8

LOGICAL PROBLEMS ASSOCIATED WITH INCLUDING AND EXCLUDING CHARACTERS DURING TREE RECONSTRUCTION AND THEIR IMPLICATIONS FOR THE STUDY OF MORPHOLOGICAL CHARACTER EVOLUTION

One of the driving forces behind a renewed interest in phylogeny is the realization that a phylogenetic context is critical for the study of a wide variety of evolutionary questions (e.g., Eldredge and Cracraft 1980; Felsenstein 1985; O'Hara 1988; Donoghue 1989; Funk and Brooks 1990; Brooks and McLennan 1991; Harvey and Pagel 1991; Maddison and Maddison 1992; Eggleton and Vane-Wright 1994; Harvey et al. 1995). Of particular interest are questions about character evolution and adaptation—an area of investigation now commonly known as comparative biology. A prominent and unresolved problem in comparative biology is whether it is appropriate to include certain characters as part of the data used to reconstruct a phylogenetic tree when the evolution of those same characters is the subject of investigation. One widespread belief is that including the characters of interest is logically circular and that independent phylogenies (i.e., ones based on data other than the characters under investigation) are needed to properly analyze character evolution (e.g., Coddington 1988; Carpenter 1989; Lauder and Liem 1989; Olmstead 1989; Sytsma 1990; Brooks and McLennan 1991; McKey 1991; Vane-Wright et al. 1992; Hedges and Maxson 1996). An alternative and also widespread opinion is that the best inferences about character evolution must be based on the best estimates of phylogeny, which must in turn be based on all of the available evidence—that is, including the characters of interest (e.g., Kluge 1989; Donoghue and Sanderson 1992; Maddison and Maddison 1992; Swofford and Maddison 1992; Deleporte 1993; A. de Queiroz and Wimberger 1993; McKitrick 1993; Lee 1997). Because the characters of interest are often morphological, the issue of whether to include or exclude those characters has become entangled with another

debate about the use of morphological versus molecular data in phylogeny reconstruction (for reviews of the latter, see Hillis 1987; Patterson 1987; chapter 1 of this book). Not surprisingly, authors who favor the use of molecular data for reconstructing phylogenetic trees commonly advocate excluding the characters of interest—particularly when those characters are morphological—and optimizing them on trees reconstructed entirely from molecular data (e.g., Hedges and Maxson 1996).

In this chapter, I analyze the practices of including and excluding the characters of interest during the reconstruction of phylogenetic trees in relation to the logic of analyzing character evolution in a phylogenetic context. I explain part of the controversy about whether to include or exclude characters as a manifestation of alternative perspectives on the general nature of inferences. I argue that even within the context of a perspective that emphasizes inferential caution, both including and excluding the characters of interest can lead to logical problems, which implies that although there is a valid concern about including the characters of interest during tree reconstruction, that concern cannot be eliminated by a simple rule to always exclude those characters. In light of these conclusions, I make both general and specific methodological recommendations about including and excluding characters that can be used to increase analytical rigor in studies of character evolution. Although based largely on an earlier article (K. de Queiroz 1996), this chapter extends my earlier analysis in several ways, including considerations of how the controversy about including versus excluding the characters of interest relates to the controversy about morphological versus molecular data, to prior assessments of homology versus analogy, and to different general classes of analyses. I have also expanded and modified the discussions of logical problems and methodological recommendations, including several examples from empirical studies.

DIFFERENT GENERAL PERSPECTIVES

The controversy about whether it is preferable to include or exclude the characters of interest stems, at least in part, from alternative general perspectives concerning the nature of inferences about character evolution. The different general perspectives can be characterized as attempting to make the inference that is most likely correct versus attempting to make an inference that is unlikely to be incorrect. These two perspectives are distinct—that is, they are not merely different ways of describing the same concern. Seeking the inference or estimate that is the most likely to be correct is an inherently bold or liberal approach in that magnitude of the difference between the best and second best es-

timate (as judged by some optimality criterion) is irrelevant. The best estimate is still the best estimate regardless of how much better it is than the next best estimate. A statistical example is point estimation, in which the inference or estimate that maximizes or minimizes some quantity is taken as the best estimate regardless of the magnitude of the difference between its value and those of alternative estimates. In this context, it makes sense to include all relevant characters in the data used to reconstruct the phylogenetic tree—even the characters of interest. To the extent that the information in the characters of interest is not predominantly misleading, those characters provide evidence about the phylogenetic relationships of the taxa in which they occur and thus also about their own evolution. Excluding that information can be expected to decrease the ability of any tree-based analytical method to yield the correct inference.

In contrast, seeking an inference or estimate that is unlikely to be incorrect is an inherently cautious or conservative approach in that the magnitude of the difference between one estimate and another is crucial. An estimate that is only marginally better than an alternative is much more likely to be incorrect than one that is substantially better, and a cautious researcher may prefer to abstain from making an inference when the best estimate has more than a very low probability of being incorrect. A statistical example is significance testing, which adopts the convention of rejecting the null hypothesis only when the data have a very low probability (<0.05) if the null hypothesis is true; but by that same convention, necessitates that in many cases one will be unable to reject the null hypothesis (i.e., make an inference) even though it is false. In this context, it sometimes makes sense to exclude the characters of interest from the data used to reconstruct the phylogenetic tree (see below). To the extent that including the characters of interest influences the outcome of an analysis in one direction or another, that practice can be expected—at least in some cases—to compromise the cautious or conservative nature of the inference.

Although these alternative perspectives lead researchers to opposite conclusions about whether to include or exclude the characters of interest, the perspectives themselves are complementary rather than contradictory. There is no contradiction in recognizing that the inference that is most likely correct may still have a relatively high probability of being incorrect—that the single best estimate of a parameter is not necessarily a conservative estimate. Moreover, there is no reason to question the validity of either general perspective; both reflect legitimate goals and concerns. In the rest of this chapter, however, I will (for the most part) restrict my discussion to the general perspective that emphasizes the avoidance of incorrect inferences, because this is the context within which concerns about including the characters of interest are most relevant.

LOGICAL FALLACIES AND RELATED INFERENTIAL PROBLEMS

Among authors who advocate excluding the characters of interest, there is a widespread opinion that including those characters is tantamount to fallacious reasoning (e.g., Felsenstein 1985; Coddington 1988; Olmstead 1989; Sytsma 1990; McKey 1991; Vane-Wright et al. 1992; Sillén-Tullberg and Møller 1993; Block and Finnerty 1994; Meyer et al. 1994; Hedges and Maxson 1996). However, just because a practice can compromise an inference does not mean that it necessarily will in all cases. I next examine the issue of logical fallacies in greater detail to show that previous statements about the problem have been oversimplified. I do so by applying some general ideas about inferential logic to the specific case of character evolution, arguing that the existence of logical problems in such studies is dependent on how the assumptions inherent in specific methods or practices bear on the specific inferences or conclusions of an analysis. This perspective is implicit in several earlier papers (e.g., Maddison 1990; Harvey and Purvis 1991; Armbruster 1992, 1993; Maddison and Maddison 1992; Swofford and Maddison 1992; Deleporte 1993; A. de Queiroz and Wimberger 1993) but has not previously been presented as a detailed and explicit argument.

Arguing in a Circle

The problem of including the characters of interest has most commonly been characterized as an example of logical circularity, or at least potential circularity (e.g., Felsenstein 1985; Coddington 1988; Olmstead 1989; Sytsma 1990; McKey 1991; Armbruster 1992, 1993; Donoghue and Sanderson 1992; Vane-Wright et al. 1992; A. de Queiroz and Wimberger 1993; Sillén-Tullberg and Møller 1993; Block and Finnerty 1994; Meyer et al. 1994; Hedges and Maxson 1996). Although there are precedents for this characterization, it is also potentially misleading. In the following discussion, I adopt the terminology of the *Oxford English Dictionary* (Second Edition, Oxford University Press, 1989), which uses different and descriptive terms to distinguish between two different, though related, classes of logical fallacies.

The fallacy termed "arguing or reasoning in a circle" (also "vicious circle" or "circular reasoning") is an invalid form of argument in which a proposition is used to establish a conclusion, and afterwards the proposition is justified by means of the very conclusion that it was previously used to establish. The description of the argument as circular derives from the fact that proposition and

conclusion depend on one another in such a way that the inferential sequence has no clear starting point. A phylogenetic example would be using a tree reconstruction method that minimizes instances of homoplasy (e.g., parsimony), obtaining a result in which homoplasy is rare, and then using that result to justify use of the method that minimizes homoplasy.

The practice (proposition) in question here is that of including certain characters as part of the data used to estimate a phylogenetic tree, which then serves as the basis for inferences about the evolution of those same characters. In this case, a circular argument would be one in which a conclusion reached in an analysis that included the characters of interest was then used to justify the very practice of including those characters. This is not, however, the kind of logical problem about which comparative biologists have been concerned. In attempting to avoid circularity by excluding the characters of interest, comparative biologists have been concerned about their conclusions regarding character evolution—not about arguments or justifications for including the characters of interest. In other words, they have been concerned only with the validity of using a particular practice to arrive at a result or conclusion rather than with the full circle that then uses the result or conclusion to justify the original practice. Strictly speaking, this is not arguing in a circle (see also Maddison and Maddison 1992; Swofford and Maddison 1992).

Begging the Question

The fallacy termed "begging the question" (also "*petitio principii*") is more relevant to the issue of including the characters of interest in that it bears on the logic of establishing conclusions (i.e., as opposed to justifying initial propositions). To beg the question is to take for granted the matter in dispute or, more specifically, to assume the conclusion in the premises of the argument. A phylogenetic example would be reconstructing phylogenies using a method that explicitly or implicitly assumes evolutionary rate constancy among lineages (e.g., phenetic clustering) and then concluding, based on those phylogenies, that rates of evolution among lineages have been constant.

Strictly speaking, most analyses of character evolution that include the characters of interest during tree reconstruction do not constitute examples of begging the question. Including the characters of interest will seldom precisely entail whatever conclusion is ultimately reached concerning the evolution of those characters, though it is at least possible to contrive an example in which it does (e.g., arguing for the single origin of a particular character when that character was the only datum used to reconstruct, using parsimony, the phylogenetic tree upon which the conclusion of a single origin is based). On the other hand, less

extreme influences on the outcome of at least some types of analyses regarding character evolution are expected to occur more commonly (see below). Therefore, in the context of an approach that favors inferential caution, there is at least a potential problem with including the characters of interest. That problem is conceptually related to the fallacy of begging the question, the difference being that the premises exert an undesirable influence on the conclusions without fully assuming them.

Inappropriately Influencing the Analysis

Various practices can systematically influence an analysis so that the analysis tends to favor a particular outcome or result. If the result that tends to be favored by a particular method or practice is the same one obtained in an analysis using that method, then a logical problem exists—namely, it is ambiguous whether the result is attributable to the data per se or whether it is an artifact of the methodological influence. I call this problem "inappropriately influencing the analysis." It does not require that the practice (premise) in question fully assumes the result (conclusion) but only that it tends to favor that result by influencing the analysis in that direction. Begging the question refers to a situation in which the practice in question makes a particular result inevitable and thus can be viewed as the most extreme form of an inappropriate influence.

The example used above to illustrate the fallacy of circular reasoning includes an example of an inappropriate influence (which nonetheless constitutes only part of the full circle illustrated in that argument). If one is going to conclude that homoplasy is rare based on the reconstruction of character state transformations on a phylogenetic tree, then the use of tree (and character state) reconstruction methods based on parsimony introduces an inappropriate influence. Although parsimony methods do not preclude the possibility of obtaining results in which homoplasy is inferred to be common (Farris 1983), they minimize and thus systematically underestimate homoplasy, which biases the analysis in favor of the conclusion that homoplasy is rare.

Problematic and Nonproblematic Influences

Not all systematic influences are logically problematic. Consider, once again, the use of parsimony methods in evaluating the commonness of homoplasy, except this time with the finding that homoplasy is common. The same bias exists as in the previous example, but this time it is not the source of inferential ambiguity. Because the method biases the analysis in favor of the inference that homoplasy is rare, obtaining the result that homoplasy is common cannot be at-

tributed to the methodological influence. Indeed, given that the method is biased against the finding that homoplasy is common, the analysis would have to be considered conservative with respect to that result. The influence is unproblematic, and it might even be considered appropriate or desirable, given the results.

Generalizing from the previous examples, we can conclude that the appropriate versus inappropriate nature of a given influence depends on the context. An influence that tends to favor a particular hypothesis is problematic and thus inappropriate if that hypothesis is supported by the results of an analysis containing the influence in question. But an identical influence is nonproblematic and appropriate, in the sense of strengthening the inference, if the hypothesis favored by the influence is contradicted and thus an alternative hypothesis is supported by the results. The concepts of appropriate and inappropriate influences are closely related to the concepts of conservative and liberal tests. An appropriate influence makes a test more conservative, whereas an inappropriate one renders a test more liberal. In any case, the critical issue in classifying a systematic methodological influence as appropriate or inappropriate is whether the results are potentially attributable to that influence, because that is what compromises the reliability of the inference.

A systematic influence can also be considered nonproblematic if it is neutral with respect to the hypotheses being evaluated. If the influence in question does not favor any of the hypotheses under consideration, then it causes no inferential ambiguity regardless of which hypothesis is supported by the results. For example, the use of tree reconstruction methods based on parsimony would not be problematic with regard to the question of whether rates of evolution have been constant or variable among lineages. Insofar as those methods bias the results in terms of the frequency of homoplasy rather than the distribution of character transformations among lineages, they do not bias the results in favor of either constant or variable rates.[1]

In short, the classification of a given influence as problematic (inappropriate) or nonproblematic (appropriate or neutral) depends on the specific circum-

---

[1]On the other hand, use of specific parsimony-based optimization methods, such as accelerated versus delayed transformation (Swofford and Maddison 1987), might well bias inferences about evolutionary rates in individual lineages. It should also be noted that although certain patterns of rate inequality may cause parsimony methods to construct the incorrect topology (Felsenstein 1978), the problem does not result from rate inequality per se but rather from certain patterns of branch length inequality (Hendy and Penny 1989). Moreover, the patterns of branch length inequality in question do not necessarily prevent parsimony methods from identifying rate inequalities, though in cases of "long edge [branch] attraction" (Hendy and Penny 1989), the lengths of long branches will be underestimated and those of certain short branches will be overestimated.

stances of an analysis. In particular, it depends on how that influence affects the results of the analysis with respect to the specific hypothesis under consideration. Because of this context dependence, the same influence can be appropriate in one situation, neutral in another, and inappropriate in yet another. Consequently, it is simplistic to reject a given method or practice without considering the specific relationship between the systematic influences inherent in that method and the hypothesis that is supported by the results of an analysis in which it is employed, and it is almost certainly incorrect to think that a given method or practice will be logically problematic in all cases.

PROBLEMS ASSOCIATED WITH INCLUDING (AND EXCLUDING) THE CHARACTERS OF INTEREST

In the context of the analysis presented above, it should not be surprising that including the characters of interest—or, for that matter, excluding them—leads to logical problems in some cases but not in others. The existence of a logical problem depends on how each practice bears on the hypotheses under consideration. Consider, for example, a study that seeks to assess whether the distribution of a character (state) among species is the result of single versus multiple evolutionary origins. In such a study, including the character of interest can be expected to influence the results in favor of a single origin, particularly when parsimony methods are used to reconstruct the phylogenetic tree. Because parsimony methods minimize total homoplasy, including the character of interest will tend to give results in which the numbers of its own reconstructed origins and losses are reduced relative to reconstructions on trees upon which that character had no influence (Maddison 1990; Maddison and Maddison 1992; Swofford and Maddison 1992). Such reductions will be logically problematic if the results support the hypothesis of a single origin. Because including the character of interest is expected to increase the chances that the results will indicate a single origin, that inference is not particularly cautious, and the practice of including the character of interest constitutes an inappropriate influence (but see below).

Nevertheless, if the results support the alternative hypothesis of multiple origins, then including the character of interest is not logically problematic. Because this practice will tend to influence the results in favor a single origin, a result indicating multiple origins is not called into question by that practice. In this case, the inference of multiple origins is conservative and the influence associated with including the character of interest is appropriate (but see below). For similar reasons, excluding the character of interest will be logically prob-

lematic if the results favor multiple origins, but it will not be problematic if they favor a single origin. Thus, with respect to the question of single versus multiple origins, both including and excluding the characters of interest can systematically influence the results. Whether the influence associated with either practice is appropriate or inappropriate, however, depends on which hypothesis is supported.

In other cases, the influence associated with including the characters of interest is not problematic, but for a different reason. Consider, for example, studies investigating the potential correlated evolution of two characters (Felsenstein 1985; Maddison 1990; Pagel 1994). Maddison (1990) pointed out that in his parsimony-based method for analyzing correlated character evolution, including the characters of interest will tend to reduce the number of gains and losses in those characters relative to optimizing them on a tree upon which they had no influence. Therefore, he concluded that including the characters of interest will likely render the test more conservative because with fewer changes it will be more difficult to obtain a significant correlation. For the same reason, excluding the characters of interest will likely render the test more liberal. It should be noted that the systematic influence in this case merely affects the number of observations, rather than specifically favoring a correlation between the characters or the lack thereof. It is also important to note that including the characters of interest—rather than excluding them—is expected to result in a more conservative test.

In still other cases, neither including nor excluding the characters of interest is expected to constitute a problematic influence. Consider an analysis of whether rates of evolution for a particular set of characters have been constant or variable among certain lineages. Although reconstructing the phylogeny either including or excluding the characters of interest may influence the numbers of inferred changes in those characters, neither practice is expected to predispose the analysis to distribute the changes equally over branches of equal temporal duration or to concentrate them in one or a few such branches. Therefore, provided that the methods of tree reconstruction and character optimization do not contain inherent rate biases, the characters of interest can be optimized on a tree derived from an entirely different set of characters, or they can be optimized on a tree based entirely on the characters of interest. Neither practice should compromise the rigor of the analysis by influencing the results in favor of either constant or variable rates.

From the examples presented above, we can conclude that both including and excluding the characters of interest can systematically influence the outcome of an analysis. Whether such an influence is logically problematic, however, depends on the specifics of the situation. In some cases, including the characters of interest constitutes an inappropriate influence; in other cases, excluding those

characters constitutes an inappropriate influence; and in still other cases, neither practice constitutes a significant influence. Consequently, inferential rigor cannot be guaranteed by adopting a simple rule to always exclude (or include) the characters of interest. Indeed, adopting a such a rule can sometimes have the opposite effect intended by the researcher—that is, it can result in a less rigorous analysis. Sound inferences require careful consideration of how each practice is expected to influence the results of a given analysis and how those influences bear on the evaluation of the specific hypotheses under consideration.

MORPHOLOGY VERSUS MOLECULES

Concerns about including versus excluding characters are most often raised in reference to morphological characters (and to a lesser extent, behavioral, ecological, and physiological ones). On the one hand, biologists commonly have an interest in the evolution of morphological characters (as well as behavioral, ecological, and physiological ones); on the other hand, morphological characters (unlike behavioral, ecological, and physiological ones) are the data most often available for reconstructing phylogenies—indeed, in many cases they are the only available data. According to biologists who believe that including the characters of interest is logically problematic, the solution is to reconstruct the tree based on an "independent" data set, most often a molecular one (e.g., Hedges and Maxson 1996). But biologists are also interested in the evolution of molecular characters, and one rarely, if ever, hears the argument that phylogenies based on an "independent" morphological data set are needed to analyze the evolution of molecular characters, though this conclusion follows from the premise that including the characters of interest is logically problematic (see also Lee 1997).

This discrepancy may be attributable, at least in part, to a difference in the kinds of questions that are being investigated concerning the evolution of molecular versus morphological characters. For example, studies about the evolution of molecular characters more commonly address issues about rates of evolution, while those about the evolution of morphological characters more commonly address issues about single versus multiple origins. Given this situation and the conclusions (see preceding section) that biases associated with including and excluding the characters of interest generally are not problematic for studies of evolutionary rates, but can be for studies of single versus multiple character state origins, perhaps including the characters of interest presents fewer problems for molecular characters. Nevertheless, the discrepancy also seems to reflect, at least in part, a prejudice on the part of certain researchers against morphological (as well as ecological, physiological, and behavioral)

characters relative to molecular ones. For example, though largely unsubstantiated (Sanderson and Donoghue 1989), there is a widespread opinion that morphological characters are more subject to homoplasy than molecular ones (e.g., Hedges and Maxson 1996). But regardless of the relative frequency of homoplasy in different classes of characters, my previous conclusions about the influences associated with including and excluding the characters of interest do not justify different standards for morphological versus molecular characters. If the practice of including (or excluding) the characters of interest introduces an inappropriate influence, then a logical problem will exist regardless of whether the characters of interest are morphological or molecular. Similarly, if the practice of including (or excluding) the characters of interest does not introduce an inappropriate influence, then no logical problem will exist regardless of whether the characters of interest are morphological or molecular.

Another manifestation of the prejudice against morphological characters is the extremely questionable practice of excluding all morphological characters on the grounds that the characters of interest are morphological (e.g., Block and Finnerty 1994; Meyer et al. 1994). As noted above, it may sometimes be justifiable to exclude the characters of interest to avoid an inappropriate influence, and it may even be justifiable to exclude other characters if they are thought to be adaptively (i.e., functionally and phylogenetically) correlated with the characters of interest. For example, someone studying the evolution of pollination systems might want to exclude certain characters, such as stamen length, that are thought to be an integral component of the pollination system. It is not justifiable, however, to exclude characters on the basis of their sharing some property with the characters of interest that has no bearing on the hypothesis or inference in question. Thus, when investigating the evolution of a particular morphological character, it is not justifiable to exclude other characters that are not thought to be adaptively correlated with the character of interest for the irrelevant reason that those characters also happen to be morphological (see also Donoghue and Sanderson 1992).

HOMOLOGS VERSUS ANALOGS

There are some cases in which it is reasonably clear that the characters of interest should not be included as part of the data used to reconstruct the phylogenetic tree. These are cases in which the characters of interest are classes of traits, often defined in terms of function, which are considered analogous rather than homologous prior to the phylogenetic analysis (usually on the grounds of similarity). For example, the class of traits referred to as floral color change may involve the petals in some species, the ovary in others, and the stamens in still

others (Weiss 1991). Similarly, the class of traits known as sexual dimorphism includes both cases in which males are larger than females and those in which females are larger than males. To the extent that such classes of traits are considered to consist of non-homologous components, they are likely to be misleading regarding phylogenetic relationships and thus should not be included—at least not as functionally defined classes—in the phylogenetic analysis (McLennan and Brooks 1993). This does not mean, however, that the individual components should not be included. Provided that those components consist of traits that are considered potentially homologous, they might well be included. Thus, although one might not want to include a character with the states "sexual dimorphism absent" and "sexual dimorphism present"; it might nevertheless be acceptable (at least in some cases) to include a character with the states "males larger than females," "males and females of equal size," and "females larger than males." Furthermore, just because the different instances of a functionally defined character are not the same (identical) does not mean that they are not potential homologs. For example, the fact that cooperative breeding in birds includes systems with both male and female helpers, as well as systems with male helpers only, does not mean those systems are necessarily non-homologous (contra McLennan and Brooks 1993). It is at least possible that a system with male and female helpers has been derived from one with male helpers only. All of these considerations apply primarily to studies that seek inferences that are the most likely to be correct. If, on the other hand, the primary concern is inferential caution, it may sometimes be permissible to include a character even when it is thought to include non-homologous components—if doing so makes the basis for an inference more conservative. For example, the inference that floral color change has evolved multiple times within flowering plants would only be strengthened by including the character "floral color change" (i.e., regardless of the specific structures involved) in the phylogenetic analysis, since this practice would tend to bias the analysis against the inference of multiple origins.

DIFFERENT TYPES OF STUDIES

Studies of character evolution can be classified in several ways with regard to some very general properties. Some generate new hypotheses, while others test existing hypotheses. Some evaluate the relative support for two or more alternative hypotheses (e.g., single versus multiple origins), while others attempt to reject a null hypothesis (e.g., no phylogenetic association between characters). Some focus on individual character transformations (e.g., origin of the avian flight apparatus), while others emphasize general classes of character

transformations (e.g., origin of cooperative breeding in all taxa in which it occurs) (see McLennan and Brooks [1993] and Edwards and Naeem [1994] for different perspectives on this issue). These different classes of studies have somewhat different consequences for the practices of including and excluding the characters of interest.

On the one hand, studies in many of the categories commonly adopt the perspective described above that seeks the estimate or inference about character evolution that is the most likely to be correct regardless of how much better it is supported than an alternative inference. For example, in proposing an hypothesis about the evolution of a particular character, one might accept the most parsimonious reconstruction of character state transformation inferred from the minimum length tree. Or one might favor whichever hypothesis, single or multiple character state origins, maximized the value of the relevant likelihood function on a given tree. The only studies that seem to be prohibited from adopting this perspective are those that explicitly attempt to reject a null hypothesis—that is, studies involving significance testing. By adopting the convention of rejecting the null hypothesis only when the data have a relatively low probability if the null is true, such studies implicitly adopt the alternative perspective of accepting an inference only if it is unlikely to be incorrect. In any case, under the perspective that seeks the single best estimate or inference about character evolution, it may generally be preferable to include the characters of interest not regardless of, but in that context because of, their potential influence on the results. Provided that the characters of interest contain phylogenetic information, that information is expected to improve the accuracy of inferences about their own evolution.

Even within the context of the perspective that emphasizes inferential caution, there are important differences between generating hypotheses and evaluating the relative merits of alternative hypotheses, on the one hand, and attempting to reject null hypotheses, on the other. When attempting to reject a null hypothesis, that hypothesis is explicitly identified prior to the analysis. Therefore, provided that the influence of excluding or including the characters of interest on the attempt to reject that hypothesis can also be determined prior to the analysis, the practice that renders the test more conservative can simply be chosen (but see below). Of course, failure to reject the null hypothesis should not be interpreted as support for the null hypothesis, since the influence associated with either including or excluding the characters of interest has been exploited to make the test more conservative—that is, to make it more difficult to reject the null hypothesis even if it is false.

In contrast, when a phylogenetic analysis is used to generate, rather than to reject, an hypothesis about character evolution, that hypothesis in many cases

will not have been identified prior to the analysis. This will make it impossible to know the influence of either including or excluding the characters of interest on the hypothesis in question until after the analysis has been performed and the hypothesis generated. Similar problems arise in analyses that simultaneously evaluate alternative hypotheses. In such analyses, it may be possible to determine which practice, including or excluding the characters of interest (if either), should tend to favor a particular hypothesis (but see below). Nevertheless, if the hypotheses are strict alternatives—such that contradiction of one implies support for the other and vice versa—then the outcome of the analysis becomes critical. In other words, it may still be impossible to identify the practice that results in a more conservative inference until after the analysis has been conducted, since one must first know which practice favors the inference that is actually supported by the analysis.

RECOMMENDATIONS

Given that both including and excluding the characters of interest can potentially compromise the conservativeness of an analysis, and given that in some cases it is not possible to identify which practice will do so prior to the analysis, how are comparative biologists to achieve logically sound inferences about character evolution? Although this question has not yet received a great deal of attention, several authors have already used methods or approaches that deal with the problem of including and excluding characters in more sophisticated ways. I will next review some these methods in an attempt to provide some practical guidance. My intent is not so much to endorse the particular methods that I discuss but rather to use those methods as examples of a general approach to analyses of character evolution whose conclusions do not suffer from inappropriate influences.

Perhaps the most straightforward approach is to conduct phylogenetic analyses both including and excluding the characters of interest. If the trees resulting from the separate analyses turn out to be identical in terms of whatever properties might affect the inference about character evolution (e.g., topology, branch lengths—though the latter are unlikely to be identical if they are estimated from the data that includes and excludes the characters of interest), then the inclusion versus exclusion of the characters of interest is irrelevant to inferences about their evolution. Evolutionary transformations in the characters of interest are simply reconstructed on the tree and the appropriate inferences about character evolution are drawn from them. On the other hand, if the trees resulting from the separate analyses differ, then transformations in the characters of interest are reconstructed on the alternative trees and inferences about

character evolution are drawn from them. If the inferences about character evolution are identical,[2] then the inclusion versus exclusion of the characters of interest is irrelevant, and the results can be accepted. If, on the other hand, the results based on the alternative trees differ, then relationships critical to evaluating the relative merits of the inferences in question are too tenuous for either to be accepted as a conservative inference. A better supported phylogeny and hence more data are needed (see also Swofford and Maddison 1992; Losos 1994).

The approach described above should be useful both for generating hypotheses about character evolution and for evaluating the relative merits of alternative hypotheses. A variant of it was used by Armbruster (1993) in his analysis of pollination and mating system evolution in *Dalechampia,* a clade of neotropical vines. Although Armbruster excluded the characters of interest from his phylogenetic analyses, he was nevertheless concerned about possible correlations between those characters and some of the characters that were used to reconstruct the phylogeny, particularly those describing certain aspects of floral morphology. Therefore, he performed separate phylogenetic analyses including and excluding the morphological characters that he believed might be correlated with pollination and mating systems. He found that the separate analyses yielded similar trees, with the main difference being that the tree for the reduced data set (presumably a consensus tree) had fewer resolved nodes. Several of Armbruster's inferences about the evolution of pollination and mating systems in *Dalechampia* were supported by both analyses, including the inference that pollination systems have been evolutionarily labile. Those inferences apparently are not dependent on including (or excluding) characters of floral morphology that are potentially related to pollination and mating systems.

Another method based on including and excluding characters was used by Ronquist (1994) in his analysis of the evolution of a type of nest parasitism in cynipid wasps. Some species of cynipids, known as inquilines, cannot induce galls in plants; instead, they are nest parasites whose young develop in the galls induced by other species of cynipids, the hosts. Ronquist wished to address the question of whether the inquilines had originated once, presumably from one of their hosts, and later spread to other hosts (inquiline monophyly), or whether each inquiline had originated separately, presumably from its host (inquiline polyphyly). He was concerned, however, that a phylogenetic analysis might be confounded by convergence, either among the inquilines, because of their similar parasitic mode of life, or between parasites and hosts, because of their similar environments. He reasoned that if the characters responsible for grouping

---

[2]This is not necessarily the same as saying that the reconstructions of individual character transformations are identical—particularly when inferences about character evolution are generalizations about classes of character transformations.

taxa together (whether different parasites or hosts and their parasites) in an initial analysis were convergent, then removal of those characters should result in a fundamental change in relationships (so that, for example, a parasite might first be grouped with its host, but afterwards with other parasites). On the other hand, if the characters were indicative of the true relationships, then their removal should result only in a loss of resolution. Ronquist performed an analysis incorporating this reasoning on 12 taxa of cynipid wasps, 4 of which were inquilines. His initial analysis supported a single origin of the inquilines, and subsequent removal of characters common to all inquilines did not change this result. Support for the hypothesis of inquiline monophyly apparently is not simply the result of including convergent characters related to inquilinism in the analysis.

Because multiple characters are generally used to reconstruct phylogenies, a potential exists for complex interactions among those characters in determining the optimal tree(s). Consequently, it is not always possible to predict the effects of including versus excluding the characters of interest on the outcome of an analysis by simple deduction. A study by Weller et al. (1995) on the evolution of breeding systems in Alsinoideae, a clade of endemic Hawaiian plants, illustrates this problem. Weller et al. used parsimony-based methods to reconstruct phylogenetic relationships among species of Alsinoideae including and excluding a character describing the breeding system, as well as several other characters describing aspects of floral morphology that were thought to be related to the breeding system. They then used the phylogeny and character optimization methods based on parsimony to infer the number of times breeding system dimorphism (gynodioecy, subdioecy, dioecy) had evolved. One might expect that exclusion of the breeding system character from the phylogenetic analysis and optimizing it on the tree afterwards would lead to inferring an equal or greater number of origins for dimorphism than optimizing it on a tree based on a data set that included the character. However, Weller et al. found that some of the most parsimonious trees required more origins of dimorphism when the breeding system character was included than when it was excluded. This result was obtained in at least some cases under both accelerated and delayed transformation methods of character optimization and treating the character as either ordered or unordered. A similar result was obtained by Armbruster (1993) regarding the consistency index of his pollination system character, which was higher (indicating less rather than more homoplasy) when the characters thought to be related to it were excluded from the phylogenetic analysis. These examples indicate that the effects of including versus excluding characters may sometimes run counter to expectations, suggesting that it is advisable to examine those effects empirically.

The analysis of character evolution on alternative trees can be extended from the optimal trees based on the analysis of data sets including and excluding the

characters of interest to larger sets of trees—such as all the resolved trees derivable from a strict consensus tree of the optimal trees for data sets including and excluding the characters of interest, or a random sample of those trees. To the extent that additional trees allow for additional reconstructions of character evolution, this approach can be used to increase the conservativeness of an analysis, though it might also be expected to decrease analytical power (i.e., the ability to make an inference). Losos (1994) and Martins (1996) have proposed using computer simulations to generate sets of trees for unresolved parts of a phylogeny, and Martins (1996) has described methods for estimating the parameter of interest, as well as confidence intervals, on sets of alternative trees.

When attempting to reject a null hypothesis, it is sometimes possible to estimate both the phylogenetic tree and the evolution of certain characters simultaneously, using the null hypothesis as a constraint in the estimation process. This approach was suggested with regard to testing for correlated character evolution by Felsenstein (1985) and Pagel (1994) and has recently found broad application in comparative phylogenetic studies (Huelsenbeck and Rannala 1997). In the approach described by Felsenstein and by Pagel, the characters of interest are included in the data used to estimate the tree, and the optimality criterion, in this case the likelihood, is a single multivariate quantity describing both the tree and the parameter of interest (in this case, the correlations). The maximum value of the likelihood function with the characters of interest unconstrained (allowing for correlated evolution) is then compared, in terms of a statistical criterion for significance, with the maximum likelihood obtained when the characters of interest are constrained to correspond with the null hypothesis (independent evolution). The significance testing approach guarantees conservativeness in terms of the probability of rejecting a true null hypothesis, but the effect of simultaneously optimizing the character correlations and the tree, as opposed to optimizing the character correlations on a tree estimated from other characters is not obvious. Once again, the most straightforward way to determine that effect is to do the analysis both ways.

CONCLUSIONS

Sometimes a distinction is made between systematics, the discipline concerned primarily with reconstructing or estimating phylogenetic trees, and comparative biology, the discipline concerned primarily with using those reconstructed trees to study character evolution. Deleporte (1993), for example, categorized the biologists associated with those respective disciplines as "phylogeny makers" and "phylogeny users." Although this dichotomy is useful for describing

different research emphases, it is also oversimplified. One thing that should be clear from the preceding analysis is that the rigorous evaluation of hypotheses about character evolution often demands the reconstruction or estimation of phylogenetic trees for that express purpose. In other words, comparative biologists cannot afford to limit themselves to using trees generated by systematists; they need to be actively involved in tree reconstruction. This need is obvious in cases where no phylogeny or only a poorly resolved one is available; but the present analysis indicates that it is also important for achieving cautious or conservative inferences even when a resolved phylogeny is available. The benefits to comparative biologists of being versed in the analysis of systematic data—or alternatively, of working closely with systematists—are tremendous. Comparative biologists who are able to produce their own trees will be able to satisfy their own unique needs, from investigating the effects of including and excluding characters (e.g., Armbruster 1993; Ronquist 1994; Weller et al. 1995), to generating sets of trees for unresolved parts of phylogenies (e.g., Losos 1994; Martins 1996), to comparing optimal unconstrained trees with those that are constrained to conform with various null hypotheses (e.g., Felsenstein 1985; Pagel 1994). But the benefits of generating trees for the express purpose of evaluating evolutionary hypotheses are not limited to promoting more rigorous analyses in this area. One of the greatest benefits will be a more thorough integration of systematics and comparative biology—two closely related fields that have much to gain from one another.

ACKNOWLEDGMENTS

REFERENCES

Armbruster, W. S. 1992. Phylogeny and the evolution of plant–animal interactions. Bioscience 42: 12–20.

Armbruster, W. S. 1993. Evolution of plant pollination systems: Hypotheses and tests with the neotropical vine *Dalechampia*. Evolution 47: 1480–1505.

Block, B. A., and J. R. Finnerty. 1994. Endothermy in fishes: A phylogenetic analysis of constraints, predispositions, and selection pressures. Environmental Biology of Fishes 40: 283–302.

Brooks, D. R., and D. A. McLennan. 1991. Phylogeny, Ecology, and Behavior: A Research Program in Comparative Biology. University of Chicago Press, Chicago.

Carpenter, J. M. 1989. Testing scenarios: Wasp social behavior. Cladistics 5: 131–144.

Coddington, J. A. 1988. Cladistic tests of adaptational hypotheses. Cladistics 4: 3–22.

Deleporte, P. 1993. Characters, attributes, and tests of evolutionary scenarios. Cladistics 9: 427–432.

de Queiroz, A., and P. H. Wimberger. 1993. The usefulness of behavior for phylogeny estimation: Levels of homoplasy in behavioral and morphological characters. Evolution 47: 46–60.

de Queiroz, K. 1989. Morphological and biochemical evolution in the sand lizards. Doctoral dissertation, University of California, Berkeley.

de Queiroz, K. 1996. Including the characters of interest during tree reconstruction and the problems of circularity and bias in studies of character evolution. American Naturalist 148: 700–708.

Donoghue, M. J. 1989. Phylogenies and the analysis of evolutionary sequences, with examples from seed plants. Evolution 43: 1137–1156.

Donoghue, M. J., and M. J. Sanderson. 1992. The suitability of molecular and morphological evidence in reconstructing plant phylogeny. Pages 340–368 *in* Molecular Systematics of Plants (ed. P. S. Soltis, D. E. Soltis, and J. J. Doyle). Chapman and Hall, New York.

Edwards, S. V., and S. Naeem. 1994. Homology and comparative methods in the study of avian cooperative breeding. American Naturalist 143: 723–733.

Eggleton, P., and R. I. Vane-Wright, eds. 1994. Phylogenetics and Ecology. Linnean Society Symposium Series, No. 17. Academic Press, London.

Eldredge, N., and J. Cracraft. 1980. Phylogenetic Patterns and the Evolutionary Process. Columbia University Press, New York.

Farris, J. S. 1983. The logical basis of phylogenetic analysis. Pages 7–36 *in* Advances in Cladistics, Volume 2: Proceedings of the Second Meeting of the Willi Hennig Society (ed. N. I. Platnick and V. A. Funk). Columbia University Press, New York.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Systematic Zoology 27: 401–410.

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

Funk, V. A., and D. R. Brooks. 1990. Phylogenetic systematics as the basis of comparative biology. Smithsonian Contributions to Botany 73: 1–45.

Harvey, P. H., and M. D. Pagel. 1991. The Comparative Method in Evolutionary Biology. Oxford University Press, Oxford.

Harvey, P. H., and A. Purvis. 1991. Comparative methods for explaining adaptations. Nature 351: 619–623.

Harvey, P. H., A. J. Leigh Brown, and J. Maynard Smith, eds. 1995. New uses for new phylogenies. Philosophical Transactions of the Royal Society of London B 349: 1–118.

Hedges, S. B., and L. R. Maxson. 1996. Re: Molecules and morphology in amniote phylogeny. Molecular Phylogenetics and Evolution 6: 312–314.

Hendy, M. D., and D. Penny. 1989. A framework for the quantitative study of evolutionary trees. Systematic Zoology 38: 297–309.

Hillis, D. M. 1987. Molecular versus morphological approaches to systematics. Annual Review of Ecology and Systematics 18: 23–42.

Huelsenbeck, J. P., and B. Rannala. 1997. Phylogenetic methods come of age: Testing hypotheses in an evolutionary context. Science 276: 227–232.

Kluge, A. G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). Systematic Zoology 38: 7–25.

Lauder, G. V., and K. F. Liem. 1989. The role of historical factors in the evolution of complex organismal functions. Pages 63–78 *in* Complex Organismal Functions: Integration and Evolution in Vertebrates (ed. D. B. Wake and G. Roth). John Wiley and Sons, New York.

Lee, M. S. Y. 1997. Re: Molecules, morphology, and phylogeny: A response to Hedges and Maxson. Molecular Phylogenetics and Evolution 7: 394–395.

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

Maddison, W. P. 1990. A method for testing the correlated evolution of two binary characters: Are gains and losses concentrated on certain branches of a phylogenetic tree? Evolution 44: 539–557.

Maddison, W. P., and D. R. Maddison. 1992. MacClade: Analysis of Phylogeny and Character Evolution. Sinauer, Sunderland, Mass.

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

McKey, D. 1991. Phylogenetic analysis of the evolution of a mutualism: *Leonardoxa* (Caesalpiniaceae) and its associated ants. Pages 310–334 *in* Ant–Plant Interactions (ed. C. R. Huxley and D. F. Cutler). Oxford University Press, Oxford.

McKitrick, M. C. 1993. Trends in the evolution of hindlimb musculature in aerial-foraging birds. Auk 110: 189–206.

McLennan, D. A., and D. R. Brooks. 1993. The phylogenetic component of cooperative breeding in perching birds: A commentary. American Naturalist 141: 790–795.

Meyer, A., J. M. Morrissey, and M. Schartl. 1994. Recurrent origin of a sexually selected trait in *Xiphophorus* fishes inferred from a molecular phylogeny. Nature 368: 539–542.

O'Hara, R. J. 1988. Homage to Clio, or, toward an historical philosophy for evolutionary biology. Systematic Zoology 37: 142–155.

Olmstead, R. 1989. Phylogeny, phenotypic evolution, and biogeography of the *Scutellaria angustifolis* complex (Lamiaceae): Inference from morphological and molecular data. Systematic Botany 14: 320–338.

Pagel, M. 1994. Detecting correlated evolution on phylogenies: A general method for

the comparative analysis of discrete characters. Proceedings of the Royal Society of London B 255: 37–45.

Patterson, C., ed. 1987. Molecules and Morphology in Evolution: Conflict or Compromise? Cambridge University Press, Cambridge.

Ronquist, F. 1994. Evolution of parasitism among closely related species: Phylogenetic relationships and the origin of inquilinism in gall wasps (Hymenoptera, Cynipidae). Evolution 48: 241–266.

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

Sillén-Tullberg, B., and A. P. Møller. 1993. The relationship between concealed ovulation and mating systems in anthropoid primates: A phylogenetic analysis. American Naturalist 141: 1–25.

Swofford, D. L., and W. P. Maddison. 1987. Reconstructing ancestral character states under Wagner parsimony. Mathematical Biosciences 87: 199–229.

Swofford, D. L., and W. P. Maddison. 1992. Parsimony, character-state reconstructions, and evolutionary inferences. Pages 186–223 *in* Systematics, Historical Ecology, and North American Freshwater Fishes (ed. R. L. Mayden). Stanford University Press, Stanford, Calif.

Sytsma, K. J. 1990. DNA and morphology: Inference of plant phylogeny. Trends in Ecology and Evolution 5: 104–110.

Vane-Wright, R. I., S. Schulz, and M. Boppré. 1992. The cladistics of *Amauris* butterflies: Congruence, consensus, and total evidence. Cladistics 8: 125–138.

Weiss, M. R. 1991. Floral colour changes as cues for pollinators. Nature 354: 227–229.

Weller, S. G., W. L. Wagner, and A. K. Sakai. 1995. A phylogenetic analysis of *Schiedea* and *Alsinidendron* (Caryophyllaceae: Alsinoideae): Implications for the evolution of breeding systems. Systematic Botany 20: 315–337.