STEVEN POE AND JOHN J. WIENS

2

CHARACTER SELECTION AND THE METHODOLOGY OF MORPHOLOGICAL PHYLOGENETICS

Estimation of phylogeny involves a series of interdependent steps: exploratory work (e.g., selection of characters and taxa), data collection and organization (e.g., character coding), and analysis (e.g., finding optimal trees). Many of the details and much of the rationale for choices made in these steps commonly are not made explicit in published papers; the analysis phase is often the only part of this series that is described (Thiele 1993). Morphological phylogenetics entails—in addition to those aspects that are characteristic of almost all phylogenetic analyses (such as using an algorithm to find a tree)—its own suite of usually implicit steps, such as the evaluation of character independence with reference to functional and growth constraints (Hennig 1966) and the methodology for coding observed anatomical variation for analysis (Pogue and Mickevich 1990).

Several researchers have noticed the lack of and stressed the need for explicitness in character selection criteria (e.g., Pimentel and Riggins 1987; Stevens 1991; Kesner 1994; Wiens 1995). However, the extent of this problem had not been formally studied and documented. Therefore, we did a systematic examination of more than a decade's worth of contributions to the primary literature of morphology-based phylogenetics. We investigated how morphologists chose characters for phylogenetic analysis, whether they were explicit about the criteria they set, and how much scientific merit the various character selection criteria have.

Table 2.1

Journals surveyed for morphological phylogenetic studies and the total number of applicable articles from each

Journal	No. articles
American Museum Novitates	71
Annals of the Entomological Society of America	25
Annals of the Missouri Botanical Garden	33
The Auk	11
Canadian Journal of Zoology	37
Cladistics	33
Condor	7
Copeia	22
Evolution	8
Herpetologica	8
Herpetological Monographs	4
Journal of Crustacean Biology	12
Journal of Herpetology	4
Journal of Mammalogy	3
Journal of Natural History	39
Journal of Zoology	11
Phycologia	4
Plant Systematics and Evolution	22
Systematic Botany	52
Systematic Zoology/Biology	13
Taxon	4
Wilson Bulletin	2
Zoological Journal of the Linnean Society	87

For each of the journals listed, our survey covered all issues dated 1986–1997, with the following exceptions: *American Museum Novitates*, 1995 volume not included; *Journal of Crustacean Biology*, 1996 and 1997 volumes not included.

METHODS

We collected data on morphological phylogenetic studies published in the years 1986–1997 in the 23 journals listed in table 2.1. To be included, an article had to be primary literature about a study of relationships, and the morphological characters had to have been selected for that study. We thus excluded papers that were reanalyses of data from the literature. In addition, we excluded studies in which trees were constructed by intuition; a numerical (but not necessarily computer-assisted) analysis of characters was required. Also excluded were studies in which the researchers did not seek independent characters (e.g., studies that involved principal-components analyses).

Table 2.2

Examples of criteria used in choosing characters

Reference	Criteria
Livezey 1986	"I chose to exclude mensural characters because . . . polarities and character states . . . are difficult to determine[.]"
Kraus 1988	"Two criteria were used for accepting characters for this analysis. First, the character had to vary discontinuously within the study group. Second, the character states had to be largely invariant within taxa. By 'largely invariant' I mean present in about 80% or more of the adult individuals of the species I examined. . . . Despite the second criterion, I occasionally allowed for the use of polymorphic characters in the analysis if the character was polymorphic for only one or two species and was invariant in the remaining species."
Cox and Urbatsch 1990	"Quantitative characters were generally rejected because the character states are circumscribed arbitrarily and because such arbitrary states probably have nothing to do with homology and analysis."
Bateman et al. 1992	Characters rejected when "known for less than ⅔ of OTUs."
Gensel 1992	Character excluded if "too variable (and either autapomorphic or homoplasious) to be included."
Hufford and Dickson 1992	"[C]haracters were deleted in the course of the investigation because of excessive missing data or polymorphism within groups."
Hood et al. 1994	"[C]haracters were selected for phylogenetic analysis because of their low intraspecific variation, presence in most taxa, ability to be scored unequivocally, and phylogenetic informativeness."
Kesner 1994	"[Desired] characters distinctive, biologically significant, and sufficiently constant within OTUs to allow for reliable character state assignment."

Each study that met our requirements for inclusion was evaluated as to whether or not criteria for character selection were mentioned, and any criteria were noted. Specifically, we scored papers positively if they mentioned exclusion of characters for any reason (other than to fit the two basic assumptions noted below), and we scored papers negatively if they did not mention character selection or if they explained character selection only in terms of which types were included. We assumed that independence and putative homology (common ancestry of states) are a part of any character selection strategy and therefore gave no credit for statements to this effect. Examples of character selection criteria are exclusion of a character because of high levels of intraspecific variation and exclusion because of missing data. Table 2.2 gives examples from the literature.

It is possible to distinguish between operational and conceptual criteria. For example, requiring "low" intraspecific variation is a conceptual criterion, whereas allowing only characters that have fixed states in terminal taxa is an operational criterion. However, because so few of the papers (<1%) included operational criteria, we chose to combine these two subcategories. It is also possible in some cases to distinguish between a criterion and a justification for that criterion. For example, a researcher could choose to reject all quantitatively coded characters (a criterion) because the continuous nature of such characters is thought to be unsuitable for cladistic analysis (a justification). However, justifications can be used as criteria, and vice versa, depending on context. So we treated criteria and justifications as one category in our scoring.

Although we focused on the methods section of articles in looking for the general character sampling schemes, we also examined the character descriptions. Some authors gave no general sampling scheme but did include detailed discussions of variation and of decisions concerning particular characters. Such discussions were scored positively for offering criteria. However, given the large number of articles we surveyed, we may have missed a few listings of criteria, particularly if they were buried in lengthy character descriptions.

We used the data that we collected to investigate four possibilities. First, we determined whether morphologists generally were explicit about character selection criteria. Second, we divided the data by journal type to assess whether authors publishing in general systematic journals were more conscientious than those publishing in taxon-oriented journals. Third, we compared our data on plant studies with our data on animal studies to detect any differences between the two fields. Fourth, we mapped trends over time to assess whether the advancement of phylogenetic methods has been paralleled by an increase in documented rigor of character selection.

RESULTS AND DISCUSSION

A Lack of Explicitness

The results of our survey suggest that, in general, morphologists are not explicit about how they choose systematic characters: Of the 512 morphological phylogenetic studies examined, 102 (20%) mentioned criteria for exclusion of char-

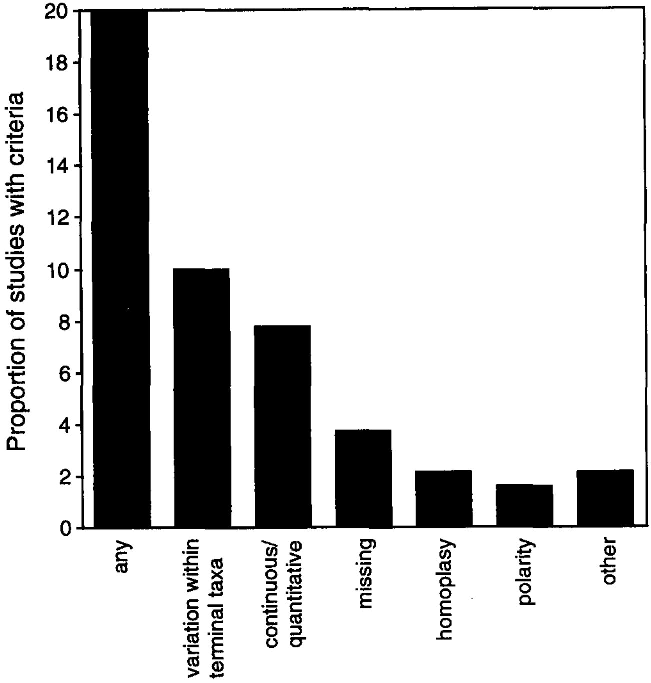

Figure 2.1. Explicitness of the morphological phylogenetic studies in our survey and the frequency with which the various character exclusion criteria were used. "Any" (meaning "any criterion for character exclusion was mentioned") is the overall percentage of studies that were explicit about character exclusion criteria. Some studies listed more than one criterion. "Other" refers to criteria that matched none of the others listed.

acters (figure 2.1). Papers in general systematics (or conceptual) journals were no more explicit than papers in taxon-oriented (organismal) journals, nor was there a difference between animal and plant studies ($P > 0.10$ for both comparisons, G-test of independence; figure 2.2). Furthermore, morphological phylogenetics has not shown a consistent trend toward more explicitness over the years ($P > 0.10$, Spearman's coefficient of rank correlation; figure 2.3).

Why might this state of affairs be maintained, despite pleas in the literature for change (e.g., Pogue and Mickevich 1990; Wiens 1995)? One possible explanation (which could not be addressed in a literature survey) is that morphological systematists believe that a common filter exists among them (Thiele 1993), such that discovering morphological characters is an objective and repeatable process even without listed criteria. Such commonality of observation seems unlikely, especially considering that different observers set dissimilar

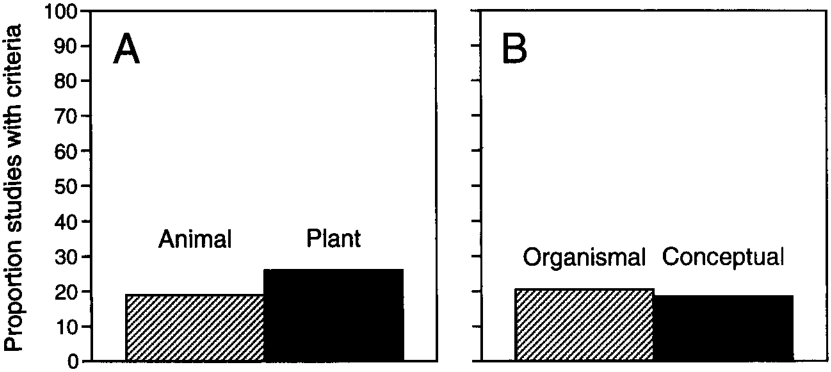

Figure 2.2. Explicitness of morphological phylogenetic studies according to type of journal in which they were published: **(A)** animal-studies or plant-studies journal; **(B)** taxon-oriented (organismal) or general systematic journal.

character state boundaries when presented with the same data on variation (Gift and Stevens 1997). Although there is probably little disagreement over what constitutes an ideal character, there is evidence of greatly differing opinions among morphologists as to where the dividing line is between a marginally acceptable and unacceptable character. The results of our survey suggest that people select characters differently. Comparing studies of the same organisms by different researchers is also telling (e.g., compare Good 1988 with J. A. Campbell and Frost 1993).

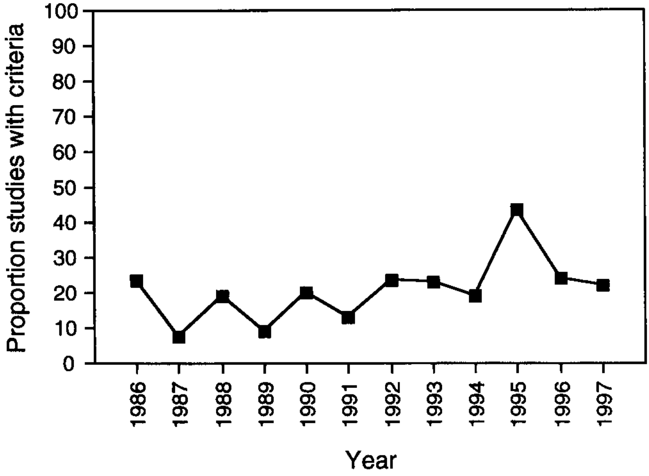

Figure 2.3. Explicitness of morphological phylogenetic studies by year of publication.

Inconsistency of character selection criteria can occur even within a single study. Stevens (1991) pointed out several examples of researchers who claimed to reject "quantitative" characters but who nevertheless accepted characters that could be described with an arbitrary quantitative cutoff, such as "calyx: >5 mm [0], <5 mm [1]" (or, less obviously, "calyx: large [0], small [1]"). A common filter is a myth.

Why Be Explicit?

One reason to be explicit about character selection criteria is simply to increase the rigor of morphological systematics. We recognize that an element of subjectivity may always be present in character selection—no two systematists are likely to view variation in exactly the same way. But minimizing subjectivity in methods is a goal of science. If the selection of characters remains a "black box" between selecting taxa and forming a matrix of zeros and ones, morphological phylogenetics will continue to be vulnerable to attack from those who accuse researchers of manipulating data to reach a preconceived phylogeny (e.g., Pritchard 1994) and from those who see such methods as sound in theory but flawed in implementation (e.g., Gould 1983). There is evidence that preconceived ideas about relationships affect the character choice of some systematists (e.g., those who screen out characters that show "too much homoplasy"; see below). The extent to which characters are picked to produce the desired tree is open to question if no selection criteria are stated. Furthermore, the differences between phylogenetic trees constructed by different authors for the same taxa may be difficult to interpret without knowing what criteria the authors used for including and excluding characters.

A second reason to be explicit is to allow for testing of the validity of character selection criteria and the properties of particular types of characters. For example, intraspecific variation is one of the most widely cited criteria for excluding characters, but the merit of this criterion has been tested only recently (e.g., J. A. Campbell and Frost 1993; Wiens 1995, 1998b; Wiens and Servedio 1997). Attempts to analyze the properties of polymorphic characters have been hampered by the fact that many morphological data matrixes do not indicate any intraspecific variation for the taxa and characters coded. Without explicit criteria for character selection, it is impossible to tell whether the absence of polymorphism in a matrix is due to actual rarity of intraspecific variation, the researcher's exclusion of polymorphic characters, or the researcher's deliberate failure to report uncommon variants within species (Wiens 1995). Even if polymorphic characters are explicitly included, many characters may still be excluded because "too much" intraspecific variation was observed (e.g.,

J. A. Campbell and Frost 1993). Thus, without explicit criteria for character selection, it is difficult to tell whether or not a given morphological study provides an unbiased sample of polymorphic characters, and what the nature and extent of the biases (if any) are. Because of this lack of information, empirical studies of polymorphic characters have so far been based on only a handful of data sets that included an explicit and seemingly unbiased sampling protocol for intraspecifically variable characters. A comparable situation exists for quantitative characters (Thiele 1993).

Similarly, comparisons of homoplasy (or some other measure) between different character "types" may be confounded by the methods by which an investigator chooses particular types of characters. For example, the finding of de Queiroz and Wimberger (1993) that behavioral characters are comparable in homoplasy levels to morphological characters could be explained by at least two hypotheses. First, there may be some inherent similarity in evolutionary plasticity in these two classes of data. Alternatively, systematists may be evaluating behavioral and morphological data through much the same filter, such that they tend to choose only characters that vary in an acceptable way, regardless of data type. The interpretation of comparative results such as these is crucial: the first of these hypotheses tells us something about evolution, whereas the second tells us something about systematists. These hypotheses can be tested only if investigators are explicit about character selection criteria.

Comparisons of rates of morphological evolution also may be rendered uninterpretable if no criteria for accepting or rejecting characters are listed. For example, Slowinski (1995) used seven characters in a study of 18 species of coral snakes and concluded that the group is "morphologically conservative." In contrast, Kluge (1989) used 53 morphological characters in his analysis of the relationships of 10 boid snakes. Does the discrepancy in number of characters included indicate that coral snakes are less variable than boas? It might, or it might mean only that Kluge and Slowinski have very different filters for character acceptability. We cannot test either of these explanations because neither of these authors gave explicit general criteria for how they chose characters.

Character Selection Criteria

In this section, we discuss the merits of various criteria for selecting morphological phylogenetic characters. Each of these criteria was mentioned by several authors of morphological phylogenetic papers; criteria listed only once are not discussed. In evaluating these criteria, our main concern is whether use of each criterion is likely to contribute to greater accuracy in morphological phylogenetics (where accuracy is the probability of estimating the true organismal phylogeny). However, we acknowledge that in most empirical studies the true phylogeny is unknown, that our predictions about accuracy come largely from statistical analyses, computer simulations, and congruence studies (see Hillis 1995 for a review), and that not everyone is concerned about phylogenetic accuracy (e.g., Kluge 1997; Siddall and Kluge 1997) or convinced by studies that attempt to address it.

In discussing these criteria, we point out some perceived disadvantages of particular character types that in fact appear to be applicable to morphological phylogenetics in general. Nonetheless, these potential problems should not be taken as indictments of the quality of morphological data relative to other types of data, for three main reasons. First, as discussed below, most of the perceived disadvantages are unfounded for morphology in general, as well as for particular types of morphological data. Second, some of the disadvantages discussed here are the same for or have clear analogs in other types of data. For example, coding intraspecific variation can pose problems regardless of the data type. Third, although some problems are for the most part limited to morphology (e.g., continuous variation), other data types have their own problems as well (e.g., alignment in DNA sequence data).

Some of the criteria discussed below overlap in content or in the sense that one criterion may be used as a justification for another. Our categories are thus somewhat arbitrary, and discussions of some criteria may bear on others.

Variation within terminal taxa. We found that the most common reason given for excluding characters was variation within terminal taxa (figure 2.1), particularly variation within species (i.e., polymorphism). Although there is evidence that polymorphic characters are more homoplastic than fixed characters (J. A. Campbell and Frost 1993; Wiens 1995), there is also evidence (from the same data sets) that polymorphic characters contain useful phylogenetic information. These two observations appear to make for a difficult choice in terms of whether or not to include intraspecifically variable characters (i.e., is it better to include more characters despite the increase in homoplasy or fewer characters with less homoplasy?). However, studies of real and simulated data sets (Wiens and Servedio 1997; Wiens 1998b) suggest that the choice is clear: Analyses that include polymorphic characters are consistently more accurate than those that exclude them, even when the data sets show a strong positive relationship between levels of homoplasy and intraspecific variability. Even though intraspecifically variable characters do seem to be more homoplastic, they evidently contain information that more than compensates for their homoplasy.

Similarly, characters are often excluded because they vary among the species within a higher level terminal taxon. In a simulation study, Wiens (1998a) showed that excluding such characters consistently decreased the accuracy of phylogenetic estimates, relative to including such characters. Furthermore, as discussed in more detail in chapter 5 of this book, Wiens's analysis of published morphological data for squamate reptiles (from Estes et al. 1988) suggested that (1) characters that vary among the species of a given higher taxon hold significant phylogenetic information for recovering higher level relationships (but are somewhat less informative than characters that are fixed within the higher taxon) and (2) there is generally no significant relationship between levels of homoplasy and the extent to which characters vary among the species within a given higher taxon. Thus, there is little support for the practice of excluding characters merely because they vary within the terminal units of a phylogenetic analysis.

Missing data. The desire to avoid having "missing data entries" (unknown or unscorable character states for some taxa) is sometimes used as a justification for excluding characters (figure 2.1). A recent simulation study (Wiens 1998c) provided little support for this practice. Under a wide variety of simulated conditions, adding characters with missing data significantly increased the phylogenetic accuracy of a given data set. Therefore, excluding characters with missing data must significantly decrease accuracy. Even when 75% of the data cells are scored as unknown, including such characters can still significantly increase accuracy.

Missing data entries themselves appear to be relatively harmless. However, taxa with missing data for a given character are effectively unsampled for that character, and limited taxon sampling may lead to positively misleading results through long branch attraction (Swofford et al. 1996). Under some conditions, including characters that cannot be scored for many taxa may decrease phylogenetic accuracy. This result seems most likely when very few taxa are sampled and the taxa that are sampled are distributed on the true phylogeny such that there are long terminal branches connected by a short internal branch (i.e., the situation described by Felsenstein 1978). Nevertheless, when unknown data entries are distributed randomly among taxa or are confined to a monophyletic subset of taxa, inclusion of characters with incomplete data is much more likely to increase accuracy than decrease it (Wiens 1998c). Therefore, we reject the general practice of excluding characters simply because they have missing data.

Continuous and quantitatively coded variation. Many authors who gave criteria mentioned a desire to avoid "continuous" variation in favor of "discrete" variation and to avoid "quantitative" coding in favor of "qualitative" coding. The terms "continuous" and "quantitative" often are used interchangeably but have distinct definitions (Thiele 1993). "Qualitative" and "quantitative" are ways of describing variation (descriptively and numerically, respectively) rather than properties of variation; "continuous" and "discrete" are potential properties of frequency distributions. Many types of variation can be coded either qualitatively or quantitatively, and many qualitatively coded data are "based on a quantitative phenomenological base filtered through the reified semantic discontinuities of . . . [anatomical] terminology" (Stevens 1991). Although these terms are distinct, we discuss them together because the reported justifications for exclusion of each of these character types are similar.

Quantitative coding and continuous variation have been rejected for any of a variety of reasons: because continuous variation is theoretically unsuitable for cladistic analysis (e.g., Pimentel and Riggins 1987), because state delimitation is arbitrary (e.g., Cox and Urbatsch 1990), because there is no consensus on a suitable coding method for quantitative data (see reviews listed below), because some types of continuous variation are difficult to characterize objectively (e.g., Wiens 1993), or because accuracy is reduced by use of such characters (e.g., C. S. Campbell 1986). These criticisms, which in general are poorly founded, are each discussed below.

Certainly phylogenetic methods for morphology were designed with discrete variation in mind (Hennig 1966; Wiley 1981). However, it does not follow from this intention that useful phylogenetic information cannot be extracted from continuous variation. In fact, numerous methods have been devised to code continuous variation in ways that satisfy cladistic assumptions of independence and discreteness of states (e.g., Archie 1985; Chappill 1989; Thiele 1993; Strait et al. 1996). Furthermore, it is not clear just how discrete characters must be to be considered useable. Stevens (1991) gave examples of variation that could be justifiably coded with either discrete cutoffs or more quantitative methods and showed that characters cannot be sharply distinguished as either discrete or continuous; rather, they fall along a continuum (see also figure 2 of Thiele 1993). There appears to be nothing uniquely undesirable about continuous variation in phylogenetic analysis, and there seems to be no objective means to distinguish "bad" ("continuous") variation or "good" ("discrete") variation in practice. Furthermore, Thiele (1993) has shown that quantitative characters can be comparable to more traditionally scored characters in homoplasy levels, phylogenetic informativeness, and resulting tree topology.

If arbitrary delimitation of states is a problem at all, it is for phylogenetic analysis of morphology in general, not just for analysis of quantitative data. Arbitrary delimitation of states is common in qualitatively coded variation, as in the decision to code a size difference as "large (0), small or absent (1)," or as "present (0), absent (1)," or as "large (0), small (1), absent (2)." In presence/

absence coding, which might be viewed as nonarbitrary, variation in size and shape is ignored under the rubric of "present." Is the presence/absence division really more significant, evolutionarily, than the large/small division? In fact, quantitatively coded variation may be less subject to vagaries in coding than is qualitatively coded variation, because explicit statistical or methodological grounds are often used in quantitative coding (Thiele 1993).

Although we do not advocate use of characters for which state delimitation is completely arbitrary, we recognize that stochastic genetic and environmental variation will produce situations where a "correct" coding scheme is not obvious but where potentially useful phylogenetic information remains. An example of so-called arbitrary coding that does not necessarily mislead is seen in Stevens (1991). His table 1 depicts a set of states coded in four different ways. He argued from this example that such characters are not suitable for cladistic analysis because there is no nonarbitrary way of deciding which coding scheme is best. However, although the different coding schemes do not produce identical groupings, the various possible relationships are largely compatible and in many cases differ only in the degree of resolution. For example, three of the four coding schemes group taxa A, B, and C together, three of four group A′, B′, and C′ together, and all find the maximum difference (e.g., one step) between taxa A and C′. Methods such as gap weighting (e.g., Thiele 1993) and frequency coding (e.g., Wiens 1995), which give greater weights to larger differences between taxa and lower weights to differences between similar taxa should be useful in cases where state delimitation has a potentially large arbitrary component.

The lack of a consensus regarding quantitative coding methods is well documented (see Archie 1985; Chappill 1989; Farris 1990; Strait et al. 1996). However, no accepted protocol exists for coding qualitative characters either, and examples of varying methods for coding discrete or qualitative variation abound (e.g., binary versus multistate, composite versus reductive; Wilkinson 1995). Clearly, the lack of a consensus for coding methods is not restricted to quantitative characters.

In our survey, a few authors rejected characters because of difficulty in objectively assigning character states (e.g., when there was continuous variation in shape or color among taxa). We argue that even if characters cannot be coded into discrete character states, they still can be characterized quantitatively in a phylogenetically useful way. For example, the thin-plate spline method may be useful in converting shape differences into phylogenetic characters (Zelditch et al. 1995) and one could conceivably assign numerical values to shades of color. However, some characters may be intractable to both quantitative and qualitative coding—for example, continuously varying shape characters lacking obvious landmarks for making either qualitative character state boundaries or measurements. For this reason, we consider difficulty in characterization to be a legitimate but perhaps overly used reason to reject characters.

The idea that inclusion of quantitative characters decreases phylogenetic accuracy has yet to be addressed directly. However, an empirical study by Thiele (1993) showed that continuous characters (at least when coded using gap weighting) can display significant nonrandom covariation and can result in trees that are significantly similar to or identical to trees based on qualitative characters. We take these results as evidence that including continuous characters increases accuracy, and we know of no examples that support the opposite conclusion.

Unknown polarity. At one point in the history of phylogenetics, it was thought that the plesiomorphic state of a character had to be determined a priori for that character to be useful in phylogenetic analysis (e.g., Wiley 1981). Consequently, many authors rejected characters that could not be polarized, and some have given this as an explicit criterion for excluding data. However, the idea that characters must be polarized to be included in a phylogenetic analysis has been shown to be misguided (e.g., Swofford and Olsen 1990). If an analysis includes a hypothetical ancestor taxon or outgroup to root the ingroup tree, then an unpolarized character merely represents a character with an ambiguous or unknown state for that taxon and still contributes to the resolution of ingroup relationships. Simulations suggest that excluding characters from analysis simply because their states cannot be determined in a few taxa significantly decreases phylogenetic accuracy (Wiens 1998c). Furthermore, unknown polarity of the characters is not likely to be an issue in a study design that includes outgroup taxa as terminal units in a global phylogenetic analysis (as advocated by Nixon and Carpenter 1993), in contrast to a design that requires a priori evaluation of character polarity (e.g., Maddison et al. 1984).

Level of homoplasy. Some authors stated that they chose characters that are conservative or not too homoplastic. This practice is flawed in that our perception of whether a character is homoplastic or not depends on some a priori notion of phylogeny (Hennig 1966; Wiley 1981), and the authors citing "too much homoplasy" as a criterion for character exclusion did not state how they determined these characters to be too homoplastic. Nevertheless, choosing characters that evolve at an appropriate rate for the level of analysis has become recognized as an important step in phylogenetic analyses of molecular data (Hillis et al. 1996). Practical means have been developed to identify and address these issues for molecular sequence data (e.g., comparing types of changes, such as transitions versus transversions; differential weighting to accommodate rate heterogeneity; identifying hypervariable regions of alignment), but many of these practices do not have clear analogs for morphological data. Researchers attempting to identify general classes of morphological characters that are too homoplastic for inclusion in phylogenetic studies have consistently arrived at the same conclusion: that these questionable character types do contain useful phylogenetic information and thus should not be excluded (e.g., integumentary characters in birds [Chu 1998]; reductive characters [Begle 1991; Buckup 1993]; morphometric characters [Thiele 1993]; and polymorphic characters [J. A. Campbell and Frost 1993; Wiens 1995, 1998b]). Until certain classes of morphological characters are convincingly identified as generally misleading, or until a method is developed to consistently identify misleading characters in particular cases, we reject the practice of excluding "excessively homoplastic" characters.

SUMMARY AND RECOMMENDATIONS

Our main conclusion from our survey is that most morphologists are not explicit about how they choose systematic characters. Our most basic recommendation from this work is not controversial: quite simply, morphological systematists need to discuss character selection and give their operational criteria for rejecting characters. The methodology of winnowing observed morphological variation into a data matrix of character states should be given as standard practice, and reviewers and advisers should ask questions about this methodology as a matter of course. The lack of explicit methodology for character selection for most morphological data sets seriously compromises (1) the claimed objectivity of modern morphological phylogenetics, (2) the comparison of phylogenies published by different authors, (3) tests of the properties of different types of morphological characters, (4) the comparison of morphological data to other types of characters (e.g., molecular, behavioral), and (5) the comparison of morphological evolution between taxonomic groups.

We found that systematists use a variety of criteria to exclude characters. However, there is little evidence to support the validity of most of the commonly cited criteria. In fact, simulation and congruence studies suggest that application of many of these criteria may actually decrease phylogenetic accuracy by discarding useful data. Although our discussion of character exclusion criteria has been largely from a perspective of maximizing phylogenetic accuracy, we suggest that a philosophically based "total evidence" approach to phylogenetics (e.g., Kluge 1989) would likewise require abandoning these exclusion criteria.

How should morphologists choose characters? The lack of evidence to justify most of the listed character rejection criteria suggests that much more variation could be included in phylogenetic analyses than is used presently. We hope that this review will encourage systematists to be more explicit about their methodology and to consider carefully their exclusion of potentially informative data.

ACKNOWLEDGMENTS

Thanks to Peter Stevens and Brad Livezey for helpful comments in review. S.P. was supported by fellowships from the National Science Foundation and the Smithsonian Institution.

REFERENCES

Archie, J. W. 1985. Methods for coding variable morphological features for numerical taxonomic analysis. Systematic Zoology 34: 326–345.

Bateman, R., W. A. di Michele, and D. Willard. 1992. Experimental cladistic analysis of anatomically preserved arborescent lycopsids from the Carboniferous of Euramerica: An essay on paleobotanical phylogenetics. Annals of the Missouri Botanical Gardens 79: 500–559.

Begle, D. P. 1991. Relationships of the osmeroid fishes and the use of reductive characters in phylogenetic analysis. Systematic Zoology 40: 33–53.

Buckup, P. A. 1993. Phylogenetic interrelationships and reductive evolution in neotropical characidiin fishes (Characiformes, Ostariophysi). Cladistics 9: 305–341.

Campbell, C. S. 1986. Phylogenetic reconstructions and two new varieties in the *Andropogon virginicus* complex (Poaceae: Andropogoneae). Systematic Botany 11: 280–291.

Campbell, J. A., and D. R. Frost. 1993. Anguid lizards of the genus *Abronia:* Revisionary notes, descriptions of four new species, phylogenetic analysis, and key. Bulletin of the American Museum of Natural History 216: 1–121.

Chappill, J. A. 1989. Quantitative characters in phylogenetic analysis. Cladistics 5: 217–234.

Chu, P. C. 1998. A phylogeny of the gulls (Aves: Larinae) inferred from osteological and integumentary characters. Cladistics 14: 1–43.

Cox, P. B., and L. Urbatsch. 1990. A phylogenetic analysis of the coneflower genera (Asteraceae: Heliantheae). Systematic Botany 15: 394–402.

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

Estes, R., K. de Queiroz, and J. A. Gauthier. 1988. Phylogenetic relationships within Squamata. Pages 119–281 *in* Phylogenetic Relationships of the Lizard Families (ed. R. Estes and G. K. Pregill). Stanford University Press, Stanford, Calif.

Farris, J. S. 1990. Phenetics in camouflage. Cladistics 6: 91–100.

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

Gensel, P. 1992. Phylogenetic relationships of the zosterophylls and lycopsids: Evidence from morphology, paleoecology, and cladistic methods of inference. Annals of the Missouri Botanical Gardens 79: 450–473.

Gift, N., and P. F. Stevens. 1997. Vagaries in the delimitation of character states in quantitative variation: An experimental study. Systematic Biology 46: 112–125.

Good, D. A. 1988. Phylogenetic relationships among gerrhonotine lizards. University of California Publications in Zoology 121: 1–138.

Gould, S. J. 1983. Hen's Teeth and Horse's Toes: Further Reflections in Natural History. W. W. Norton, New York.

Hennig, W. 1966. Phylogenetic Systematics. University of Illinois Press, Urbana.

Hillis, D. M. 1995. Approaches for assessing phylogenetic accuracy. Systematic Biology 44: 3–16.

Hillis, D. M., B. K. Mable, and C. Moritz. 1996. Applications of molecular systematics: The state of the field and a look to the future. Pages 515–543 *in* Molecular Systematics, 2nd edition (ed. D. M. Hillis, C. Moritz, and B. K. Mable). Sinauer Associates, Sunderland, Mass.

Hood, S. B., A. Reznick, and J. D. Palmer. 1994. Phylogenetic relationships in *Anemone* (Ranunculaceae) based on morphology and chloroplast DNA. Systematic Botany 19: 169–200.

Hufford, L., and W. Dickson. 1992. A phylogenetic analysis of Cunoniaceae. Systematic Botany 17: 181–200.

Kesner, M. H. 1994. The impact of morphological variants on a cladistic hypothesis with an example from a myological data set. Systematic Biology 43: 41–57.

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

Kluge, A. G. 1997. Testability and the refutation and corroboration of cladistic hypotheses. Cladistics 13: 81–96.

Kraus, F. 1988. An empirical evaluation of the use of the ontogeny polarization criterion in phylogenetic inference. Systematic Zoology 37: 106–141.

Livezey, B. C. 1986. Phylogeny and historical biogeography of steamer ducks (Anatidae: *Tachyeres*). Systematic Zoology 35: 458–469.

Maddison, W. P., M. J. Donoghue, and D. R. Maddison. 1984. Outgroup analysis and parsimony. Systematic Zoology 33: 83–103.

Nixon, K. C., and J. M. Carpenter. 1993. On outgroups. Cladistics 9: 413–426.

Pimentel, R., and R. Riggins. 1987. The nature of cladistic data. Cladistics 3: 201–209.

Pogue, M., and M. Mickevich. 1990. Character definitions and character-state delimitations: The bête noire of phylogenetic inference. Cladistics 6: 319–361.

Pritchard, P. 1994. Cladism: The great delusion. Herpetological Review 25: 103–111.

Siddall, M. E., and A. G. Kluge. 1997. Probabilism and phylogenetic inference. Cladistics 13: 313–336.

Slowinski, J. 1995. A phylogenetic analysis of the New World coral snakes (Elapidae: *Leptomicrurus, Micruroides,* and *Micrurus*) based on allozymic and morphological characters. Journal of Herpetology 29: 325–338.

Stevens, P. F. 1991. Character states, morphological variation, and phylogenetic analysis: A review. Systematic Botany 16: 553–583.

Strait, D., M. Moniz, and P. Strait. 1996. Finite mixture coding: A new approach to coding continuous characters. Systematic Biology 45: 67–78.

Swofford, D. L., and G. J. Olsen. 1990. Phylogeny reconstruction. Pages 411–501 *in* Molecular Systematics (ed. D. M. Hillis and C. Moritz). Sinauer Associates, Sunderland, Mass.

Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Phylogeny reconstruction. Pages 407–514 *in* Molecular Systematics, 2nd edition (ed. D. M. Hillis, C. Moritz, and B. Mable). Sinauer Associates, Sunderland, Mass.

Thiele, K. 1993. The holy grail of the perfect character: The cladistic treatment of morphometric data. Cladistics 9: 275–304.

Wiens, J. J. 1993. Phylogenetic systematics of the tree lizards (genus *Urosaurus*). Herpetologica 49: 399–420.

Wiens, J. J. 1995. Polymorphic characters in phylogenetic systematics. Systematic Biology 44: 482–500.

Wiens, J. J. 1998a. The accuracy of methods for coding and sampling higher-level taxa for phylogenetic analysis: A simulation study. Systematic Biology 47: 381–397.

Wiens, J. J. 1998b. Testing phylogenetic methods with tree-congruence: Phylogenetic analysis of polymorphic morphological characters in phrynosomatid lizards. Systematic Biology 47: 411–428.

Wiens, J. J. 1998c. Does adding characters with missing data increase or decrease phylogenetic accuracy? Systematic Biology 47: 625–640.

Wiens, J. J., and M. R. Servedio. 1997. Accuracy of phylogenetic analysis including and excluding polymorphic characters. Systematic Biology 46: 332–345.

Wiley, E. O. 1981. Phylogenetics: The Theory and Practice of Phylogenetic Systematics. John Wiley, New York.

Wilkinson, M. 1995. A comparison of two methods of character construction. Cladistics 11: 297–308.

Zelditch, M. L., W. L. Fink, and D. L. Swiderski. 1995. Morphometrics, homology, and phylogenetics: Quantified characters as synapomorphies. Systematic Biology 44: 179–189.