Why Morphometrics is Not Special: Coding Quantitative Data for Phylogenetic Analysis

Donald L. Swiderski; Miriam Leah Zelditch; William L. Fink

Systematic Biology, Vol. 47, No. 3. (Sep., 1998), pp. 508-519.

Stable URL:
http://links.jstor.org/sici?sici=1063-5157%28199809%2947%3A3%3C508%3AWMINSC%3E2.0.CO%3B2-A

Systematic Biology is currently published by Society of Systematic Biologists.

sets: A case study in the living iguanas (Squamata, Iguanidae). Mol. Biol. Evol. 13:1087–1105.

STEEL, M. 1992. The complexity of reconstructing trees from qualitative characters and subtrees. J. Classif. 9:91–116.

TEHLER, A. 1995. Morphological data, molecular data, and total evidence in phylogenetic analysis. Can. J. Bot. 73:S667–S676.

VAN ZYLL DE JONG, C. G. 1987. A phylogenetic study of the Lutrinae (Carnivora; Mustelidae) using morphological data. Can. J. Zool. 65:2536–2544.

VANE-WRIGHT, R. I., S. SCHULZ, AND M. BOPPRÉ. 1992. The cladistics of *Amauris* butterflies: Congruence, consensus and total evidence. Cladistics 8:125–138.

VRANA, P. B., M. C. MILINKOVITCH, J. R. POWELL, AND W. C. WHEELER. 1994. Higher level relationships of the arctoid Carnivora based on sequence data and "total evidence." Mol. Phylogenet. Evol. 3:47–58.

WHEELER, W. C., P. CARTWRIGHT, AND C. Y. HAYASHI. 1993. Arthropod phylogeny: A combined approach. Cladistics 9:1–39.

WILKINSON, M. 1994. Common cladistic information and its consensus representation: Reduced Adams and reduced cladistic consensus trees and profiles. Syst. Biol. 43:343–368.

WILLIAMS, D. M. 1994. Combining trees and combining data. Taxon 43:449–453.

YODER, A. D. 1994. Relative position of the Cheirogaleidae in strepsirhine phylogeny: A comparison of morphological and molecular methods and results. Am. J. Phys. Anthropol. 94:25–46.

ZHANG, Z.-Q. 1995. A cladistic analysis of Trombidiidae (Acari: Parasitengona): Congruence of larval and adult character sets. Can. J. Zool. 73:96–103.

Received 22 July 1996; accepted 10 September 1997
Associate Editor: D. Cannatella

Syst. Biol. 47(3): 508–519, 1998

Why Morphometrics Is Not Special: Coding Quantitative Data for Phylogenetic Analysis

DONALD L. SWIDERSKI,[1] MIRIAM LEAH ZELDITCH,[2] AND WILLIAM L. FINK[3]

[1]*Museum of Zoology,* [2]*Museum of Paleontology, and* [3]*Department of Biology and Museum of Zoology, University of Michigan, Ann Arbor, Michigan 48109, USA; E-mail: dlswider@umich.edu (D.L.S.), zelditch@umich.edu (M.L.Z.), wfink@umich.edu (W.L.F.)*

Systematists often use qualitative descriptions of shape in phylogenetic analyses, but several biologists object to phylogenetic analyses using quantitative descriptions of those same shapes (Pimentel and Riggins, 1987; Felsenstein, 1988; Mickevich and Weller, 1990; Garland and Adolph, 1994). In a previous paper (Zelditch et al., 1995), we argued that the problem with phylogenetic analysis of quantitative shape data lies in the particular methods traditionally used to quantify shapes, not in quantification per se. In addition, we demonstrated that some of the more serious objections to using morphometric data in phylogenetic analyses are removed by using landmark-based morphometric methods developed by Bookstein (1991). Although we demonstrated that phylogenetic analysis of quantified shape variables is valid in theory, some practical problems remain. In this paper, we address the major remaining problem, that of coding: specifically, the problem of recognizing divergent character states.

Even a brief survey of the literature shows that coding is a complicated task in which several obstacles must be overcome (see Mickevich and Weller, 1990; Mabee and Humphries, 1993; Wilkinson, 1995). In this paper, we focus on one particular obstacle: evaluation of the diversity of a feature to determine which sets of taxa are similar in that feature. These judgements of similarity (and differences) are the foundations on which inferences of homology and monophyly are based. If these judgements employ inappropriate criteria, then those inferences are apt to be misled, and the resulting phylogeny is likely to be wrong.

Several biologists have argued that there can be no valid criteria for dividing quantitative data into discrete states because quantitative traits are inherently continuous (Pimentel and Riggins, 1987; Felsenstein, 1988; Garland and Adolph, 1994). In fact, they claim that coding quantitative data introduces artificial distinctions even if the observed distribution is discontinuous. This claim has even been parlayed

into arguments against cladistic parsimony (Felsenstein, 1988; Garland and Adolph, 1994), or against phylogenetic analysis of all morphometric data (Pimentel and Riggins, 1987; Mickevich and Weller, 1990). Therefore, we begin this paper by showing that the arguments against coding quantitatively described traits are not supported by theory. Rather, the obstacles posed by continuity are only practical problems and are not unique to quantitative data.

In the remainder of this paper, we address the practical problems of recognizing different states. Several systematists have proposed methods of coding that are designed to recognize states despite the lack of discontinuities between taxa. We review some commonly used methods and show that the criteria most methods employ to delimit states are not appropriate for phylogenetic analysis. Consequently, the character states produced by these methods do not support hypotheses of homology. We did find one method that is suitable for phylogenetic analysis, which we illustrate by using it to code features of adult body shape in six species of piranha. We prefer this method because it does not rely on arbitrary distance criteria or on statistical hypotheses that are irrelevant to the inference of homology.

CONTINUITY

Thiele (1993) suggested that some objections to coding quantitative data can be removed by making a distinction between terms that indicate how the trait was described and terms that indicate how the trait varies. Four terms, (qualitative, quantitative, continuous, and discontinuous) indicate how a trait is described. As did Wiley (1981), Thiele argued that quantitative should mean only that the trait was described by a numerical scale, i.e., by counting or measurement. In contrast, qualitative should mean only that the trait was described by using words. Continuous characters are a subset of quantitative characters, specifically those described by using an infinitely divisible numerical scale (e.g., real numbers). Discontinuous refers to the subset of quantitative characters that are described by using a numerical scale that is not infinitely divisible (e.g., integers). Used in this way, these terms imply nothing about how a trait varies. They imply nothing about biology because they refer only to our measurement scales.

Objections to coding morphometric data are not really concerned with the use of a continuous quantitative scale, but with what Thiele called "overlapping;" i.e., the range of variation of a trait in one taxon contains values that are also within the range of variation of that trait in another taxon. The contrasting pattern is disjunct, meaning that none of the values within the range of one taxon lie within the range of the other taxon. The words overlapping and disjunct can be applied whether the trait is described quantitatively or qualitatively; however, the comparison of ranges of qualitatively described features is necessarily subjective.

Morphometric data are usually reported as though the measurements were taken on a continuous scale. In reality, the scale of any instrument is discontinuous (e.g., 0.01, 0.02, 0.03, . . .), reflecting the limit of the resolving power of the instrument. A report utilizing the instrument's discontinuous scale is interpreted as an approximation to a continuous scale, not as an indication of steplike behavior of the character. Traits that are customarily reported on a discontinuous scale are counts for which fractional values are excluded conceptually (e.g., the number of teeth, for which incompletely formed teeth are either counted or not counted).

Thiele's discussion of the semantic issues clarifies the point that chains of overlapping ranges are the primary obstacle to coding morphometric traits. However, Thiele, as did Stevens (1991), also pointed out that this problem is not unique to quantitative data. In fact, Thiele and Stevens argued that this is one of many of the problems that are the same for quantitative and qualitative data. For example, one issue that must be resolved for every feature is the comparability of that feature across all taxa in the study; another is the recognition of distinct conditions of the feature. Thus, Thiele and Stevens argued for applying the same criteria to quantitative and qualitative data, and for making the criteria explicit for all data.

Pimentel and Riggins (1987) were quite explicit. They argued that features with overlapping ranges should not be coded as having distinct states. This position is evident from

their definition of a cladistic character as "a feature of organisms that can be evaluated as a variable with two or more *mutually exclusive* and ordered states" (p. 201, emphasis added). It is also clear that this definition was meant to apply to all kinds of data, because Pimentel and Riggins stated it at the beginning of the paper, and again, in their discussion of quantitative data (p. 207). In the latter context, they elaborated on the requirement that characters have mutually exclusive states and argued that the only valid basis for coding any character is a gap, a hiatus in the distribution of a character, such that no individuals are observed to have those values. In their view, the ideal case would be a gap between ranges (Thiele's "disjunction"). Pimentel and Riggins did allow coding if a few taxa contain individuals that are on each side of the gap (these taxa would be polymorphic), but no taxon can have individuals within the gap. For Pimentel and Riggins, the gap is absolutely required for coding because it unambiguously demarcates mutually exclusive sets of values, without any statistical or mathematical manipulation. They characterized these gaps as "natural" (p. 207), implying that any distinction which is not based on disjunction in the raw data is artificial.

Felsenstein (1988) agreed that division of overlapping ranges into separate states creates artificial distinctions. However, he argued that coding based on observed gaps also imposes artificial distinctions. Felsenstein claimed this argument is supported by theoretical predictions that polygenic characters will exhibit gradual, incremental change. Thus, even if a trait evolves rapidly, it still passes from one value to the next with no values skipped (i.e., saltation does not occur). This implies that a descendant population will overlap the immediately ancestral population. From this implication, Felsenstein inferred that disjunction of terminal taxa represents missing data (e.g., unrecovered fossils), because if all the ancestral populations were known they would form links in an unbroken chain connecting the terminals. Felsenstein argued that observed gaps between terminals are not real and should not be used as the foundation for any coding scheme.

Felsenstein's argument about the reality of gaps overextends a legitimate theory. That theory describes anagenetic change, transformation in a single unbranched lineage. However, lineages branch (speciation occurs), and the branches are genetically and evolutionarily independent. Because they are independent, the chains representing the descendant lineages will eventually become distinct from each other, as individuals within the lineages acquire novelties. This divergence is simply a consequence of independent evolution within separate lineages. The unbroken links of the chain connect ancestors to descendants, not terminal taxa to each other.

We conclude that there is no obstacle in theory to coding taxa with overlapping ranges. In fact, Felsenstein's argument provides grounds for us to argue that phylogenetic systematists need an approach to coding that does not require gaps. The ranges of populations representing nascent branches can be expected to overlap each other and the range of their common ancestor. Obviously, a gap would be useful, but a lesser amount of differentiation can also indicate evolutionary independence. The goal of phylogenetic systematics is to infer evolutionary independence (branching) from evidence of divergence. When divergence is relatively small and ranges overlap, the real obstacle to coding is distinguishing between differences due to poor sampling and differences due to evolution. We address this problem below.

Methods of Coding

In this section, we review some of the most widely used methods of coding. For each method, we focus on the criterion used to divide a series of taxa with overlapping ranges into smaller groups, and on the validity of that criterion as a basis for inferring homology.

We begin with gap coding (Mickevich and Johnson, 1976), both because it is one of the oldest methods of coding and because most newer methods of coding are intended to improve on gap coding. This method is illustrated with the five hypothetical populations shown in Figure 1. The smallest mean is assigned state 0. The next largest mean is assigned a new state only if the difference between means is greater than the value of the pooled standard deviation (s_p). Then the third mean is compared to the second, and so

on, until all pairs of adjacent means have been evaluated. In this context, "gap" refers to the difference between means, not the disjunction between ranges.

The principal problem with gap coding is that it provides a small amount of unreliable information from which to judge the similarity of taxa. The information is the similarity of means, as indicated by s_p. This information is unreliable because variances of taxa are often dissimilar, making s_p a poor indicator of the actual overlap between two taxa. Some will overlap more than expected; others, less. Gap coding also misrepresents the amount of overlap when distributions are skewed or otherwise deviate from normality. For these reasons, we do not recommend gap coding as a basis for inferring similarity of taxa.

Most critiques of gap coding focus on other problems (e.g., Thorpe, 1984; Archie, 1985; Chappill, 1989). One common complaint is that taxa may be quite different but still be assigned the same state because they are ends of a long series of closely spaced taxa. This problem is illustrated in Figure 1 by taxa B, C, and D. The distance between B and C is small ($<s_p$), as is the distance between C and D. Consequently, all three taxa are assigned the same state, even though the distance between B and D is large ($>s_p$).

To solve this problem, Archie (1985) proposed a method of defining subsets of taxa and a method of coding overlapping subsets. To define a subset, Archie used the mean of a taxon (x_i) and s_p to define an interval (x_i to $x_i + s_p$). The subset includes all taxa with means in that interval. For Figure 1, the subsets would be {A}, {B, C}, {C, D}, {D}, and {E}. The method of coding begins by deleting any subset that is completely included in another,

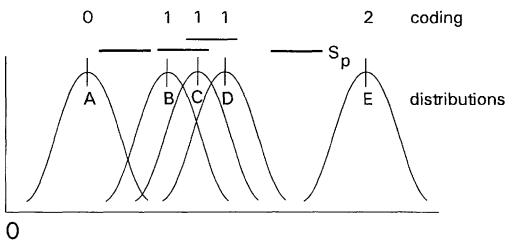

FIGURE 1. Gap coding of five idealized populations (A–E).

such as {D}. Then, state 0 is assigned to the subset that includes the lowest mean, {A}. Codes increment by 1 at the beginning or end of a subset (or by 2 if the subsets are disjunct). Thus, the Archie coding for Figure 1 would be A = 0, B = 2, C = 3, D = 4, E = 6.

Archie's methods solve the problem of long series, but still rely on the dubious information provided by the mean and s_p. More importantly, Archie's methods have a serious problem of their own. Farris (1990) criticized Archie's methods because subsets are defined by some criterion in the first step, but then that criterion is ignored in the second step. In the example above, B and C are placed in the same subset because they are not different from each other, but are assigned different states because C is not different from D. The inconsistent logic is particularly clear when subsets are defined by statistical analyses, as in homogeneous subset analysis (Simon, 1983; Farris, 1990). Using Archie coding on these subsets would assign different states to taxa despite statistical tests showing that their means are not significantly different.

A somewhat different solution to the problem of long series is incorporated in methods proposed by Colless (1980), Thorpe (1984), and Chappill (1989). In these methods, the morphometric distance between the most widely separated means or individuals is divided into two or more equal segments. The segments are numbered in order, and the code assigned to a mean or individual is the number of the segment in which it is located (Fig. 2). In effect, segment coding rescales the original measurements to a smaller number of larger increments.

Segment coding solves the problem posed by long chains of closely spaced means. However, it replaces that problem with a more fundamental one. It distorts the similarities and differences among the taxa. Means or individuals near the limits of a segment may be more similar to those in the adjacent segment than they are to the ones in their own segment. In Figure 2, C is no closer to B than it is to D, but B and C are assigned the same state and D is assigned a different state. Because segment coding does not reflect similarity, it cannot be used as a basis for inferring homology.

Archie coding and segment coding create bigger problems than the one they solve. This

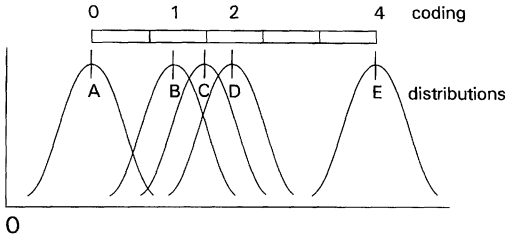

FIGURE 2. Segment coding of five idealized populations (A–E).

is because these methods are designed to solve the wrong problem. That problem, failure to distinguish between taxa at the ends of a long series, does not arise from a defect of gap coding; rather, gap coding reveals the reality that sometimes intermediate taxa bridge the difference between the ends. This problem does not have a solution. Any consistent method that evaluates ranges of variation will encounter cases in which intermediate taxa form a bridge between taxa that would otherwise be considered different. Such cases may bring attention to the criterion used to judge whether ranges are similar or different, but they are not grounds for replacing evaluation of similarity with computation of arbitrary distance metrics.

Farris (1990) argued that the criterion of similarity used in gap coding (s_p) should be replaced with explicit statistical tests of the differences between means. Then, a series of pairwise tests (with appropriate adjustments for multiple comparisons) is used to construct homogeneous subsets: groups in which no two sample means are statistically significantly different (Simon, 1983). Because homogeneous subset coding uses an explicit statistical test of similarity, rather than an unreliable indicator of overlap, it eliminates one source of error that affects gap coding. In addition, this method eliminates some ambiguity by using a statistical test, rather than a proxy for a statistical test. Taxa in mutually exclusive subsets can be assigned different codes because the mean for each taxon in one set *is* significantly different from the mean of each taxon in the other set. Equally obvious, homogeneous subsets that intersect (share taxa) cannot be assigned different states because the means of the shared taxa cannot be distinguished from the means of any taxa in either set.

The improvements incorporated in homogeneous subset coding are important, but the flaw it retains is more important. Like gap coding, homogeneous subset coding uses a minimal description of the variability within each taxon: the mean and standard deviation. Consequently, both methods of coding are prone to errors when the observed distribution within a taxon departs from the expected normal distribution. This is not a trivial or purely formal objection. Several factors may account for deviation from normality, and many of them are commonly encountered in systematic studies (e.g., allometry, geographic variation, sexual dimorphism, biased collecting methods). Additional sources of biased distributions may be encountered when multiple species are combined into higher taxa (e.g., in studies of evolutionary trends or differential extinction). Given these common sampling problems, it is crucial that a method of coding uses as much information as possible about the distribution of individuals within each taxon.

Almeida and Bisby (1984) also recognized that coding should be based on more information than a comparison of means and standard deviations. They used box plots to show the entire range of each species divided into quartiles. Figure 3a shows box plots for five hypothetical taxa similar to those in Figure 1. Almeida and Bisby used the box plots to find regions where there was no overlap (Fig. 3a, zone a) and regions where only the outer quartiles of the taxa overlap (Fig. 3a, zone b). These regions delimit the sets of taxa that can be assigned the same character state code for that trait.

Almeida and Bisby's use of quartiles is an improvement over the other methods because it conveys some information about deviations from normality. However, it produces a coarse-grained analysis, in which taxa that overlap as much as 25% can be assigned different character states. Almeida and Bisby were uncomfortable with allowing this much overlap (p. 408), as are we. This problem could be remedied by using a different cutoff (e.g., outer 5th percentiles), but a more important problem would remain: the lack of any a priori justification for applying a fixed standard to all comparisons. Unfortunately, the use of a fixed standard is probably unavoidable when the distribution

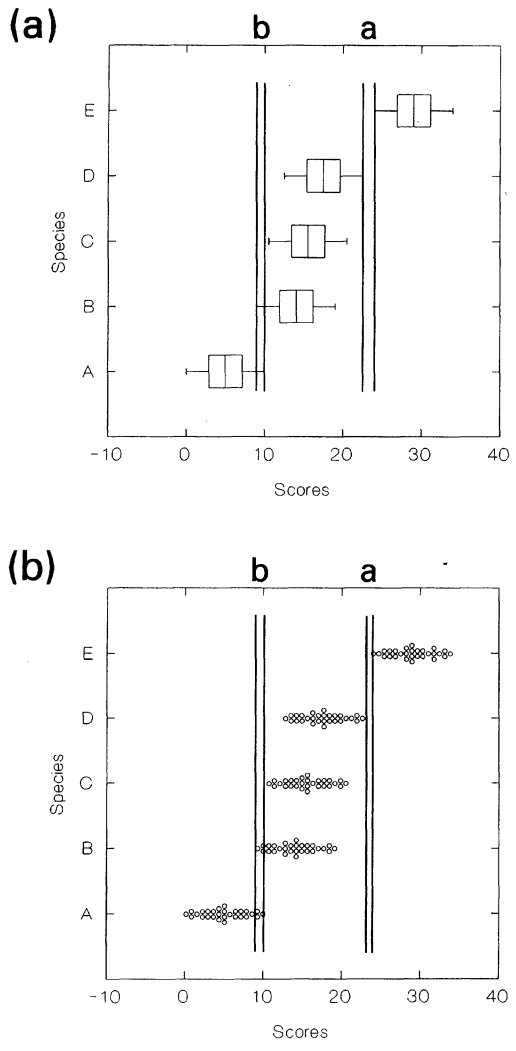

FIGURE 3. Coding based on overlap of taxon ranges. (a) Box plots dividing ranges into quartiles. (b) Dot plots of individual scores. The paired lines denote: a) the gap between taxa D and E, b) the overlap between taxa A and B.

is represented by a range bar divided into categories, as done by Almeida and Bisby. Such graphs omit the number and distribution of individuals in each sample, forcing systematists to base coding decisions on the number and size of the overlapping categories. Without a rule for these decisions, systematists are likely to base coding on subjective impressions of the pattern of overlap among all taxa, which is even less justified (Gift and Stevens, 1997).

To improve Almeida and Bisby's method, we suggest using dot plots (Fig. 3b), which are essentially symmetrical histograms with large numbers of small intervals. These diagrams illustrate the spread of individuals rather than the clustering shown by conventional histograms. Dot plots produce finer-scale descriptions, which allow coding decisions to be based on analysis of the individuals in the study, not on the numbers of individuals within coarse classes.

Almeida and Bisby's approach, with our modification, puts the coding of quantitative data on the same footing as traditional analysis of qualitative data: The diversity within each group is evaluated and then is compared with the membership of other groups to see if there is overlap. Then, the overlap is evaluated to determine whether a hypothesis of evolutionary transformation is justified. Our preference is to determine whether the density of individuals decreases near the edge of the observed range of a taxon, which would be an indication that the edge of the observed range is close to the edge of the actual range. If overlap involves only individuals from these fringes, then we would recognize different states. Other systematists may prefer different criteria; one advantage of the dot plots is that readers can apply their own criteria to the same data.

One problem that is not solved by using dot plots is the one caused by intermediate taxa overlapping the ranges of taxa that do not overlap each other. In Figure 3b, the ranges of taxa A and C do not overlap, but the range of taxon B overlaps both. This problem cannot be solved by any method that consistently applies a criterion for recognizing differentiation. If B cannot be distinguished from either A or C, the character should be considered phylogenetically uninformative for those taxa.

Some systematists may prefer a method of coding that incorporates a rigid, automatic criterion for recognizing different states. We have not proposed any such rules, because none can be realistically applied to all cases. Several of the methods discussed above represent attempts to employ rules; their failures demonstrate that the rules do not apply universally. We see no reason to obey rules to code quantitatively described traits when we would not obey those rules to code the same traits if they

were qualitatively described. Instead, we suggest that the coding of each trait should be decided on its own merits, by examining the distribution of individuals in each taxon.

Analysis of Piranhas

Methods

Below, we illustrate overlap coding of data from real populations rather than the hypothetical constructions used above. Most of these data come from analyses of the ontogeny of piranha shape (Fink and Zelditch, 1995; Zelditch and Fink, 1995). Descriptions of the morphometric methods, including a selection of landmarks, are presented in those papers (see also, Bookstein, 1989, 1991; Zelditch et al., 1992; Swiderski, 1993). Because our purpose is to demonstrate the coding method, not the morphometric methods, here we present only a brief description of the morphometric methods, highlighting departures from previous studies or details that are particularly relevant to coding.

Fink and Zelditch (1995) analyzed ontogenetic shape change in five species: *Pygopristis denticulata, Serrasalmus gouldingi, Pygocentrus cariba, Pygocentrus nattereri,* and *Pygocentrus piraya.* In this study, we use the adults from that study, and add new data by including the adults of a sixth species, *S. elongatus.* We define adults as specimens with centroid size >100 (corresponding to a standard length >75 mm, which is approximately the size at which the juvenile phase of growth ends (centroid size is defined by Slice et al., 1996). We restricted this study to adults because many studies are limited to adults, given the difficulties of obtaining juveniles, and because the description of ontogenies adds several problems that are beyond the scope of this demonstration (cf., Zelditch et al., 1992; Mabee and Humphries, 1993; Fink and Zelditch, 1995).

Shape was described by using the thin-plate spline analysis, which can be implemented with either of the following programs: F. J. Rohlfs TPSPLINE or J. M. Humphries JSPLINE (both are available at http://life.bio.sunysb.edu/morph/).

Each adult in this study was compared to the same starting form, an average juvenile of the outgroup, *Pygopristis denticulata.* Even in studies of allometry, comparison of adults to the juvenile of an outgroup is unusual, but we do it here because the starting form defines the variables used in the morphometric analysis (principal warps). By using the same starting form that was used in the ontogenetic studies, we insure that our descriptions of adult shape can be compared to the descriptions of shape ontogenies.

Principal warps differ from conventional measurements in many ways (see Bookstein 1991; Zelditch et al., 1992, 1995; Swiderski, 1993), but one difference that is particularly relevant here is that principal warps are two-dimensional variables. The observed values, called *partial* warps, reflect not only the magnitude of shape change, but also its direction with respect to the organism. Partial warps commonly are reported as x, y coordinate pairs, representing amounts of change in two directions of an orthogonal grid system. We aligned the starting form so that x is the anteroposterior axis and y is the dorsoventral axis.

The results of the spline analysis, the partial warps scores, are presented in two formats for coding. The first format is one-dimensional, and describes the anteroposterior and dorsoventral components separately. In this format, the distributions of individuals are displayed by dot plots, as suggested above. The second format is two-dimensional, in which the dot plots are replaced with scatter plots showing the distributions of the anteroposterior (x) and dorsoventral (y) components jointly. (Other methods of coding can also be adapted for use with two-dimensional data, by computing a set of ellipses or computing an appropriate multivariate test statistic. The logic of our argument favoring overlap coding was not contrived so as to favor the only method that could be applied to two-dimensional data.)

Results

Figure 4a shows the pattern of landmark displacements for the largest-scale principal warp of the starting form. In this pattern, landmarks near the middle move in one direction, and landmarks near the ends move in the opposite direction. This pattern is illustrated with an arbitrarily chosen $+y$ multiplier to show the proportions of relative displacements that this

component would represent. This pattern of landmark displacement can be described verbally as a change in dorsal convexity. Figure 4b shows the observed scores for this dorsoventral component of shape change for each individual in this study. A score of zero indicates that, in this component, the specimen is not different from the average juvenile *Pygopristis denticulata*. Figure 4b indicates that the ranges for these six species are very similar. Based on their broad overlaps, we infer that there has been no differentiation of this feature in the dorsoventral direction.

The scores in Figure 4b represent only one component of the changes described by the pattern in Figure 4a. The same pattern oriented in the anteroposterior direction (all arrows rotated 90° clockwise) represents a graded pattern of change in which one end of the body is expanded relative to the other end. Positive scores represent anterior elongation; negative scores, posterior elongation. Figure 4c shows the distribution of scores for this feature. There is a broad overlap between *Pygocentrus* and *Serrasalmus*. Line A marks the edge of the *Pygocentrus* range, and a third or more of each *Serrasalmus* species is on the left of this line. *Serrasalmus* also overlaps *P. denticulata*: Two specimens of *S. elongatus* are on the right of line B and two *Pygopristis denticulata* are on the left of the line. This distribution could be interpreted as indicating two evolutionary transitions, one across each line, but neither line unambiguously demarcates two sets of taxa. In both cases, the source of ambiguity is uncertainty about the limits of the *Serrasalmus* species. For example, at line B the density of *Pygopristis denticulata* specimens drop abruptly, and the two individuals to the left of the line can be reasonably interpreted as lying on the fringe of the distribution. In contrast, the two *S. elongatus* on the right of the line cannot be interpreted as lying on the fringe of their distribution, because the sample size for *S. elongatus* (13), is too small to reliably infer the distribution of within-group variation. The two *S. elongatus* on the right of the line may only appear to be unusual because variation within that species is not adequately described. Consequently, we do not feel that these distributions justify an inference of separate character states for this shape feature.

Figure 4d shows the two-dimensional distribution of shape changes described by the pattern in Figure 4a. A specimen with coordinates (0,-0) would not be different from the average juvenile *Pygopristis denticulata* in this feature in any direction. Coordinates of $(+x, +y)$ would indicate that the specimen differs from the juvenile *Pygopristis denticulata* in both greater anterior elongation and greater dorsal convexity. As in the previous plots, the two-dimensional plot shows that this component of shape varies almost entirely in the anteroposterior direction, and that the species ranges overlap too broadly to recognize separate states. The ranges of the three *Pygocentrus* species are nearly identical. Most specimens of the two *Serrasalmus* species, and all specimens of *Pygopristis denticulata* lie outside of the *Pygocentrus* range, but several specimens of both *Serrasalmus* species lie within the *Pygocentrus* range. More importantly, some *Serrasalmus* are found near dense clusters of *Pygocentrus* individuals, well beyond the edge of the *Pygocentrus* range. Few *Serrasalmus* are found near clusters of *Pygopristis denticulata*, but the widely scattered *S. elongatus* surround much of the *P. denticulata* range. The relatively sparse distribution of *S. elongatus* suggests that additional samples should be expected to have individuals that fall within the *P. denticulata* range. Consequently, we conclude that *S. elongatus* bridges the gap between *Pygopristis* and *Pygocentrus*, and that this shape feature is not phylogenetically informative.

Figure 5a shows the pattern of landmark displacement for a small-scale feature localized to part of the head. The dorsoventral scores for this feature (Fig. 5b) show broad overlaps for all species. The anteroposterior scores (Fig. 5c) indicate that there is some differentiation between *Pygopristis denticulata* and the other species along this axis (relative length of the snout and jaws). The two-dimensional plot (Fig. 5d) provides better evidence of a shape change. All *Pygopristis denticulata* but one are on the left of the line, whereas all *Serrasalmus* and all *Pygocentrus* but one are on the right of the line. In addition, a sparsely populated space lies to the right of this line. Based on the space between the two groups and the small number of specimens that have crossed that space, we recognize two states for this shape feature (one unique to *Pygopristis denticulata*).

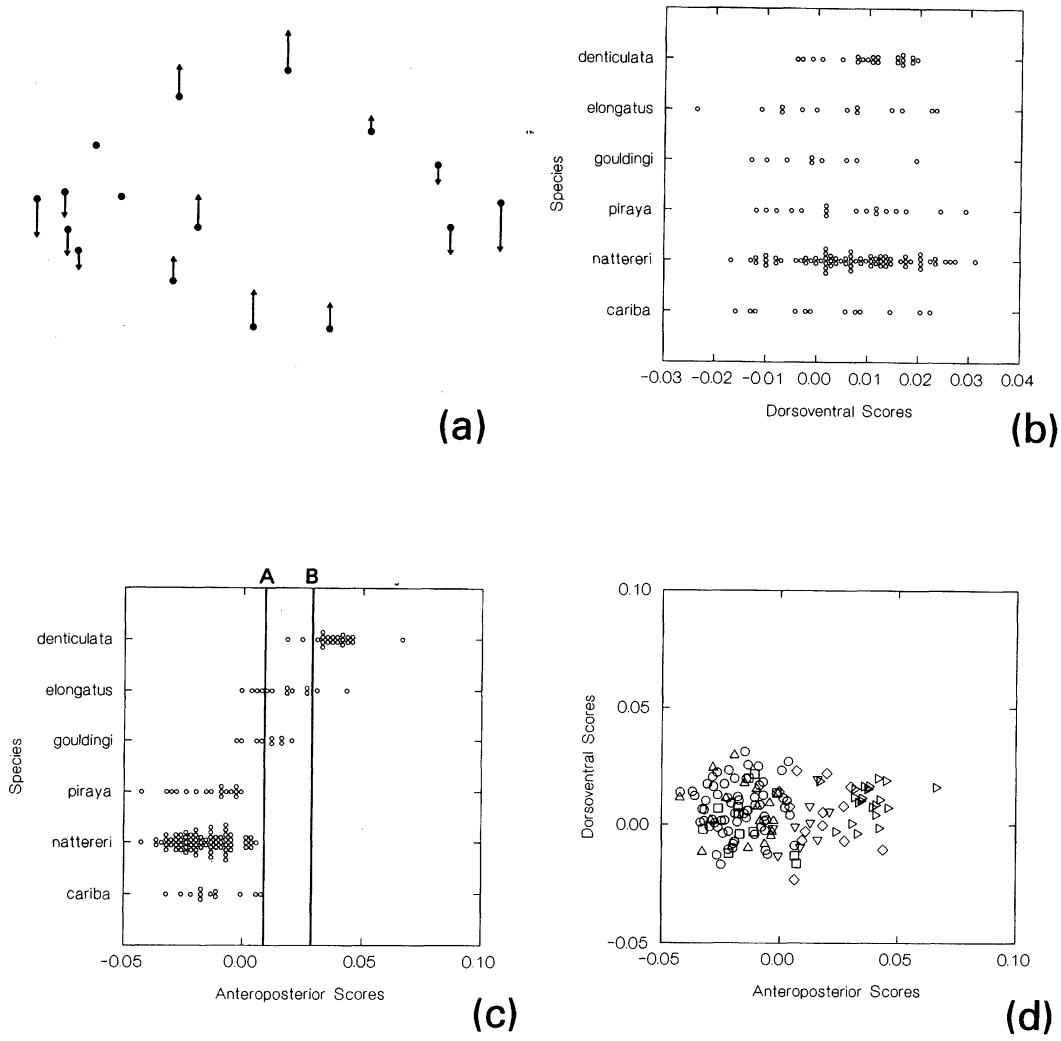

FIGURE 4. Variability of a large-scale principal warp. ▷ = *Pygopristis denticulata*, ◇ = *S. elongatus*, ▽ = *S. gouldingi*, ○ = *Pygocentrus. nattereri*, △ = *Pygocentrus piraya*, □ = *Pygocentrus cariba*. (a) Pattern of landmark displacement. (b) Partial warp scores for displacement along the dorsoventral axis. (c) Partial warp scores for displacement along the anteroposterior axis. A marks one edge of the *Pygocentrus* range; B separates most *Pygopristis denticulata* from most *Serrasalmus*. (d) Bivariate plot of anteroposterior and dorsoventral partial warp scores.

These two states could not be recognized in the one-dimensional plot because the direction of change is not aligned with the anatomical axes.

Figure 6a illustrates a small-scale feature that describes changes in the region extending from the base of the dorsal fin through the caudal peduncle at the base of the tail fin. The one-dimensional plots for this feature have been omitted; the two-dimensional plot (Fig. 6b) again indicates a change that is not aligned with the anatomical axes (across line B). However, the main reason we show this feature is because it appears to have transformations in two different directions. The ranges of *Serrasalmus* and *Pygocentrus* are completely disjunct. Both overlap the range of *Pygopristis denticulata*, but from different sides, and neither overlap is enough to prevent recognition of distinct states. Based on these distributions, we infer two independent character changes.

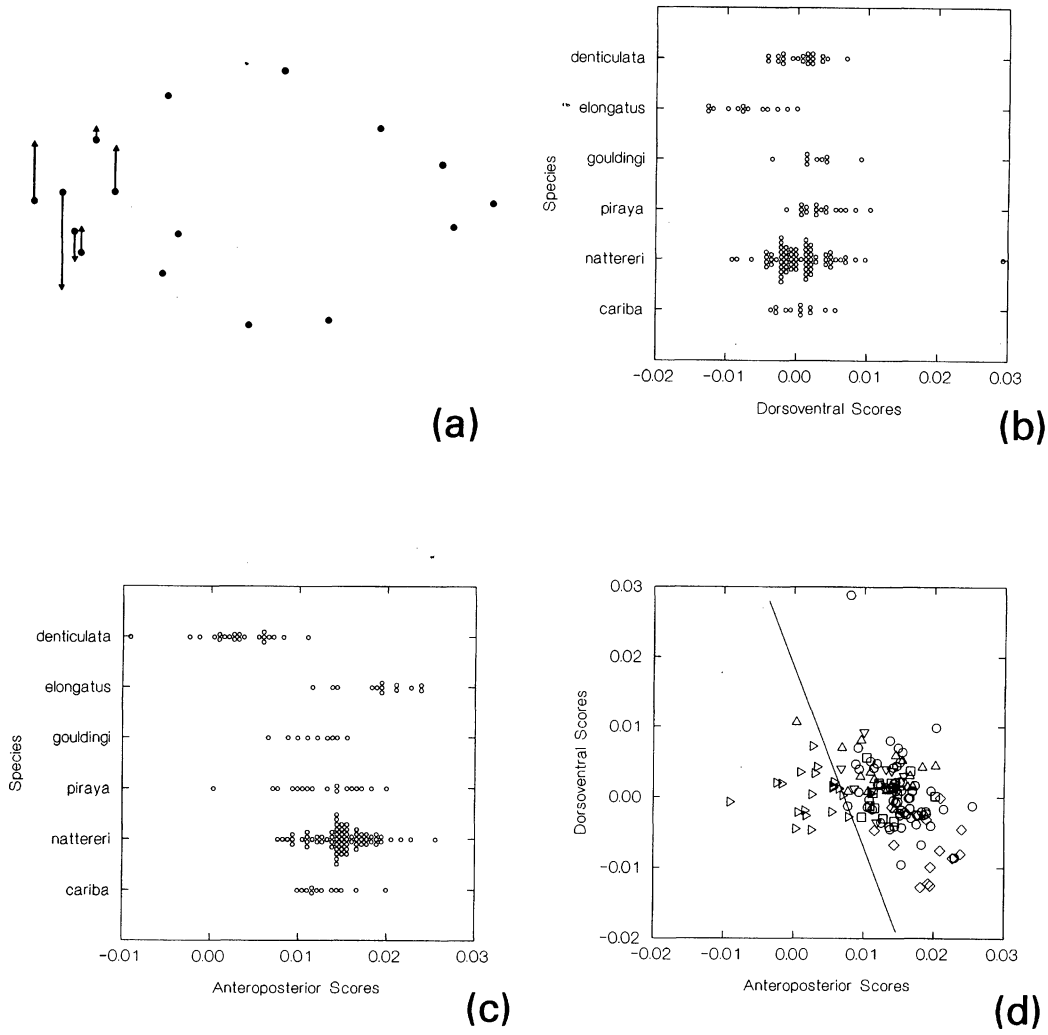

FIGURE 5. Variability of a small-scale principal warp. (a) Pattern of landmark displacement. (b) Partial warp scores for displacement along the dorsoventral axis. (c) Partial warp scores for displacement along the anteroposterior axis. (d) Bivariate plot of anteroposterior and dorsoventral partial warp scores. The line delimits the sets of taxa inferred to have different character states. Symbols as in figure 4.

The anteroposterior transformation of *Serrasalmus* is primarily a relative elongation of the region between the dorsal fin and the caudal peduncle. In contrast, the dorsoventral transformation of *Pygocentrus* is primarily a relative thickening of the caudal peduncle.

There are several other features we could show, but these three are sufficient to demonstrate the approach we advocate, for both one-dimensional and two-dimensional characters. Our preliminary analysis of the distribution of 14 shape features indicates there may be more than 10 shape transformations among these six species. Six of the changes are in features that underwent two changes. Some of the changes may be autapomorphies, but at least half are potentially informative for resolving phylogenetic relationships.

SUMMARY

Thorpe (1984) and Chappill (1989) argued that selection of a coding method should be

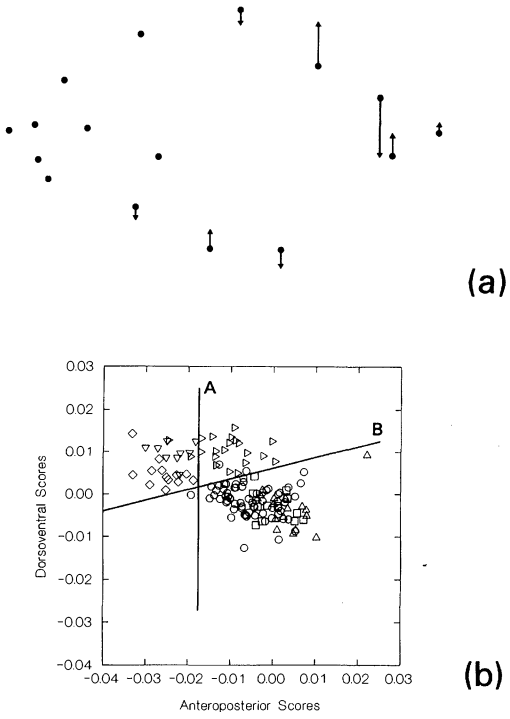

FIGURE 6. Variability of another small-scale principal warp. (a) Pattern of landmark displacement. (b) Bivariate plot of anteroposterior and dorsoventral partial warp scores. Lines A and B delimit sets of taxa inferred to have different character states. Symbols as in Figure 4.

based on the purpose of coding. In our view, the purpose of coding is dictated by the principles of phylogenetic systematics. The foundation of phylogenetic systematics is the observation that monophyletic groups can be recognized if homologous character states, shared evolutionary novelties, can be identified (Hennig, 1966). Unfortunately, characters do not have labels indicating their homology. Instead, a systematist must propose a hypothesis of homology and evaluate its congruence with independent hypotheses based on other traits. In this context, the purpose of coding is to represent those hypotheses.

The major obstacle to coding is that the a priori groups under analysis (i.e., taxa) often have ranges of variation that overlap to some degree. This is true whether traits are described qualitatively or quantitatively. One advantage of quantitative description is that it permits a more detailed analysis of how much the ranges of variation overlap. It may seem appropriate to use statistical methods to summarize the amount of overlap and even to decide objectively (on a priori grounds) whether taxa are similar or different. Above, we demonstrated some of the problems resulting from these uses of statistical analysis. In our view, the most important problem is the implication that similarity of the feature across taxa is the basis for inferring homology. The similarity that is relevant to phylogenetic analysis is not proximity in morphospace, but shared novelty. Statistical methods can describe proximity, but they cannot recognize novelty.

The method of coding we recommend uses graphical displays of individual values. Coding decisions are based on all of the individuals in each taxon, not on summaries derived from models of expected distributions. Then, the evidence for inferring divergence is independently evaluated for each pair of overlapping taxa. Coding decisions are not based on a priori rules that have no bearing on recognition of evolutionary novelty. This is the same approach that is used to code qualitatively described traits.

ACKNOWLEDGMENTS

We thank Sharon Jansa, Barb Lundrigan, and two anonymous reviewers for comments on previous drafts of this paper. Piranha studies were supported by NSF DEB-9509195 to (M.L.Z. and W.L.F.).

REFERENCES

ALMEIDA, M. T., AND F. A. BISBY. 1984. A simple method for establishing taxonomic characters from measurement data. Taxon 33:405–409.

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

BOOKSTEIN, F. L. 1989. Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Trans. Pattern Anal. Mach. Intell. 11:567–585.

BOOKSTEIN, F. L. 1991. Morphometric tools for landmark data: Geometry and biology. Cambridge Univ. Press, New York.

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

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

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

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

FINK, W. L., AND M. L. ZELDITCH. 1995. Phylogenetic analysis of ontogenetic shape transformations: A reassessment of the piranha genus *Pygocentrus* (Teleostei). Syst. Biol. 44:343–360.

GARLAND, T., JR., AND S. C. ADOLPH. 1994. Why not to do two species comparative studies: Limitations on inferring adaptation. Physiol. Zool. 67:797–828.

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.

HENNIG, W. 1966. Phylogenetic systematics. Univ. Illinois Press, Urbana.

MABEE, P. M., AND J. HUMPHRIES. 1993. Coding polymorphic data: Examples from allozymes and ontogeny. Syst. Biol. 42:166–181.

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

MICKEVICH, M. F., AND S. J. WELLER. 1990. Evolutionary character analysis: Tracing character change on a cladogram. Cladistics 6:137–170.

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

SIMON, C. 1983. A new coding procedure for morphometric data with an example from periodical cicada wing veins. Pages 378–382 *in* Numerical taxonomy. (J. Felsenstein, ed.). NATO ASI Series G, Volume 1. Springer-Verlag, Berlin.

SLICE, D. E., F. L. BOOKSTEIN, L. F. MARCUS, AND F. J. ROHLF. 1996. Appendix I. A glossary for geometric morphometrics. Pages 531–551 *in* Advances in morphometrics. (L. F. Marcus, M. Corti, A. Loy, G. J. P. Naylor, and D. E. Slice, eds.). NATO ASI Series A, Volume 284. Plenum, New York.

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

SWIDERSKI, D. L. 1993. Morphological evolution of the scapula in tree squirrels, chipmunks and ground squirrels (Sciuridae): An analysis using thin-plate splines. Evolution: 47:1854–1973.

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 38:244–355.

WILEY, E. O. 1981. Phylogenetics: The theory and practice of phylogenetic systematics. J. Wiley and Sons, New York.

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

ZELDITCH, M. L., F. L. BOOKSTEIN, AND B. L. LUNDRIGAN. 1992. Ontogeny of integrated skull growth in the cotton rat *Sigmodon fulviventer*. Evolution 46:1164–1180.

ZELDITCH, M. L., AND W. L. FINK. 1995. Allometry and developmental integration of body growth in a piranha, *Pygocentrus nattereri* (Teleostei: Ostariophysi). J. Morphol. 223:341–355.

ZELDITCH, M. L., W. L. FINK, AND D. L. SWIDERSKI. 1995. Morphometrics, homology and phylogenetics: Quantified characters as synapomorphies. Syst. Biol. 44:179–189.

Received: 10 April 1997; accepted 20 December 1997
Associate Editor: D. Cannatella

Syst. Biol. 47(3): 519–537, 1998

Systematic Bias in Phylogenetic Analysis: Is the Strepsiptera Problem Solved?

JOHN P. HUELSENBECK

Laboratory of Molecular Systematics, Smithsonian Museum Support Center, 4210 Silver Hill Road, Suitland, Maryland, 20746, USA; E-mail: johnh@onyx.si.edu

A phylogenetic method is inconsistent if it converges to an incorrect tree as characters (e.g., the columns or sites in a DNA sequence data matrix) are added to a phylogenetic problem. Inconsistency was first identified as a potential problem in phylogenetics by Felsenstein (1978) who showed that parsimony and compatibility methods could become inconsistent for four taxa under a restricted set of circumstances. However, the inconsistency problem was later shown to occur under less stringent conditions by Hendy and Penny (1989) who demonstrated the inconsistency of parsimony for trees of more than four species when the data obeyed a molecular clock. Other methods of phylogenetic estimation were later shown to be inconsistent under some conditions (DeBry, 1992; Huelsenbeck and Hillis, 1993; Gaut and Lewis, 1995; Huelsenbeck, 1995a; Waddell, 1995; Yang, 1996). For example, when the assumptions of distance and maximum-likelihood methods are