

THE USEFULNESS OF BEHAVIOR FOR PHYLOGENY ESTIMATION: LEVELS OF HOMOPLASY IN BEHAVIORAL AND MORPHOLOGICAL CHARACTERS

ALAN DE QUEIROZ¹ AND PETER H. WIMBERGER²

Section of Ecology and Systematics, Cornell University, Ithaca, NY 14853 USA

Abstract.—It is widely believed that behavior is more evolutionarily labile and/or more difficult to characterize than morphology, and thus that behavioral characters are not as useful as morphological characters for estimating phylogenetic relationships. To examine the relative utility of behavior and morphology for estimating phylogeny, we compared levels of homoplasy for morphological and behavioral characters that have been used in systematic studies. In an analysis of 22 data sets that contained both morphological and behavioral characters we found no significant difference between mean consistency indices (CIs, which measure homoplasy) within data sets for the two types of characters. In a second analysis we compared overall CIs for 8 data sets comprised entirely of behavioral characters with overall CIs for 32 morphological data sets and found no significant difference between the two types of data sets. For both analyses, 95% confidence limits on the difference between the two types of characters indicate that, even if given the benefit of the doubt, morphological characters could not have substantially higher mean CIs than behavioral characters. These results do not support the idea that behavioral characters are less useful than morphological characters for the estimation of phylogeny.

Key words.—Behavior, character evolution, consistency index, homoplasy, morphology, phylogeny.

Received July 12, 1991. Accepted May 14, 1992.

One of the basic premises of ethology is that behavior evolves in essentially the same fashion as morphology. This idea both arose from and promoted the use of behavioral characters in studies of phylogeny. Although Darwin (1859) clearly realized that instincts evolve, in modern ethology the idea that behavior and structure should be treated equivalently traces most clearly to Whittman (1899) and Heinroth (1911), whose views stemmed largely from the application of behavior to the systematics of birds. With the rise of ethology the use of behavioral characters in studies of phylogeny became relatively commonplace (Hinde and Tinbergen, 1958; Bekoff, 1977; McLennan et al., 1988). Tinbergen spoke for many ethologists when he claimed "Behaviour characters are in principle neither more nor less useful [for systematics] than morphological or other characters; they merely add characters to the total by which overall likeness is judged" (1959, p. 328).

Despite the ethological tradition of using behavior in systematics, among biologists in general it is widely believed that behavioral characters are inferior to morphological characters as indicators of phylogenetic relatedness. Two main arguments have been forwarded to support this notion. First, it is claimed that preliminary (i.e., nonphylogenetic) criteria for homologizing characters—for example, the position of the character in relation to other features—are difficult and perhaps even impossible to apply to behavioral traits (Atz, 1970; Hodos, 1976; Aronson, 1981). Second, it is held that behavior is so evolutionarily labile that it is suspect as an indicator of relationships (Atz, 1970; Urbani, 1989).

These two arguments are closely related, thus it is worthwhile to clarify the distinction between them from the outset. Both arguments imply that, following a phylogenetic analysis, behavioral characters will be found to exhibit more homoplasy—i.e., character convergence or reversal—than will morphological characters. The first argument implies that such homoplasy—or mistaken homology as some would call it—occurs because behavioral characters lack some of the properties that allow us to make reasonably good preliminary homology assess-

¹ Present address: Department of Ecology and Evolutionary Biology, Biological Sciences West Building, University of Arizona, Tucson, AZ 85721 USA.

² Present address: Department of Biology, University of Michigan, Ann Arbor, MI 48109 USA.

ments for morphological characters. Put another way, the behavioral characters that we would a priori call homologues are not as easily compared as morphological characters, thus we are more likely to make mistakes in assessing behavioral homology. The second argument, on the other hand, implies that, even if we could somehow equalize the a priori criteria for homology applied to behavior and morphology, we would still find that behavioral characters show more homoplasy than morphological characters. In other words, a given degree of similarity is more likely to arise or be lost multiple times for behavior than for morphology.

The fact that behavioral characters can give apparently good estimates of phylogeny falsifies the notion that such characters absolutely cannot be homologized (e.g., McLennan et al., 1988; Arntzen and Sparreboom, 1989; Prum, 1990). It is also clearly true that *some* behavioral characters can be quite conservative in their evolution. For example, chemosensory tongue protrusion apparently arose in a common ancestor of squamate reptiles no later than the Jurassic and has been retained in all lineages of squamates since then (Schwenk, 1988). Nonetheless, it is possible that a priori criteria of homology are *usually* more difficult to apply to behavior than morphology and/or that behavioral characters are *usually* more intrinsically labile than morphological characters.

Studies of sticklebacks (McLennan et al., 1988) and manakins (Prum, 1990) have shown that behavioral homoplasy is no more prevalent than morphological homoplasy for characters used to estimate phylogeny in these taxa. However, a more inclusive test is required to assess in more general terms the relative usefulness of these two kinds of characters for the estimation of phylogeny.

In this study we use cladistic methods to compare relative amounts of behavioral and morphological homoplasy in a wide variety of taxa and characters. Our study complements an earlier study by Sanderson and Donoghue (1989) comparing levels of homoplasy in morphological and molecular data sets. We discuss the results with respect to both the usefulness of behavior in systematics and the question of whether be-

havior in general is more evolutionarily labile than morphology.

MATERIALS AND METHODS

General

We performed two comparisons of levels of homoplasy in behavioral and morphological characters. In the first, we used data sets that contained both behavioral and morphological characters and compared levels of homoplasy for the two kinds of characters within each data set. In the second, we compared overall amounts of homoplasy for data sets that contained only behavioral characters with those for data sets that contained only morphological characters. The first analysis controls for peculiarities of particular taxa by pairing behavioral and morphological data points by study group. The second analysis lacks this control but provides a larger sample of taxa and a more powerful statistical test.

The data sets that include behavioral characters are all the published and unpublished systematic data sets we could find that allowed calculations of homoplasy as detailed below. The purely morphological data sets in the second analysis were taken from Sanderson and Donoghue (1989). This sampling scheme resulted in a bias toward studies of certain taxonomic groups and taxonomic ranks. All of the data sets containing behavioral characters are for either arthropods or vertebrates, and within these two groups hymenopterans and birds are particularly overrepresented. Most of the data sets are from studies of interspecific or intergeneric relationships.

We define a behavioral character broadly as one representing movement (or the lack of movement—e.g., “freezing” in response to a predator) of the organism as a whole or any of its external parts. The behavioral data include characters representing a wide assortment of functional categories, including courtship, nesting, territorial and other social behavior, antipredator responses, and feeding behavior. A wide range of organizational levels of behavior are also represented, from simple, stereotyped movements to more inclusive behaviors such as nesting dispersion (e.g., solitary or colonial) and habitat type. Physical manifestations of

behavior, principally characters of nest architecture, are also included. Notwithstanding this wide range of behavioral character types, mating display and nest architecture characters make up the majority of the behavioral data.

The morphological characters are also varied and include color patterns and other gross external characteristics, many osteological features, and characteristics of the soft anatomy. We excluded karyological and biochemical characters from the analyses of levels of homoplasy. These latter characters might be considered morphological; however, the ideas regarding behavioral and morphological characters that we are trying to address were not formulated with these kinds of characters in mind. This is not to suggest that we believe such characters are more or less useful than classical morphological features. Sanderson and Donoghue (1989; see also Wyss et al., 1987) found no difference in levels of homoplasy between classical morphological characters and molecular characters.

When entire data sets were taken from a single systematic study we used the tree(s) obtained by the original author(s). If two or more data sets from the literature were combined, or if only part of an original data set was amenable to cladistic analysis, we obtained trees using the computer programs PAUP version 3.0 (Swofford, 1989) or Hennig 86 (Farris, 1988). In these cases we retained the ordering scheme of the original authors; if no obvious ordering scheme had been used we treated the characters as unordered. A data set was included only if the number of phylogenetically informative characters was large enough to at least potentially give a fully resolved tree; thus the number of informative binary-character equivalents had to be at least equal to the number of taxa minus two. (The number of binary-character equivalents is the number of characters that a data set would contain if all the characters were converted to binary characters.)

We used the consistency index (CI; Kluge and Farris, 1969) as our measure of homoplasy. The CI is defined as the minimum number of character state changes required by a given data set divided by the number of character state changes required for the

same data given the tree in question. This can be restated as the number of state changes required (given the number of characters and character states but without reference to a tree) if no character exhibits any homoplasy divided by the number of state changes, including homoplasious changes, required by the tree in question. CIs can be calculated both for individual characters and for all characters combined (Farris, 1989b). A CI of 1.0 indicates no homoplasy in the character(s) in question for the given tree. As homoplasy increases the CI decreases asymptotically toward a theoretical minimum value that is determined by the number of characters, the number of character states, and the number of taxa (Archie, 1989). As an illustration of the calculation of the CI, consider a character that, within the group of taxa under consideration, has two character states, a primitive state A and a derived state A'. The numerator of the CI for this character is always 1 because the minimum number of character state changes is 1 (i.e., a single change from A to A'). If, for the given tree, A' is required to arise twice independently from A, then the denominator is 2 and the CI is 0.5. If, in addition to arising twice, A' in one lineage must revert back to A, then the denominator is 3 and the CI is 0.333. Insofar as homoplasy is a form of "noise" in phylogenetic analysis, higher CIs indicate greater usefulness for the estimation of phylogeny.

The consistency index has been criticized as a comparative measure of homoplasy on the grounds that it varies with number of characters and number of taxa and is sensitive to the inclusion of unique derived (autapomorphic) characters, which by definition are phylogenetically uninformative (Archie, 1989). We have used the CI because it is easily calculated, its meaning is generally understood by systematists, and its inherent biases can be accounted for in comparative analyses (see below). Furthermore, the alternative measures of homoplasy whose properties are well known—the homoplasy excess ratio, homoplasy excess ratio maximum, and character retention index (Archie, 1989, 1990; Farris, 1989a)—all take into account not only the number of states for each character but the number of taxa that have each state. Part of our aim

TABLE 1. Systematic data sets used for comparison of character CIs within data sets. See text for description. RANK is the taxonomic rank of the terminal taxa among which relationships were investigated. MCHAR and BCHAR are the number of morphological and behavioral characters, respectively. MCI and BCI are the mean CIs for morphological and behavioral characters, respectively. The references include those from which the morphological and behavioral characters being analyzed were taken as well as any others that contributed information used to construct trees.

Set	Taxon	RANK	#TAXA	MCHAR	BCHAR	MCI	BCI	REF.
1	Polistine wasps	genera	28	34	18	0.60	0.62	1
2	Eumenine wasps	genera	25	25	1	0.86	1.00	2
3	Vespine wasps	spp/gen	7	17	8	0.88	0.71	3, 4
4	Vespid wasps	subfam	6	20	8	0.78	0.96	5-7
5	<i>Apis</i> bees	species	6	9	3	0.94	0.83	8
6	Meloid beetles	genera	9	21	2	0.87	1.00	9
7	Arachnids	orders	11	63	1	0.71	1.00	10
8	Shrimps	orders	5	3	1	0.83	1.00	11
9	Gasterosteiform fishes	genera	6	18	17	0.86	0.96	12, 13
10	No. Amer. Hylid frogs	species	13	9	4	0.94	0.66	14-18
11	Pipid frogs	spp/gen	7	52	2	0.88	0.75	19, 20
12	<i>Triturus</i> newts	species	9	2	11	0.52	0.81	21-26
13	Seaducks	spp/gen	10	12	2	0.83	0.75	27-29
14	Alcid birds	species	23	31	2	0.82	0.70	30
15	Pelecaniform birds	families	13	26	5	0.92	1.00	31
16	Ochthoeca flycatchers	genera	5	5	2	0.90	1.00	32
17	Tody-tyrant flycatchers	genera	9	8	2	1.00	0.75	33
18	<i>Myiobius</i> flycatchers	genera	5	8	2	0.85	1.00	34
19	Empidonax flycatchers	genera	7	6	3	1.00	0.83	32
20	Manakin birds	species	19	36	29	0.90	0.84	35, 36
21	Squamate reptiles	families	19	61	3	0.68	0.67	37-39
22	Sand lizards	species	10	36	2	0.87	0.67	40

References: 1. Carpenter and Wenzel, unpubl. data; 2. Carpenter and Cumming, 1985; 3. Carpenter, 1987; 4. Carpenter, 1989a; 5. Carpenter, 1982; 6. Carpenter, 1988; 7. Carpenter, 1989b; 8. Alexander, 1991; 9. Pinto, 1984; 10. Schultz, 1990; 11. Schram, 1986; 12. McLennan et al., 1988; 13. Nelson, 1971; 14. Gaudin, 1969; 15. Gaudin, 1974; 16. Wiley, 1982; 17. Hedges, 1986; 18. Cocroft, unpubl. data; 19. Cannatella and Truett, 1988; 20. de Sa and Hillis, 1990; 21. Hellmich, 1962; 22. Nikol'skii, 1962; 23. Thorn, 1968; 24. Arnold and Burton, 1978; 25. Obsi et al., 1988; 26. Arntzen and Sparreboom, 1989; 27. Johnsgard, 1960; 28. Livezey, unpubl. data; 29. Livezey, 1986; 30. Strauch, 1985; 31. Cracraft, 1985; 32. Lanyon, 1986; 33. Lanyon, 1988a; 34. Lanyon, 1988a; 35. Prum, 1990; 36. Prum, unpubl. data; 37. Estes et al., 1988; 38. Presch, 1988; 39. Schwenk, 1988; 40. de Queiroz, 1989.

was to measure the evolutionary lability of characters and this lability can be associated with the number of taxa that have each character state. For example, the number of taxa in which a derived character state occurs is also the maximum number of independent origins of that state that can be inferred from the data. Thus we did not want to factor out the number of taxa that have each state in our measure of homoplasy. The CI does not account for this variable and thus is more appropriate than the alternative measures as an indicator of evolutionary lability.

For calculations of CIs we used only data coded as discrete states; CIs can be calculated for continuous characters (Kluge and Farris, 1969), but it is unclear how one should compare such CIs with those obtained for discretely coded characters. Phylogenetically uninformative characters—i.e., autapomorphies and characters invariant within the ingroup—were excluded for the

calculations of CIs. Inclusion of such characters in calculations of CIs gives an inflated indication of the phylogenetic usefulness of the data. In addition to assessing the usefulness of characters in phylogenetic studies a goal of our study was to measure the evolutionary lability of characters. Autapomorphies and shared derived characters (synapomorphies) common to all members of the ingroup do indicate evolutionary change. However, because some of our sources presented only informative characters, we did not perform a separate analysis with uninformative characters included.

Statistical analyses were performed using Systat version 5.1 (Wilkinson, 1987).

Comparison of Character CIs within Data Sets

For each of the 22 data sets listed in Table 1 we obtained a most parsimonious tree or set of trees (the most trees for any one study

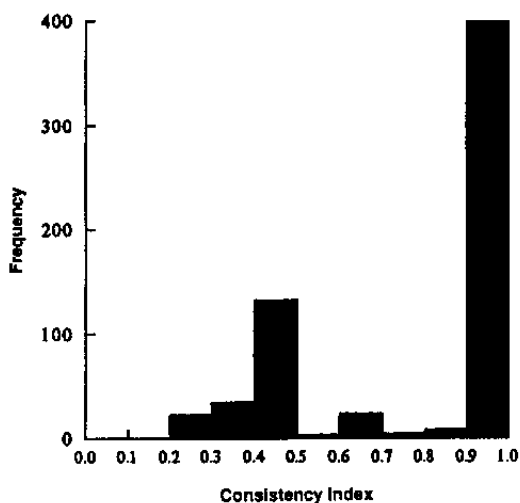


FIG. 1. Combined distribution of individual CIs of informative characters for all the data sets listed in Table 1.

being six) as outlined above. Because our goal was to measure actual levels of homoplasy we excluded data sets (not referenced) that gave highly unresolved trees. We did not use a specific algorithm for rejecting data sets on these grounds. However, decisions to reject data sets were made without knowledge of CIs for these data, thus their exclusion should not have biased the results. To obtain the best estimate of the true tree, we included any available karyological and molecular data that could be coded as characters with discrete states. (As noted above, levels of homoplasy were not calculated for such characters.) In estimating the phylogeny we included all types of data in a single parsimony analysis rather than performing separate analyses and obtaining a consensus tree (for the rationale for this approach see Barrett et al., 1991).

We tested the null hypothesis that behavioral characters are no more homoplasious than morphological characters by comparing mean CIs for the two kinds of characters within each of the data sets. In compiling the sets of character CIs we excluded characters that were polymorphic within any terminal taxon on the grounds that CIs for such characters may be heavily influenced by the unknown cladistic structure within these terminal taxa. In addition, we excluded behavioral characters that obviously depend on unique morphological features. We

also left out characters that had unknown states for more than 20% of the taxa. This 20% rule allowed us to retain the great majority of characters without including what seemed an unreasonable number of unknown states, but this threshold is obviously somewhat arbitrary. Including characters with unknown states for certain taxa tends to raise CIs because parsimony algorithms assign the most parsimonious state to such unknowns. A greater percentage of the behavioral than the morphological characters had unknown states, thus the 20% threshold (rather than exclusion of all characters with any unknown states) may have biased the results toward higher CIs for behavior. However, the great majority of behavioral characters were known for all taxa so this effect is likely to be slight. For cases in which more than one most parsimonious tree was found, we used the mean CI averaged over all the characters in all the trees.

For phylogenetically informative data sets in general the distribution of character CIs tends to be heavily skewed toward 1.0 (perfect consistency). In addition, because of the limited set of numerators and denominators only certain values for character CIs are possible and in reality only very few values are at all frequent. A typical distribution of character CIs has a large peak at 1.0 with several smaller peaks at other likely outcomes (e.g., 0.5 for a binary character with one extra step; 0.333 for a binary character with two extra steps; see Fig. 1). This distribution is obviously highly nonnormal. The distributions of the *mean* character CIs for the behavioral and morphological samples are also nonnormal, although they are more normal than the CI distribution. However, the distribution of differences between mean morphological and behavioral CIs within data sets does not differ significantly from a normal distribution (Lilliefors test, $N = 22$, maximum distance = 0.126, $P > 0.4$). This justified the use of a parametric test; we used a paired *t*-test to assess differences between morphological and behavioral CIs, with each data set providing a paired comparison in the analysis.

Because of the small sample size (i.e., number of data sets) and the substantial variance in the difference between mean morphological and behavioral CIs, a type II