Syst. Biol. 50(5):689–699, 2001

Character Analysis in Morphological Phylogenetics: Problems and Solutions

JOHN J. WIENS

Section of Amphibians and Reptiles, Carnegie Museum of Natural History, Pittsburgh, Pennsylvania 15213-4080, USA; E-mail: wiensj@carnegiemuseums.org

Abstract.—Many aspects of morphological phylogenetics are controversial in the theoretical systematics literature and yet are often poorly explained and justified in empirical studies. In this paper, I argue that most morphological characters describe variation that is fundamentally quantitative, regardless of whether they are coded qualitatively or quantitatively by systematists. Given this view, three fundamental problems in morphological character analysis (definition, delimitation, and ordering of character states) may have a common solution: coding morphological characters as continuous quantitative traits. A new parsimony method (step-matrix gap-weighting, a modification of Thiele's approach) is proposed that allows quantitative traits to be analyzed as continuous variables. The problem of scaling or weighting quantitative characters relative to qualitative characters (and to each other) is reviewed, and three possible solutions are described. The new coding method is applied to data from hoplocercid lizards, and the results show the sensitivity of phylogenetic conclusions to different scaling methods. Although some authors reject the use of continuous, overlapping, quantitative characters in phylogenetic analysis, quantitative data from hoplocercid lizards that are coded using the new approach contain significant phylogenetic structure and exhibit levels of homoplasy similar to those seen in data that are coded qualitatively. [Character coding; morphology; phylogenetics; quantitative characters; weighting.]

Good science requires clearly explained, repeatable methods. Yet, practitioners of morphological phylogenetics tend not to be explicit about their methodology, specifically, how morphological characters are selected, and how states are defined, delimited, coded, and ordered (a process I refer to as "character analysis"). The lack of methodological explanation in published morphological studies has been discussed by several authors (e.g., Pimentel and Riggins, 1987; Pogue and Mickevich, 1990; Stevens, 1991; Thiele, 1993; Wiens, 1995) and has been documented for character selection (Poe and Wiens, 2000). This is a particularly serious problem, because in contrast to analysis of DNA sequence data, in which character definition and character state delimitation are virtually automatic (the nontrivial problem of alignment notwithstanding), morphological character analysis requires considerable effort, involving many methodological decisions and implicit assumptions at every step in the process.

Many aspects of morphological character analysis are controversial, including the way in which characters are constructed (e.g., Maddison, 1993; Pleijel, 1995; Wilkinson, 1995; Hawkins et al., 1997; Lee and Bryant, 1999; Strong and Lipscomb, 1999), whether intraspecifically variable characters can be included (Pimentel and Riggins, 1987; Nixon and Wheeler, 1990; Stevens, 1991; Campbell and Frost, 1993; Thiele, 1993; Wiens, 1995, 1998; Rae, 1998), how within-species variation is coded (Archie, 1985; Campbell and Frost, 1993; Thiele, 1993; Wiens, 1995, 1999; Swiderski et al., 1998 Smith and Gutberlet, 2001), how character states are ordered (Hauser and Presch, 1991; Lipscomb, 1992; Wilkinson, 1992; Slowinski, 1993), and how different types of morphological characters are weighted relative to each other (e.g., Farris, 1990; Campbell and Frost, 1993; Wiens, 1995, 1998). Different choices and assumptions are important because they can lead to radically different trees (e.g., Wiens, 1995).

In this paper, I suggest that for many morphological characters, these problems and controversies in the selection, definition, delimitation, and ordering of characters may have a common solution. Many, if not most, morphological characters describe variation in quantitative traits (e.g., differences in size, shape, or counts of serially homologous structures), regardless of whether systematists choose to code them quantitatively or qualitatively (Stevens, 1991; Thiele, 1993). Given this, three fundamental problems of character analysis (character state definition, delimitation, and ordering) potentially can be solved by simply

coding these quantitative traits as continuous, quantitative variables. I propose a parsimony method that allows quantitative traits to be analyzed directly as continuous variables (a modification of the gap-weighting method of Thiele [1993]). I then discuss the problem of scaling (or weighting) quantitative characters relative to qualitative characters and to each other and describe three possible solutions to this problem. I demonstrate the new approach to coding, using an empirical data set for hoplocercid lizards and show that phylogenetic results can be highly sensitive to different scaling methods. Finally, although many authors have advocated excluding continuous quantitative characters from phylogenetic analyses, I show that quantitative data from hoplocercid lizards coded by this new approach do contain significant phylogenetic signal and exhibit levels of homoplasy similar to those for data that are coded qualitatively.

TERMINOLOGY

There is often confusion surrounding the terminology of different types of morphological characters. As noted in Thiele's (1993) review, quantitative characters are described using numbers, whether those numbers describe the relative size or shape of a structure (morphometric characters) or a count of serially homologous traits (meristic characters, such as the number of teeth, limbs, or vertebrae). Qualitative characters are described with words (e.g., short, long, present, absent). Continuous characters are characters that can take on any real number value, whether a measurement of a morphometric character in a specific individual, or the mean value of an intraspecifically variable, quantitative trait for a given species (including meristic characters). Discrete characters are those that can take on only a limited subset of all possible values; these can refer to character states (e.g., 0, 1), or raw values for meristic traits (e.g., 20 maxillary teeth). In much of the systematics literature, however, "discrete" is often used to mean characters that show some degree of disjunction between species in ranges of within-species variation (i.e., they have nonoverlapping ranges), and "continuous" is used to mean characters that show little disjunction.

ADVANTAGES OF TREATING MORPHOLOGICAL CHARACTERS AS CONTINUOUS QUANTITATIVE VARIABLES

Morphological characters reported in the phylogenetics literature typically describe variation that is fundamentally quantitative, whether it is variation in relative size or shape of structures or in counts of meristic characters. However, as noted by Stevens (1991) and Thiele (1993), quantitative variation is often coded as discrete through qualitative description or use of a quantitative cutoff (e.g., state 0 = 2–4 scales; state 1 = 5–7 scales). Explicit quantitative coding methods, such as M-coding (Goldman, 1988) and gap weighting (Thiele, 1993), have been used by some systematists (e.g., Boughton et al., 1991; Chu, 1998; Gutberlet, 1998; Poe, 1998). Other systematists exclude quantitative characters because of so-called continuous variation (meaning extensive overlap in ranges of trait values between species), given the idea that such data are unsuitable for phylogeny reconstruction (e.g., Pimentel and Riggins, 1987). However, most published studies discuss neither how variation was coded nor what criteria were used for character selection (see review by Poe and Wiens, 2000).

I make three proposals. First, morphological systematists should explain clearly, and justify, their criteria for selection of characters and their methods of character analysis (i.e., defining, delimiting, coding, and ordering character states). Second, excluding characters because of overlapping ranges of intraspecific variation ("continuous") is unjustified. Intraspecifically variable, overlapping traits can contain useful phylogenetic information, whether the characters are coded quantitatively (Thiele, 1993; this study) or qualitatively (i.e., polymorphic characters; Wiens, 1995, 1998). Furthermore, the distinction between intraspecifically "fixed" and variable characters may be an artifact of small sample size and qualitative character definition (few characters will be intraspecifically invariant if defined quantitatively). Even though characters showing greater intraspecific variation tend to be more homoplastic (Archie, 1985; Campbell and Frost, 1993; Wiens, 1995), results from real and simulated data sets show that—given a finite sample of characters—including polymorphic

characters consistently increases phylogenetic accuracy relative to excluding them (Wiens and Servedio, 1997; Wiens, 1998). Third, coding quantitative variation as continuous quantitative characters (i.e., weighting character state transformations on the basis of differences in mean trait values using the method proposed in this study) may be preferable to qualitative coding because it can potentially solve three common problems in morphological phylogenetics. These problems are explained in the sections below.

Vague character definitions.—The language of many character state descriptions is extremely vague. For example, states are commonly described simply as "wide" versus "narrow", "small" versus "large", or "long" versus "short". The problem is that this description is not clear on how a given specimen is assigned a given state. This problem can be solved by defining the trait quantitatively.

Arbitrary character state delimitation.—Many systematists provide explicit quantitative criteria for determining whether a given specimen has a given character state for a given character. In many cases, however, it is not clear how the states were delimited. Typically, a range of values is given for each state, but usually without explanation for why a given set of ranges was chosen, or why a given number of intervals was used (e.g., state 0 = 1–3 scales, state 1 = 4–6 scales; state 0 = olecranon process <20% humerus length, state 1 process >40% humerus length). Similar cutoffs may also be defined by using qualitative morphological landmarks, such as the presence or absence of contact between two features and whether one structure is longer or shorter than another. In some cases, even the presence or absence of a feature (a "classic" qualitative character) may be an arbitrary cutoff for a broad range of continuous quantitative variation in the size of the feature (Poe and Wiens, 2000). Gift and Stevens (1997) have shown experimentally how different researchers can divide the same quantitative variation in very different ways, leading to very different character states. This problem can be solved by not using a cutoff at all and instead coding the character as a continuous variable.

Use of cutoffs or ranges may lead to three additional problems that have not been widely appreciated. First, considerable variation within character state ranges may be ignored. For example, given a character with state 0 consisting of 11–14 vertebrae and state 1 of 15–20 vertebrae, two hypothetical taxa with species values of 11 and 14 vertebrae would be coded as identical. Second, differences within intervals may be larger than between intervals. For the above character, a change from 11 to 14 vertebrae is ignored, but a change from 14 to 15 receives maximum weight. Third, use of cutoffs and ranges may not reflect the differences in the amount of change between character states. For example, some range-coded characters are treated such that all gaps between ranges are equal, when this is clearly not the case. If a character is coded such that state 0 = mean species values from 0 to 2, state 1 = 3– 4, and state 2 = 11–15, a simple unweighted analysis would not reflect the similarity between states 0 and 1 relative to state 2 (see also Fig. 1). Yet, the degree of similarity between values found in each species was the criterion used in delimiting the states in the first place. All of the problems noted above for the use of quantitative cut-offs are also potentially present in quantitative characters that are treated qualitatively, and all of them can be solved by treating quantitative characters as continuous.

Ordering of character states.—The question of whether or not to order character states is controversial. For many characters describing quantitative variation, systematists generally assume that trait values that are similar but not identical between taxa can be lumped into the same state. The assumption underlying this approach, that there is a special similarity between taxa with similar trait values, also supports analyzing these characters as continuous variables and provides a logical basis for ordering quantitative characters, regardless of whether they are coded qualitatively or quantitatively.

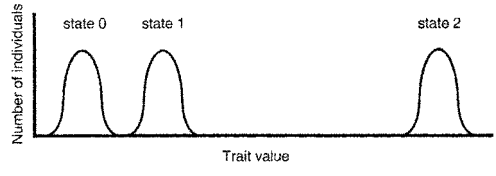

FIGURE 1. Hypothetical data showing differences between size of gaps between mean values of species, and the potential importance of gap-weighting.

Objections to Quantification

Most of the putative disadvantages of quantitative coding are shared with qualitative coding (Zelditch et al., 2000). For example, systematists may be concerned about how one derives characters from morphometric data, and if quantitative traits are correlated with body size (or with each other), exhibit variation caused by phenotypic plasticity, or require large sample sizes to be included. However, none of these potential problems are created by quantification; they exist independently of whether the characters are treated quantitatively or qualitatively (Zelditch et al., 2000).

Coding Continuous Morphological Variables Using Step-Matrix Gap-Weighting

Several authors have proposed treating intrinsically quantitative variables quantitatively and have developed various methods to do this (e.g., gap-coding: Mickevich and Johnson, 1976; generalized gap-coding: Archie, 1985; segment-coding: Colless, 1980; Thorpe, 1984; Chappill, 1989; M-coding: Goldman, 1988; gap-weighting: Thiele, 1993; finite-mixture coding: Strait et al., 1996; overlap-coding: Swiderski et al., 1998). Thiele (1993) proposed gap-weighting as a method for treating continuous variables as more-or-less continuous, by giving large weights to large differences in trait means between species, and small weights to small differences. Thiele's implementation of gap-weighting involves finding, for a given character, the mean value of the trait in each species in the analysis, the range of mean species values among taxa (i.e., the species with the greatest mean value and the species with the lowest), and then dividing this range into smaller ranges or segments equal to the maximum number of character states allowed by the phylogenetic software program (e.g., 32 for PAUP*). Species are then assigned states based on these ranges, and the character is ordered. Evolving from low to high mean trait values (or vice versa) therefore requires passing through many intermediate states and requires many steps, whereas smaller changes in trait values involve fewer state changes and fewer steps.

An important advantage of the gap-weighting method is that it incorporates information on the distance between states, weighting the changes according to the difference between mean species values (hence the name). For example, an analysis of the data in Fig. 1 using gap-coding, finite-mixture coding, or overlap-coding might reveal evidence for three character states. However, given that the degree of similarity between trait values is important and phylogenetically informative (in fact, it is the criterion used to delimit states in the first place), changes between states 0 and 1 should be much easier than evolving from either of these states to state 2. Yet, all methods but gap-weighting and segment-coding ignore this information.

Where Thiele's method falls short of treating continuous variables as continuous is in the limited number of states. I propose a method that circumvents this limitation by weighting the gaps between mean species values with step matrices. I call this approach step-matrix gap-weighting. For a given character, each taxon with a unique mean trait value is assigned a unique character state, and the costs of changes between these states are specified with a step matrix, based on the difference in mean trait values between each pair of species. The maximum cost between states in a step matrix is 1,000 in PAUP* (Swofford, 1998), and 999 in MacClade (Maddison and Maddison, 1992); using the largest value possible allows the most-fine-grained weighting. To implement the method for a given character, the mean trait value (x) for a given species is converted to a score (x_S) between 0 and 1,000 (or 999) by range-standardizing the data according to the following formula (from Thiele, 1993)

$$x_S = \frac{x - \min}{\max - \min} \times 1{,}000$$

where "min" is the minimum (lowest) mean species value of the trait across all species and "max" the maximum. The cost of a transformation between each character state (or taxon) in the step matrix is simply the difference between these scores. A simplified example of this coding method is shown in Figure 2, and a program to implement this method is available from the author.

Analysis of quantitative characters using step matrices does have some disadvantages, however. First, analyses are potentially

Taxon	Mean	Score	Code
species A	1.00	0	0
species B	2.21	403	1
species C	2.67	556	2
species D	4.00	1000	3

Step matrix

	0	1	2	3
0	0	403	556	1000
1	403	0	153	597
2	556	153	0	444
3	1000	597	444	0

FIGURE 2. Hypothetical example showing how a quantitative character is coded with step-matrix gap-weighting.

constrained by the number of distinct states allowed by the computer software package. This makes it difficult to include very large numbers of taxa (>32 for PAUP or PAUP*) with unique trait means. If the number of taxa with unique means is too large, I recommend using Thiele's (1993) gap-weighting method. This method uses less-fine-grained information but has no limits on the number of taxa that can be coded. Second, when using step-matrix gap-weighting, the only states that are reconstructed at ancestral nodes are those that occur within terminal taxa. However, to what extent (if any) this negatively impacts tree reconstruction is unclear, and simulations and congruence studies with polymorphic characters coded with frequency-based step matrices do not suggest that this problem limits phylogenetic accuracy (Wiens and Servedio, 1997, 1998; Wiens, 1998).

Some systematists may object to step-matrix gap-weighting because it requires assumptions about evolutionary processes. However, as stated above, this method is simply a logical extension of the same assumption that is widely used by morphological systematists when they code intrinsically quantitative characters. Morphological character states typically describe ranges of trait values, regardless of whether the states are defined quantitatively (e.g., state 0 = frontal process length >50% nasal length vs. state 1 = frontal process <40% nasal length) or qualitatively (e.g., frontal process long vs. short). Thus, systematists implicitly assume that taxa sharing similar but nonidentical trait values should be more closely related than taxa sharing more dissimilar trait values. They assume that traits will generally evolve gradually, rather than leaping from

low to high trait values and vice versa (i.e., they assume no a priori homoplasy in quantitative trait values). This assumption is little more than an extension of parsimony to character state definition; the minimum amount of change is assumed a priori. This assumption is also supported by the fields of empirical and theoretical quantitative genetics (Lynch and Walsh, 1998), which show that a character is generally more likely to evolve to a similar trait value (e.g., from a low mean number of ventral scales to a different low number) than to a dissimilar value (e.g., from a low to a high number of ventral scales).

Distance and likelihood methods have also been developed that can treat continuous morphological data as continuous (e.g., Felsenstein, 1981, 1988; Schluter, 1984; Lynch, 1989), and these methods may be advantageous relative to parsimony under some conditions (e.g., Wiens and Servedio, 1998). However, current applications of these methods do not readily allow for combining qualitative and quantitative traits, which may make them difficult to apply to many real data sets.

Scaling and Weighting Quantitative Characters

Not all characters are readily treated quantitatively, and a morphological analysis may contain a mixture of characters coded qualitatively and quantitatively (e.g., Chu, 1998; Gutberlet, 1998; Poe, 1998). How we weight or scale characters of different types relative to each other is an important issue that has received relatively little discussion (Farris, 1990; Thiele, 1993). For example, explicitly treating the relative length of a bone as a morphometric character with Thiele's (1993) gap-weighting method results in 32 ordered character states. Treating the same character as a qualitative trait (e.g., long vs. short) yields only two states. If the character is given equal weight relative to qualitative characters in both, the weight of the maximum change in the same character is 31 times greater when treated quantitatively rather than qualitatively. The problem is even worse with step-matrix coding; the maximum length of the character is 1 when treated qualitatively and 1,000 when treated quantitatively. These dramatic differences in weight clearly are unjustified. Three approaches might be used to adjust the weight of quantitative characters: between-character scaling, between-state scaling, and statistical scaling.

Between-character scaling.—Various authors have recommended weighting or scaling quantitative characters to be equal to each other and to qualitative characters (e.g., Thiele, 1993). The goal is to ensure that quantitative and (binary) qualitative characters have the same maximum length, an approach I label between-character scaling. For quantitative characters coded using step matrices with a maximum weight of 1,000, this equal weighting can be achieved simply by giving non-step-matrix characters a weight of 1,000. This seems to be a reasonable approach, particularly for morphometric characters.

The appropriate scaling for meristic characters is less clear. When meristic characters are viewed simply in terms of comparing mean species values, they are clearly continuous traits that are similar to morphometric characters. Under this view, between-character scaling may be most appropriate for meristic characters. When we consider the raw meristic data within species, meristic characters can also be viewed as discrete characters that are typically polymorphic and have many states. For characters involving the number of serially homologous structures, a continuum exists, running from binary characters, to multistate characters, to meristic characters; where a character falls on this continuum depends largely on the range of trait values, such that a small range implies a small number of states (Fig. 3). This continuum brings to mind Farris' (1990) question: Should meristic characters be downweighted merely because they have many states? For example, say that we observe taxa fixed for vertebral numbers of 10 and 11 among the species of a given group (Fig. 3). This is an

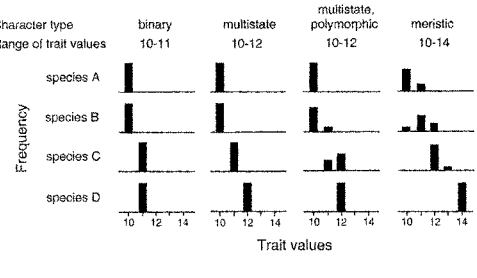

FIGURE 3. Hypothetical example illustrating the continuum from binary to multistate to meristic characters.

obviously discrete binary character that would have the same weight as any other traditional qualitative character. If we also observe taxa fixed for 12 and 13 vertebrae in the same group, and we assume the character is ordered, then applying between-character scaling to this character would make the cost (weight) of going from 10 to 11 vertebrae decrease to 33% of its original weight (i.e., because the cost of going from 10 to 13 is scaled to be equal to the cost of going from 0 to 1 in a fixed character, the cost of going from 10 to 11 decreases to one-third). But if the standard for weighting characters is the change in the frequency of adjacent character states from 0 to 100% (i.e., a change from 0 to 1 in a fixed character), then a change from 10 vertebrae in all specimens of a given species to 13 vertebrae in another should instead be equivalent to three changes between fixed traits, and should not be downweighted.

Between-state scaling.—In the context of step-matrix analysis of quantitative variables, what we need is a weighting scheme in which transformations between species with fixed, adjacent values of meristic variables (e.g., 10 to 11 vertebrae) receive the same weight as changes in binary variables (0 to 1,000), and more variable species with intermediate mean values (e.g., 10.5) receive proportionally intermediate weights. This can be accomplished by weighting each meristic character by the difference between the maximum and minimum mean species trait values (across all species in the analysis) for that character. I call this approach between-state scaling. An important advantage of this method relative to between-character scaling is that the cost of transformation between fixed, adjacent trait values (e.g., from 10 to 11 vertebrae) remains constant, regardless of the values that occur in other species.

A disadvantage of between-state scaling is that if the range of character state values is extremely high (e.g., 10 to 200), then characters weighted by this approach may have a very powerful influence on the phylogenetic results. Unfortunately, exactly where one draws the line in this continuum from multistate to meristic characters is unclear.

Statistical scaling.—A third approach that might be applied to scaling both morphometric and meristic characters is to combine the step-matrix gap-weighting method with the statistical methods designed for determining distinct states (e.g., finite-mixture coding, gap-coding). The statistical methods could be used to determine the number of distinct states for each character, and the number of distinct states minus one could be used as a weighting function for each step-matrix coded character. Using this method, the cost of a change between the lowest and highest mean species trait values would be equivalent to the maximum length of an ordered qualitative character; whether it was equivalent to a qualitative character with two states, four states, or more would depend on how many states were determined to be statistically distinct. The cost of changes between taxa with intermediate trait means would remain proportional to the difference between trait means, as for all characters coded using step-matrix gap-weighting. Thus, a character in which there are only two distinct states would receive a weight of one (equivalent to a fixed, binary character). A character with three statistically distinct states would receive a weight of two, such that changes from the lowest mean to the highest mean would have a cost of 2,000; this would be equivalent to two steps, or a change from 0 to 1 to 2 in a fixed, ordered, multistate, qualitative character. This weighting scheme, called statistical scaling, has the advantage of incorporating all the relevant information on the distance between species means, as well as some information on the variability of traits within taxa. However, this approach shares the same disadvantages of the statistical methods for character state delimitation. For example, if sample sizes are small or there are few gaps between taxa (despite a large difference in range of mean values between species), the character will receive little or no weight, even though these same restrictions are not applied to qualitatively coded characters.

In some ways, these three scaling methods do not really represent differential character weighing. Instead, they represent different ways of maintaining equal weights among characters, with each method based on a different concept of what the common currency of equal weighting should be, namely, overall character length (between-character scaling), transformations between fixed, discrete states (between-state scaling), or transformations between statistically distinct states (statistical scaling). The best overall currency is at present unclear. (Note: All three scaling methods are also applicable to data that are gap-weighted by Thiele's [1993] method.)

The uncertainty over the best scaling method, combined with the sensitivity of phylogenetic results to different scaling methods (Fig. 4), might be seen as a serious drawback of treating data quantitatively. But this is a case where quantitative analysis calls for explicit treatment of a general problem that is present but typically ignored with qualitative coding. For example, without quantitative methods for delimiting character states, a phylogenetic analysis of qualitatively coded data can be influenced strongly by a single character, if the author chooses to divide the character into a large number of states (in an analysis in which all character state transformations are given equal weight).

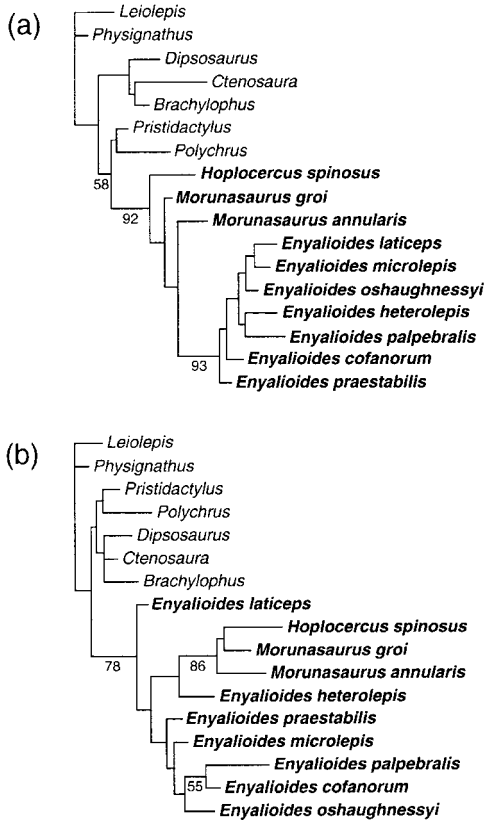

FIGURE 4. The impact of different weighting schemes for meristic characters on phylogenetic hypotheses for hoplocercid lizards. (a) Between-character scaling (meristic characters have the same maximum length as fixed, binary characters). (b) Between-state scaling (changes between numerically adjacent, fixed, trait values of meristic characters have the same length as fixed, binary characters). Numbers at nodes indicate bootstrap values >50%. Hoplocercid taxon names are in bold face. Outgroup taxa include acrodontans (*Leiolepis, Physignathus*), polychrotids (*Polychrus, Pristidactylus*), and iguanids (*Brachylophus, Ctenosaura, Dipsosaurus*). Monophyly of acrodontans, polychrotids, and iguanids was constrained during tree searches but not the relationships within or between them. See Appendices 1–3 at www.systbiol.org and Wiens and Etheridge (unpubl. manuscript) for further details.

AN EMPIRICAL EXAMPLE OF QUANTITATIVE CODING AND SCALING

I have recently applied the step-matrix gap-weighting approach outlined in this paper to a phylogenetic analysis of morphological data in hoplocercid lizards (Wiens and Etheridge, unpubl. manuscript). This is a family of 10 Neotropical species that are currently divided among three genera (*Enyalioides, Hoplocercus, Morunasaurus*). A total of 46 informative characters (squamation, coloration, osteology) were scored for 10 ingroup taxa and 7 outgroup taxa. Seventeen characters were qualitative and intraspecifically invariable, 19 were qualitative and polymorphic, 8 were meristic, and 2 were morphometric. Qualitative polymorphic characters were coded with the step-matrix frequency approach (Wiens, 1995, 1999; Berlocher and Swofford, 1997), and meristic and morphometric characters were coded with the step-matrix gap-weighting method described in the present paper. Two methods for scaling meristic characters were used: between-character scaling and between-state scaling. The third method (statistical scaling) outlined in the previous section was not attempted because of the very small sample sizes available for most species of hoplocercids, particularly for osteological characters. The list of characters, the traits means and frequencies for quantitative and polymorphic characters, and the coded data matrix are available as Appendices 1–3 at the Society of Systematic Biologists website (www.systbiol.org).

Several authors have stated that characters with extensively overlapping values between species (i.e., polymorphic, meristic, and morphometric characters) do not contain useful phylogenetic information and should therefore be excluded from phylogenetic analyses (e.g., Pimentel and Riggins, 1987; Stevens, 1991). I tested whether or not these three data types contained significant phylogenetic information relative to random data, using randomization tests on two measures of phylogenetic signal: the g_1 index (Hillis and Huelsenbeck, 1992) and

the consistency index (ci; Kluge and Farris, 1969). Seven data sets were analyzed: (1) all characters (meristic characters weighted with between-state scaling), (2) all characters (meristic characters with between-character scaling), (3) meristic characters only (with between-state scaling), (4) meristic characters only (with between-character scaling), (5) fixed characters only, (6) polymorphic characters only, and (7) morphometric characters only (with between-character scaling). Each data set was randomized 100 times, by randomly shuffling states among taxa within a given character (using a program supplied by J. P. Huelsenbeck). The number of states and the ordering and weighting of states and characters were maintained in each of the randomized data sets. The ci for each randomized data set was obtained by using a heuristic search to find the shortest tree (with tree-bisection-reconnection branch swapping and 20 random addition sequence replicates per search). The g_1 index for each randomized data set was calculated by taking a random sample of 10,000 trees from among all possible trees for that data matrix. For each of the original seven data sets, the 99% confidence interval of the mean g_1 index and ci was calculated for the 100 randomized data matrices. If the observed statistic (for the nonrandomized data) fell outside of this confidence interval, the data set was considered to contain significant, nonrandom phylogenetic information. To confirm that the phylogenetic structure occurred within the ingroup, the outgroup taxa were removed from all data sets for these analyses. In addition to the analyses of phylogenetic structure, I also qualitatively compared the average ci's of the different character types (i.e., fixed, polymorphic, meristic, morphometric) in the trees from the between-state scaling and between-character scaling. Step matrices were constructed using MacClade, and phylogenetic analyses were conducted with PAUP* (version 4.0.0d63). Support for individual branches was evaluated with nonparametric bootstrapping (Felsenstein, 1985; Hillis and Bull, 1993), using 500 pseudoreplicates per analysis with five random-addition sequence replicates per bootstrap pseudoreplicate.

Different methods for scaling meristic characters produced very different trees (Fig. 4). With between-character scaling, the tree within hoplocercids is highly incongruent with previous taxonomy but relatively similar to the only other phylogenetic study of the group (Etheridge and de Queiroz, 1988), with the genus *Enyalioides* forming a paraphyletic series of lineages at the base of the tree leading to a clade containing the genera *Morunasaurus* (which is paraphyletic) and *Hoplocercus*. In the tree based on between-state scaling, *Hoplocercus* and a paraphyletic *Morunasaurus* are at the base, and *Enyalioides* is a well-supported monophyletic group. These trees also differ in numerous placements of individual species of *Enyalioides*, although many branches of both trees (except for the monophyly of the family and the branch separating *Enyalioides* from the other genera) are relatively weakly supported.

The data sets have significant phylogenetic structure using both scaling methods as do the separately analyzed meristic, fixed, polymorphic, and morphometric characters (Table 1). Much of this signal may be concentrated in a few nodes, because many branches of both trees are weakly supported. Nevertheless, the average ci values among the different types of characters are generally similar (Table 2). Regardless of the scaling method used, the meristic characters have the highest ci values and the morphometric characters have the lowest.

This study demonstrates that the quantitative character data for these lizards do contain significant phylogenetic information,

TABLE 1. Results of randomization tests showing significant phylogenetic structure in different types of morphological data from hoplocercid lizards, using two statistics (g_1 and consistency index [ci]). The critical value refers to the 99% confidence interval from 100 randomized data matrices. When observed values for a given statistic fall outside the confidence interval for randomized data, the data are considered to contain significant phylogenetic structure.

Data type	Statistic	Observed value	Critical value
All data (between-state scaling)	g_1	−0.985	−0.214
	ci	0.625	0.565
All data (between-character scaling)	g_1	−0.925	−0.250
	ci	0.631	0.555
Meristic (between-state scaling)	g_1	−0.559	−0.304
	ci	0.700	0.672
Meristic (between-character scaling)	g_1	−0.930	−0.530
	ci	0.692	0.646
Polymorphic	g_1	−0.609	−0.240
	ci	0.577	0.510
Fixed	g_1	−0.646	−0.428
	ci	0.621	0.576
Morphometric	g_1		−1.014
	ci	0.829	0.778

TABLE 2. Consistency indices of the different types of characters for the trees in Figure 4 (n = number of parsimony-informative characters of each type).

Character type	n	Between-state scaling	Between-character scaling
Fixed	17	0.439 ± 0.238 ($0.167 - 1.000$)	0.308 ± 0.128 ($0.167 - 0.500$)
Polymorphic	19	0.406 ± 0.208 ($0.167 - 0.945$)	0.316 ± 0.237 ($0.142 - 0.945$)
Meristic	8	0.490 ± 0.057 ($0.417 - 0.599$)	0.349 ± 0.059 ($0.258 - 0.417$)
Morphometric	2	0.390 ± 0.110 ($0.312 - 0.467$)	0.307 ± 0.082 ($0.249 - 0.365$)

at least when coded with step matrices. The desire to avoid "continuous" variation is one of the most widely cited criteria for excluding characters in morphological phylogenetic studies (Poe and Wiens, 2000), and many authors have condemned the use of overlapping quantitative character data in phylogenetic analysis (e.g., Pimentel and Riggins, 1987; Stevens, 1991). Yet, the only other study to test statistically for phylogenetic structure (or lack thereof) in such data was Thiele's (1993) study in plants (genus *Banksia*). Thiele (1993) also found significant phylogenetic information in the quantitative characters he examined, and coded these characters with a method (gap-weighting with bins) similar to the step-matrix approach used in the present study. The results also show that levels of homoplasy in the meristic and morphometric data can be similar to those observed in the qualitative characters. In fact, the meristic characters are the least homoplastic set of characters in this analysis. The results of this study and the study of Thiele (1993) support the inclusion of overlapping meristic and morphometric data in phylogeny reconstruction.

SUMMARY

Many of the characters used by morphological systematists describe variation in continuous, quantitative traits, regardless of whether these traits are coded quantitatively or not. Given this view, there may be many advantages to treating these characters explicitly as continuous quantitative characters, and I have proposed new coding and scaling methods to implement this approach. Much work remains to be done on testing the accuracy of different coding and scaling methods and comparing the accuracy of parsimony, distance, and likelihood methods for quantitative traits. Congruence analyses (e.g., Wiens, 1998), which allow phylogenetic accuracy to be addressed with empirical data sets, should be particularly useful in this area.

ACKNOWLEDGMENTS

I thank Chris Beard, Maureen Kearney, Brad Livezey, Zhexi Luo, Dick Olmstead, Steve Poe, John Rawlins, Maria Servedio, Peter Stevens, and John Wible for comments on the manuscript, and Richard Etheridge for use of our hoplocercid data set for this paper.

REFERENCES

ARCHIE, J. W. 1985. Methods for coding variable morphological features for numerical taxonomic analysis. Syst. Zool. 34:326–345.

BERLOCHER, S. H., AND D. L. SWOFFORD. 1997. Searching for phylogenetic trees under the frequency parsimony criterion: An approximation using generalized parsimony. Syst. Biol. 46:211–215.

BOUGHTON, D. A., B. B. COLETTE, AND A. R. MCCUNE. 1991. Heterochrony in jaw morphology of needlefishes (Teleostei: Belonidae). Syst. Zool. 40:329–354.

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. Bull. Am. Mus. Nat. Hist. 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.

COLLESS, D. H. 1980. Congruence between morphometric and allozyme data for *Menidia* species: A reappraisal. Syst. Zool. 29:288–299.

ETHERIDGE, R., AND K. DE QUEIROZ. 1988. A phylogeny of Iguanidae. Pages 283–368 *in* Phylogenetic relationships of the lizard families: Essays commemorating Charles L. Camp (R. Estes and G. Pregill, eds.). Stanford Univ. Press, Stanford, California.

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

FELSENSTEIN, J. 1981. Evolutionary trees from gene frequencies and quantitative characters: Finding maximum likelihood estimates. Evolution 35:1229–1242.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791.

FELSENSTEIN, J. 1988. Phylogenies and quantitative characters. Annu. Rev. Ecol. Syst. 19:445–471.

GIFT, N., AND P. F. STEVENS. 1997. Vagaries in the delimitation of character states in quantitative variation—An experimental study. Syst. Biol. 46:112–125.

GOLDMAN, N. 1988. Methods for discrete coding of variable morphological features for numerical analysis. Cladistics 4:59–71.

GUTBERLET, R. L., JR. 1998. The phylogenetic position of the Mexican black-tailed pitviper (Squamata: Viperidae: Crotalinae). Herpetologica 54:184–206.

HAUSER, D. L., AND W. PRESCH. 1991. The effect of ordered characters on phylogeny reconstruction. Cladistics 7:243–265.

HAWKINS, J. A., C. E. HUGHES, AND R. W. SCOTLAND. 1997. Primary homology assessment, characters and character states. Cladistics 13:275–283.

HILLIS, D. M., AND J. J. BULL. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42:182–192.

HILLIS, D. M., AND J. P. HUELSENBECK. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. J. Hered. 83:189–195.

KLUGE, A. G., AND J. S. FARRIS. 1969. Quantitative phyletics and the evolution of anurans. Syst. Zool. 18:1–32.

LEE, D.-C., AND H. N. BRYANT. 1999. A reconsideration of the coding of inapplicable characters: Assumptions and problems. Cladistics 15:373–378.

LIPSCOMB, D. L. 1992. Parsimony, homology, and the analysis of multistate characters. Cladistics 8:45–65.

LYNCH, M. 1989. Phylogenetic hypotheses under the assumption of neutral quantitative-genetic variation. Evolution 43:1–17.

LYNCH, M., AND B. WALSH. 1998. Genetics and analysis of quantitative traits. Sinauer, Sunderland, Massachussetts.

MADDISON, W. P. 1993. Missing data versus missing characters in phylogenetic analysis. Syst. Biol. 42:576–581.

MADDISON, W. P., AND D. R. MADDISON. 1992. MacClade Ver. 3.0. Analysis of phylogeny and character evolution. Sinauer, Sunderland, Massachussetts.

MICKEVICH, M. F., AND M. F. JOHNSON. 1976. Congruence between morphological and allozyme data in evolutionary inference and character evolution. Syst. Zool. 25:260–270.

NIXON, K. C., AND Q. D. WHEELER. 1990. An amplification of the phylogenetic species concept. Cladistics 6:211–223.

PIMENTEL, R., AND R. RIGGINS. 1987. The nature of cladistic data. Cladistics 3:201–209.

PLEIJEL, F. 1995. On character coding for phylogeny reconstruction. Cladistics 11:309–315.

POE, S. 1998. Skull characters and the cladistic relationships of the Hispaniolan dwarf twig *Anolis*. Herpetol. Mon. 12:192–236.

POE, S., AND J. J. WIENS. 2000. Character selection and the methodology of morphological phylogenetics. Pages 20–36 *in* Phylogenetic analysis of morphological data (J. J. Wiens, ed.). Smithsonian Institution Press, Washington, D.C.

POGUE, M., AND M. MICKEVICH. 1990. Character definitions and character-state delimitations: The bete noire of phylogenetic inference. Cladistics 6:319–361.

RAE, T. C. 1998. The logical basis for the use of continuous characters in phylogenetic systematics. Cladistics. 14:221–228.

SCHLUTER, D. 1984. Morphological and phylogenetic relations among the Darwin's finches. Evolution. 38:921–930.

SLOWINSKI, J. B. 1993. "Unordered" versus "ordered" characters. Syst. Biol. 42:155–165.

SMITH, E. N., AND R. L. GUTBERLET, JR. 2001. Generalized frequency coding: a method of preparing polymorphic multistate characters for phylogenetic analysis. Syst. Biol. 50:156–169.

STEVENS, P. F. 1991. Character states, morphological variation, and phylogenetic analysis: A review. Syst. Bot. 16:553–583.

STRAIT, D., M. MONIZ, AND P. STRAIT. 1996. Finite mixture coding: A new approach to coding continuous characters. Syst. Biol. 45:67–78.

STRONG, E. E., AND D. LIPSCOMB. 1999. Character coding and inapplicable data. Cladistics 15:363–371.

SWIDERSKI, D. L., M. L. ZELDITCH, AND W. L. FINK. 1998. Why morphometrics is not special: Coding quantitative data for phylogenetic analysis. Syst. Biol. 47:508–519.

SWOFFORD, D. L. 1998. PAUP*: Phylogenetic analysis using parsimony*, Ver. 4.0.0d63. Sinauer, Sunderland, Massachusetts.

THIELE, K. 1993. The Holy Grail of the perfect character: The cladistic treatment of morphometric data. Cladistics 9:275–304.

THORPE, R. S. 1984. Coding morphometric characters for constructing distance Wagner networks. Evolution 38:244–255.

WIENS, J. J. 1995. Polymorphic characters in phylogenetic systematics. Syst. Biol. 44:482–500.

WIENS, J. J. 1998. Testing phylogenetic methods with tree-congruence: Phylogenetic analysis of polymorphic morphological characters in phrynosomatid lizards. Syst. Biol. 47:411–428.

WIENS, J. J. 1999. Polymorphism in systematics and comparative biology. Annu. Rev. Ecol. Syst. 30:327–362.

WIENS, J. J., AND M. R. SERVEDIO. 1997. Accuracy of phylogenetic analysis including and excluding polymorphic characters. Syst. Biol. 46:332–345.

WIENS, J. J., AND M. R. SERVEDIO. 1998. Phylogenetic analysis and intraspecific variation: Performance of parsimony, likelihood, and distance methods. Syst. Biol. 47:228–253.

WILKINSON, M. 1992. Ordered versus unordered characters. Cladistics 8:375–385.

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

ZELDITCH, M. L., D. L. SWIDERSKI, AND W. L. FINK. 2000. Discovery of phylogenetic characters in morphometric data. Pages 37–83 *in* Phylogenetic analysis of morphological data (J. J. Wiens, ed.). Smithsonian Institution Press, Washington, D.C.

Received 7 April 2000; accepted 23 November 2000
Associate Editor: A. Brower