions.

The results illustrate the differences in reconstructions produced by different methods (Ryan & Rand used quantitatively varying traits; for a similar discussion of the parallel difficulties involving qualitative, or categorical, traits, see Cunningham et al. 1998). For illustrative purposes, I chose two of Ryan & Rand's eight variables: call duration and rise time, both of which produce large differences in ancestral reconstructions depending on the method used. I employed three methods of ancestor reconstruction: linear parsimony (LP), local squared-change parsimony with a punctuated mode of character change (LSQPP), and squared-change parsimony with a gradual mode of character change (SCPG). Table 1 indicates that each pair of methods produced reconstructions differing by more than 20% for at least one ancestral node for call duration and for at least two ancestral nodes for

rise time. Some methods differed by more than 20% in five of the seven ancestral nodes. Although character states reconstructed by LP often differed greatly from reconstructions based on other methods, there were similar differences in comparisons between the other two methods (Table 1, Fig. 2).

Figure 2 shows the reconstructions produced using the three different methods. The figure reveals not only that differences in reconstructed values for particular nodes can be substantial but also that these different reconstructions can lead to markedly different interpretations of evolutionary history. For example, values reconstructed for ancestral node b differ greatly: LSCPP produces a value intermediate between *P. caicai* and *P. enesefae*; SCPG produces a value that is also intermediate, but closer to *P. caicai*; and LP produces a reconstruction that ranges from identical to *P. roraima* to intermediate between that species and *P. caicai*. Linear parsimony often produces multiple reconstructions that are equally parsimonious (i.e. that require the same amount of change summed across all branches of the tree), although the reconstructed values for particular ancestral nodes may differ in the different reconstructions. Similarly, values reconstructed for node f differ: SCPG and LP provide reconstructions nearly identical to different species (*P. pustulatus* and *P. ephippifer*), whereas LSCPP produces a reconstruction intermediate between *P. pustulatus* and *P. coloradum*. More generally, LP reconstructs a scenario in which a major evolutionary shift occurs on the branch of the tree leading from ancestral node b to *P. enesefae* (as revealed by the large distance between node b and *P. enesefae* in Fig. 2) and in which little or no evolutionary change occurs from the root node to either node b or node d or from node d to node e (evident from the fact that these nodes are reconstructed in the same position in Fig. 2). By contrast, the other two methods produce scenarios in which a moderate amount of change occurs on many branches, but none of these changes is as large as that produced by LP. Table 2 demonstrates that when change on a branch is relatively small, SCPG and LSCPP reconstruct larger changes than LP; conversely, when change is relatively large, the reverse is true. Such differences in reconstructions could lead to substantially

Table 1. Number of ancestral node reconstructions differing by more than 20%

Call rise time	Call duration			
	Squared-change parsimony	Local squared-change parsimony	Linear parsimony	
			Minimum	Maximum
Squared-change parsimony	—	3	2	2
Local squared-change parsimony	4	—	1	1
Linear parsimony				
Minimum	2	5	—	—
Maximum	3	5	—	—

In cases in which linear parsimony reconstructed a range of equally parsimonious reconstructions, the maximum and minimum linear parsimony categories refer to the range in differences between linear parsimony reconstructions and the reconstructions produced by other methods. Differences in reconstructions for call duration are above the diagonal and differences in reconstructions for call rise time are below the diagonal.

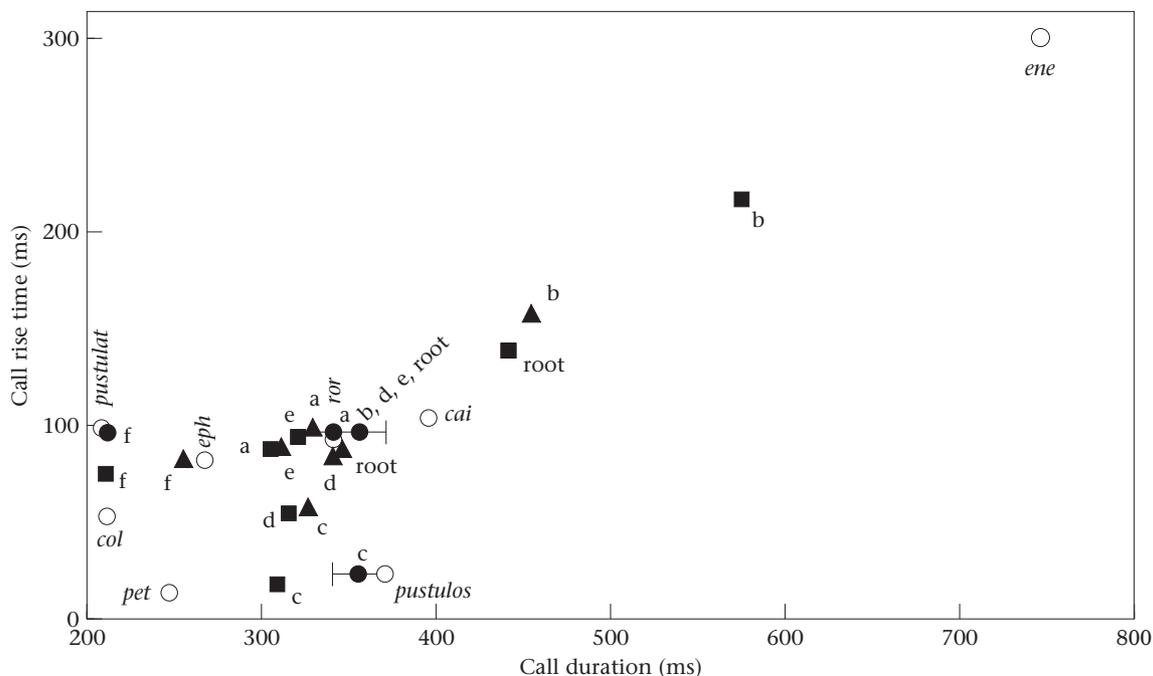


Figure 2. Ancestral reconstructions provided by three different methods (●: parsimony; ■: local squared-change parsimony; ▲: squared-change parsimony). The error bars on the linear parsimony estimates represent the range of equally parsimonious reconstructions for an ancestral node. Linear parsimony reconstructions for the root node and nodes b, d and e are identical. Extant species are indicated by an open circle. *cai*: *P. caicai*; *col*: *P. coloradum*; *ene*: *P. enesefae*; *eph*: *P. ephippifer*; *pet*: *P. petersi*; *pustulat*: *P. pustulatus*; *pustulos*: *P. pustulosus*; *ror*: *P. 'roraima'*.

different conclusions about how evolutionary change has proceeded and which processes are responsible.

The differences in the reconstructions reflect the ways in which ancestors are reconstructed. Although differing in exactly what they optimize, squared-change and local squared-change parsimony tend to spread character change throughout the tree. Some change is recon-

structed on almost all branches and relatively large changes are minimized. In these methods, mode of change also plays a role: punctuated models assume that the length of each branch of the tree (in terms of expected amounts of evolutionary change) is the same, whereas gradual modes assume that the amount of change is proportional to the length of the branch (which is often

Table 2. Changes reconstructed on each branch of the phylogeny by the different methods

Phylogeny branch (ancestor–descendant)	Call duration			Call rise time		
	Linear parsimony	Squared- change parsimony	Local squared-change parsimony	Linear parsimony	Squared- change parsimony	Local squared-change parsimony
Node f – <i>P. pustulatus</i>	–3	–49	–2	0–5	17	24
Node f – <i>P. coloradum</i>	0	–46	1	–47–42	–30	–23
Node e – <i>P. caicai</i>	25–56	84	75	5–10	17	11
Node e – node f	–161–130	–56	–112	0	–5	–18
Node c – <i>P. pustulosus</i>	0–31	44	62	0	–34	5
Node c – <i>P. petersi</i>	–124–93	–80	–62	–10	–44	–5
Node d – node c	0	–13	–6	–76–71	–27	–36
Node d – node e	0	–28	6	0	3	39
Node a – <i>P. ephippifer</i>	–73	–62	–37	–11	–16	–5
Node a – <i>P. 'roraima'</i>	0	11	36	0	–5	6
Node b – <i>P. enesefae</i>	376–407	292	178	202–207	144	86
Node b – node a	–31–0	–126	–255	–5–0	–58	–127
Root Node – node b	0	109	128	0	69	76
Root Node – node d	0	–6	–126	–5–0	–4	–85

For linear parsimony, a range of values is provided when multiple equally parsimonious reconstructions exist for one or both of the nodes involved.

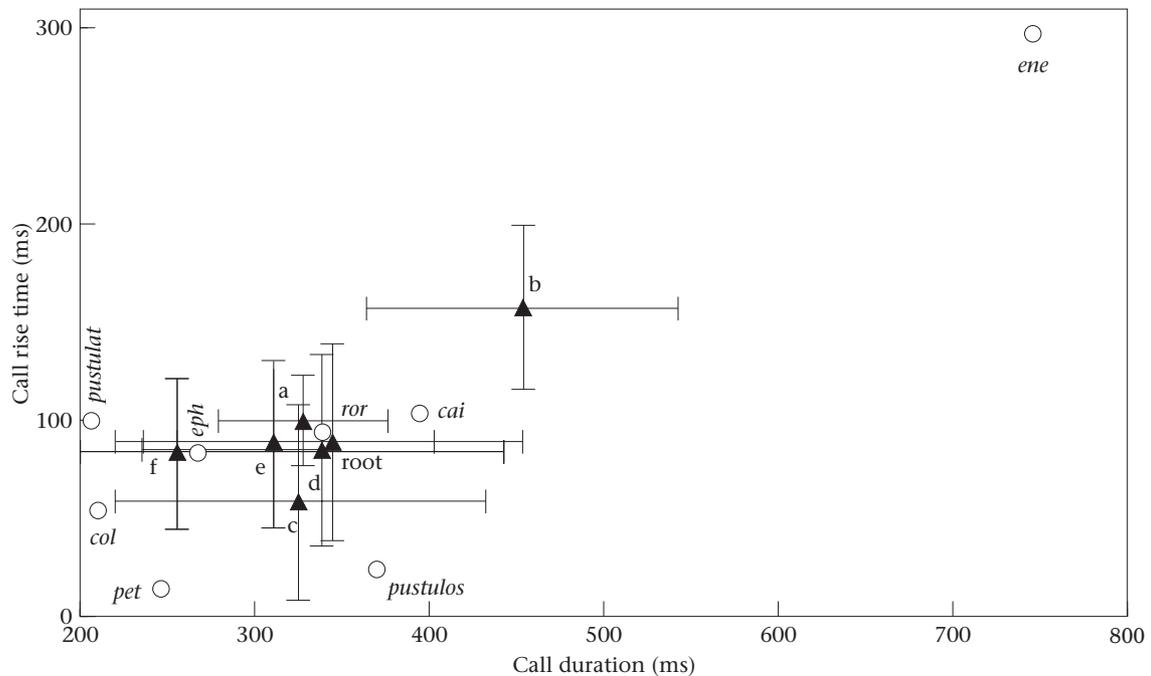


Figure 3. Confidence limits on ancestral reconstructions estimated using squared-change parsimony (\blacktriangle) with a gradual mode of evolutionary change. Extant species values represented by open circles. Species abbreviations as in Fig. 2.

equated with length of time). By contrast, linear parsimony reconstructs ancestral states to minimize the absolute amount of change summed over all branches. As a result, many branches are reconstructed with no change, whereas relatively large change occurs on a few branches. Linear parsimony also differs from the other methods in producing ambiguous reconstructions, whereas the other methods do not, and in not being able to incorporate information on branch lengths.

The results in Fig. 2 clearly illustrate that the value of reconstructed nodes can differ greatly depending on what method is used. One might think that in a particular analysis, one could choose a method based on its underlying evolutionary model. For example, squared-change parsimony methods assume some version of Brownian motion, in which the amount of change occurring on a particular branch is expected to be a function of time elapsed (although whether change occurs in a gradual or punctuated mode will make a difference). In contrast, the model of evolutionary change underlying linear parsimony seems to correspond to one of stabilizing selection with occasional peak shifts. In addition, other models of evolutionary change are conceivable (e.g. Price 1997; see Martins & Hansen 1996).

Unfortunately, data are rarely available to determine which model best describes evolutionary change in the history of a particular lineage (Gittleman & Luh 1994; Martins & Hansen 1996; Omland 1997). Consequently, a priori information about which method to use may not be available. Hence, the only option may be to use a variety of different methods. If the methods produce substantially different results, and if no basis exists for choosing among the methods based on their assumptions, then one may be left in a situation in which the

range in potential phenotypes for an ancestral node is quite large. Whether this range is so great as to make any conclusions about evolutionary pathways impossible is, of course, a testable proposition, as Ryan & Rand (1999) demonstrated. They found, at least for the range of reconstructions produced by their methods, that the conclusions of their studies were not altered qualitatively, regardless of which reconstruction was utilized.

Disregarding differences among methods, one might inquire how confident one may be in a particular reconstruction or, in statistical terms, what are the confidence limits around the estimate? Figure 3 illustrates the 95% confidence limits around the SCPG estimates calculated following Schluter et al. (1997); the method of Martins & Hansen (1997) is formally the same under the assumption of Brownian motion, but can be generalized to incorporate other assumptions (Garland et al. 1999; Martins, in press; D. Schluter, personal communication). The confidence limits are enormous; essentially, few of the ancestral reconstructions can be distinguished from each other, or from a number of the extant taxa, with any confidence. Given these wide overlaps, confidence in the actual reconstructions would appear to be quite low. Of course, as with the reconstructions themselves, these confidence limits are based on their own assumptions, the most important of which in this case are that the rate of character change conforms to a Brownian motion model and does not differ over the entire tree.

Algorithms for calculating explicit error bars do not exist for other methods of reconstructing quantitative characters. None the less, one can judge how robust an estimate is by comparing the amount of change it would require to the amount of change required to produce an alternative estimate (Schluter et al. 1997). For example,

linear parsimony reconstructs the call of ancestral node f for these two variables as being essentially identical to that of *P. pustulatus*. How confident can one be that the ancestral call at this node was not in reality identical to that of *P. coloradum*? The most parsimonious scenario requires a total amount of evolutionary change, summed across the entire phylogeny, of 356 ms in rise time and 793 ms in call duration. If, instead, ancestral states are reconstructed with the constraint that node f be reconstructed as identical to *P. coloradum*, then the scenario requiring the least amount of change implies 403 ms of change in rise time, a 13% increase, and the same amount of change in call duration. Given that only 13% change in one variable and no change in the second variable is required, one might conclude that the most parsimonious scenario is not strongly supported relative to the alternative scenario. By contrast, a scenario in which the call of node b is constrained to be identical to that of *P. enesefae* requires considerably more change: an increase in rise time of 207 ms (58%) and in call duration of 407 ms (51%). This is a substantially greater amount of change and one might conclude with greater confidence that parsimony reconstructions favour the most parsimonious scenario over this alternative.

When Should Phylogenetic Methods Be Employed?

More generally, this discussion leads to the question of whether circumstances exist in which phylogenetic approaches should not be used at all. Recent use of phylogenetic comparative methods can be traced to Felsenstein's (1985) paper in which he explicitly detailed why data for species could not be considered independent points in statistical analyses. The reason is that, because of shared ancestry, closely related species are more likely to be similar to each other than are distantly related species; data points for different species are thus not independent. This phenomenon, sometimes termed 'phylogenetic inertia' or 'phylogenetic effect', could have a number of different causes (see Harvey & Pagel 1991). A key assumption here is that, in fact, closely related species are likely to be phenotypically similar. If rates of character evolution are high, relative to rates of branching speciation, then no relationship may exist between degree of phylogenetic relationship (sometimes termed phylogenetic 'similarity' or 'distance') and phenotypic similarity. In such cases, one might expect estimates of reconstructing ancestral states to be unreliable; furthermore, the statistical argument for incorporating phylogenetic information would not hold. Indeed, given that any phylogenetic method entails some assumptions and data transformations that are difficult to test, using such methods when they are unnecessary might be inadvisable (Björklund 1997; Ackerly & Donoghue 1998).

How to determine whether the rate of character evolution is so high that one should not employ phylogenetic methods is not entirely clear. One approach is to ask whether, in fact, a relationship exists between some measure of phylogenetic similarity and phenotypic similarity. The phylogenetic autocorrelation method

(Cheverud et al. 1985; Gittleman et al. 1996; three other alternative methods have been recently proposed: Ackerly & Donoghue 1998; Diniz-Filho et al. 1998; Abouheif, in press) calculates an autocorrelation coefficient, ranging from -1 to 1 , which indicates the extent to which phenotypic similarity is related to phylogenetic similarity. When the phylogenetic correlation is very low, one might feel confident that rates of character evolution have been so high that phylogenetic relationships provide little useful information about character evolution (of course, assumptions about mode of character evolution are encoded in the this method's 'phylogenetic similarity matrix', so this method is prone to the same difficulties discussed above; also small sample sizes can be problematic with this method; Martins 1996). In fact, most of the eight variables used by Ryan & Rand have low or even negative autocorrelation coefficients, which indicates that closely related species are not necessarily phenotypically similar, which would call into question the accuracy of phylogenetic methods such as ancestor reconstruction. (The occurrence of negative autocorrelation values is actually uncommon. They could result from character displacement, in which closely related species diverge phenotypically in sympatry (Cheverud et al. 1985) or might be an artefact of sampling if investigators only choose one exemplar from clades of closely related and phenotypically similar species.) Björklund (1997) has proposed a similar approach for examining rates of evolution in qualitative character data to determine whether phylogenetic comparative approaches are warranted.

This discussion does not imply that phylogenetic information is unnecessary. Without a phylogeny, one could not assess whether phylogenetic nonindependence is likely to be an issue (Ackerly & Donoghue 1998). Hence, rather than arguing that we need not utilize phylogenetic information, this discussion simply proposes that phylogenetic information be considered at an earlier stage of analysis. If, in fact, no relationship exists between phylogenetic and phenotypic similarity, then subsequent use of phylogenetic approaches may not be warranted (Gittleman & Luh 1994). Conversely, if such a relationship does exist, then phylogenetic information should be incorporated in statistical analysis and evolutionary interpretation.

Conclusions

The last decade has seen a remarkable shift, from little attention being paid to phylogenetic information at the beginning to nearly universal attention by the end. As with many trends, the pendulum has swung too far (as Gittleman et al. 1996 previously noted) and now we are seeing critical reappraisal. Comparative biologists need to examine carefully the assumptions and implications of the particular phylogenetic methods they employ and even consider whether they are needed at all.

Acknowledgments

I thank the editors for inviting me to write this commentary, M. Ryan for providing information and a prepublication copy of his paper, D. Schluter for extensive advice

and for calculating confidence limits and J. Gittleman and T. Oakley for very useful comments on the manuscript. This work was supported by the National Science Foundation (DEB 9318642).

References

- Abouheif, E.** In press. A method for testing the assumption of phylogenetic independence in comparative data. *Evolutionary Ecology Research*.
- Ackerly, D. D. & Donoghue, M. J.** 1998. Leaf size, sapling allometry, and Corner's rules: phylogeny and correlated evolution in maples (*Acer*). *American Naturalist*, **152**, 767–791.
- Björklund, M.** 1997. Are 'comparative methods' always necessary? *Oikos*, **80**, 607–613.
- Brooks, D. R. & McLennan, D. A.** 1991. *Phylogeny, Ecology, and Behavior: a Research Program in Comparative Biology*. Chicago: University of Chicago Press.
- Cheverud, J. M., Dow, M. M. & Leutenegger, W.** 1985. The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body weight among primates. *Evolution*, **39**, 1335–1351.
- Cunningham, C. W., Omland, K. E. & Oakley, T. H.** 1998. Reconstructing ancestral character states: a critical reappraisal. *Trends in Ecology and Evolution*, **13**, 361–366.
- Diniz-Filho, J. A. F., Ramos de Santàna, C. E. & Bini, L. M.** 1998. An eigenvector method for estimating phylogenetic inertia. *Evolution*, **52**, 1247–1262.
- Felsenstein, J.** 1985. Phylogenies and the comparative method. *American Naturalist*, **125**, 1–15.
- Garland, T., Jr, Harvey, P. H. & Ives, A. R.** 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Systematic Biology*, **41**, 18–32.
- Garland, T., Jr, Midford, P. E. & Ives, A. R.** 1999. An introduction to phylogenetically based statistical methods, with a new method for confidence intervals on ancestral values. *American Zoologist*, **39**, 374–388.
- Gittleman, J. L.** 1981. The phylogeny of parental care in fishes. *Animal Behaviour*, **29**, 936–941.
- Gittleman, J. L. & Luh, H.-K.** 1994. Phylogeny, evolutionary models and comparative methods: a simulation study. In: *Phylogenetics and Ecology* (Ed. by P. Eggleton & R. Vane-Wright), pp. 103–122. London: Academic Press.
- Gittleman, J. L., Anderson, C. G., Kot, M. & Luh, H.-K.** 1996. Phylogenetic lability and rates of evolution: a comparison of behavioral, morphological and life history traits. In: *Phylogenies and the Comparative Method* (Ed. by E. P. Martins), pp. 166–205. New York: Oxford University Press.
- Greene, H. W. & Burghardt, G. M.** 1978. Behavior and phylogeny: constriction in ancient and modern snakes. *Science*, **200**, 74–77.
- Harvey, P. H. & Pagel, M. D.** 1991. *The Comparative Method in Evolutionary Biology*. Oxford: Oxford University Press.
- Lorenz, K. Z.** 1941. Vergleichende Bewegungsstudien an Anatinen. *Journal of Ornithology*, **89**, 24–32.
- Losos, J. B. & Miles, D. B.** 1994. Adaption, constraint, and the comparative method: phylogenetic issues and methods. In: *Ecological Morphology: Integrative Organismal Biology* (Ed. by P. C. Wainwright & S. M. Reilly), pp. 60–98. Chicago: University of Chicago Press.
- Maddison, W. & Maddison, D.** 1992. *MacClade 3: Interactive Analysis of Phylogeny and Character Evolution*. Sunderland, Massachusetts: Sinauer.
- Martins, E.** 1996. Phylogenies, spatial autoregression, and the comparative method: a computer simulation test. *Evolution*, **50**, 1750–1765.
- Martins, E. P.** In press. Estimation of ancestral states of continuous characters: a computer simulation study. *Systematic Biology*.
- Martins, E. P. & Hansen, T. F.** 1996. The statistical analysis of interspecific data: a review and evaluation of phylogenetic comparative methods. In: *Phylogenies and the Comparative Method* (Ed. by E. P. Martins), pp. 22–75. New York: Oxford University Press.
- Martins, E. P. & Hansen, T. F.** 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist*, **149**, 646–667.
- Omland, K. E.** 1997. Examining two standard assumptions of ancestral reconstructions: repeated loss of dichromatism in dabbling ducks (Anatini). *Evolution*, **51**, 1636–1646.
- Price, T.** 1997. Correlated evolution and independent contrasts. *Philosophical Transactions of the Royal Society of London, Series B*, **352**, 519–529.
- Ricklefs, R. E. & Starck, J. M.** 1996. Applications of phylogenetically independent contrasts: a mixed progress report. *Oikos*, **77**, 167–172.
- Ryan, M. J. & Rand, A. S.** 1995. Female responses to ancestral advertisement calls in túngara frogs. *Science*, **269**, 390–392.
- Ryan, M. J. & Rand, A. S.** 1999. Phylogenetic influence on call preferences in female túngara frogs (*Physalaemus pustulosus*). *Animal Behaviour*, **57**, 945–956.
- Schluter, D., Price, T., Mooers, A. O. & Ludwig, D.** 1997. Likelihood of ancestor states in adaptive radiation. *Evolution*, **51**, 1699–1711.
- Swofford, D. L. & Maddison, W. P.** 1987. Reconstructing ancestral character states under Wagner parsimony. *Mathematical Biosciences*, **87**, 199–229.
- Swofford, D. L. & Maddison, W. P.** 1992. Parsimony, character-state reconstructions, and evolutionary inferences. In: *Systematics, Historical Ecology, and North American Freshwater Fishes* (Ed. by R. C. Mayden), pp. 186–223. Palo Alto, California: Stanford University Press.