Patterns of Variation in Levels of Homoplasy

Michael J. Sanderson; Michael J. Donoghue

Evolution, Vol. 43, No. 8 (Dec., 1989), 1781-1795.

Stable URL:
http://links.jstor.org/sici?sici=0014-3820%28198912%2943%3A8%3C1781%3APOVILO%3E2.0.CO%3B2-1

Evolution is currently published by Society for the Study of Evolution.

Evolution, 43(8), 1989, pp. 1781–1795

PATTERNS OF VARIATION IN LEVELS OF HOMOPLASY

MICHAEL J. SANDERSON[1] AND MICHAEL J. DONOGHUE
Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721

Abstract.—Patterns of variation in levels of homoplasy were explored through statistical analyses of standardized consistency indexes. Data were obtained from 60 recent cladistic analyses of a wide variety of organisms based on several different kinds of characters. Consistency index is highly correlated with the number of taxa included in an analysis, with homoplasy increasing as the number of taxa increases. This observation is compatible with a simple model of character evolution in which 1) the probability of character-state change increases with the total number of branches in a tree and 2) the number of possible states of a character is limited. Consistency index does not show a significant relationship to the number of characters utilized in an analysis or to the taxonomic rank of the terminal taxa. When the relationship between consistency index and number of taxa is taken into account, there is no significant difference between plant and animal data sets in the amount of homoplasy. Likewise, the level of homoplasy in morphological and molecular data sets does not appear to differ significantly, although there are still too few molecular studies to be confident of this result. Future comparisons of consistency indexes, including studies along the lines established here, must take into account the influence of the number of taxa on homoplasy.

Received October 4, 1988. Accepted July 20, 1989

Phylogenetic studies of character evolution have focused on selected traits within individual groups of organisms (e.g., Ridley, 1983; Wake and Larson, 1987; Huey and Bennett, 1987; Sessions and Larson, 1987; Sillén-Tullberg, 1988; Hufford, 1988). Lauder (1981, 1982) suggested searching for general patterns of character evolution by comparing cladograms of different groups, but this has seldom been done, and such comparisons have been restricted to only a few groups (e.g., Emerson, 1988). No attempt has been made to document general patterns of character evolution by comparing cladograms of many groups at once. Here we undertake such an analysis of one general feature of character evolution, homoplasy (Lankester, 1870), which is the independent evolution of the same character state.

Although the term homoplasy is sometimes used to refer to any kind of convergence, parallelism, or reversal (Futuyma, 1986), its connotation is narrower in cladistic analyses, which provide the basic data upon which our analysis is based. Homoplasy in cladistic analyses results when features hypothesized at the outset of an analysis to be homologous are found to arise more than once on a cladogram, or to originate and then be lost (Farris, 1983; Patterson, 1982). The criteria used to assess homology a priori are a special subset of all possible criteria, since they are intended to permit recognition of character states that are similar by virtue of common ancestry. In particular, these include the positional and developmental criteria of Remane (1952; see Patterson, 1982).

Systematists often view homoplasy resulting from rejection of these initial hypotheses of homology as no more than a collection of mistakes. Indeed, the most frequently used measure of homoplasy in cladistic studies, the "consistency index" (Kluge and Farris, 1969; see below) is generally interpreted as a measure of goodness of fit of data to a tree topology, rather than a description of some evolutionary pattern. In some cases, the quality of a particular study has been judged largely by this measure (e.g., Riggins and Farris, 1983 p. 99).

Although we agree that homoplasy always represents a "mistake" in the sense of a mistaken hypothesis of homology, it is clear that not all such mistakes are of equal interest to evolutionary biologists. Mistakes due to measurement errors (e.g., misreading a caliper) do not necessitate any explanations of an evolutionary kind, whereas "mistakes" such as the discovery of an independent origin of some character do. Moreover, if the latter mistake occurs among characters that have already passed some a priori test of homology, they are especially interesting, because the evolutionary expla-

[1] Present address: 467 Mann Library, L. H. Bailey Hortorium, Cornell University, Ithaca, NY 14853.

TABLE 1. Cladistic analyses used in this study. Data types: "morph" refers to morphological data (broadly construed); "molec" refers to molecular data. See text and the Appendix for details. The numbers of characters were standardized to the numbers of binary-character equivalents (Sokal and Shao, 1985).

Data set	Kingdom	Data type	Taxonomic rank	Number of taxa	Number of characters	Number of autapomorphies	Corrected consistency index	Source
1	plant	morph	species	6	17	6	0.79	Jansen, 1981
2	plant	morph	species	9	19	0	0.79	Anderberg, 1986
3	plant	morph	species	14	31	3	0.72	Ladiges and Humphries, 1983
4	plant	morph	species	20	28	4	0.44	Campbell, 1986
5	plant	morph	species	25	51	12	0.63	Funk, 1982
6	plant	morph	species	29	65	3	0.43	Ladiges et al., 1987
7	plant	morph	species	37	65	3	0.39	Eckenwalder and Barrett, 1986
8	plant	morph	genus	29	126	0	0.31	Baum, 1983
9	plant	morph	genus	29	81	34	0.56	Bremer, 1987
10	plant	morph	genus	65	39	0	0.37	Kellogg and Campbell, 1987
11	plant	morph	genus	68	94	0	0.32	P. Stevens, pers. comm.
12	plant	morph	family	15	56	12	0.60	Dahlgren and Rasmussen, 1983
13	plant	morph	family	20	71	2	0.35	Rodman et al., 1984
14	plant	morph	family	47	61	2	0.26	Dahlgren and Bremer, 1985
15	plant	morph	order	15	37	6	0.55	Gabrielson and Garbary, 1987
16	plant	morph	class	11	41	13	0.82	Mishler and Churchill, 1985
17	plant	morph	class	20	31	0	0.62	Crane, 1985
18	plant	morph	class	20	62	0	0.50	Doyle and Donoghue, 1986
19	plant	molec[a]	species	9	119	64	0.90	Sytsma and Gottlieb, 1986*a*
20	plant	molec[b]	genus	10	117	0	0.73	Hamby and Zimmer, 1988
21	plant	molec[a]	genus	16	211	156	0.60	Jansen and Palmer, 1988
22	plant	molec[a]	genus	12	390	273	0.65	Jansen and Palmer, 1988
23	plant	molec[a]	genus	55	868	568	0.52	R. K. Jansen and J. Palmer, pers. comm.
24	plant	molec[b]	family	6	111	35	0.60	Bremer, 1988
25	plant	molec[b]	family	9	88	26	0.50	Bremer, 1988
26	plant	molec[b]	order	26	84	0	0.53	Bremer et al., 1987
27	animal	morph	species	4	23	10	1.00	Livezey, 1986
28	animal	morph	species	6	34	13	0.84	Grismer, 1983
29	animal	morph	species	9	33	3	0.81	Cracraft, 1986*a*
30	animal	morph	species	11	83	41	0.64	Jamieson et al., 1987
31	animal	morph	species	12	32	0	0.70	Fink and Fink, 1986
32	animal	morph	species	20	74	1	0.68	Collette and Russo, 1985
33	animal	morph	species	24	31	1	0.60	Guyer and Savage, 1986
34	animal	morph	species	27	94	5	0.40	Cane, 1983
35	animal	morph	species	35	72	7	0.43	Herman, 1986
36	animal	morph	genus	5	36	13	0.72	Crother et al., 1986
37	animal	morph	genus	7	26	0	0.76	Carpenter, 1987
38	animal	morph	genus	10	49	12	0.80	Schuh and Polhemus, 1980
39	animal	morph	genus	16	15	0	0.50	Wighton and Wilson, 1987
40	animal	morph	genus	17	18	5	0.54	Cutler and Gibbs, 1985
41	animal	morph	genus	27	323	78	0.49	Fink, 1985
42	animal	morph	genus	57	307	24	0.45	Kitching, 1987
43	animal	morph	family	13	59	0	0.66	Cracraft, 1985
44	animal	morph	family	22	113	43	0.63	Nelson, 1984
45	animal	morph	class	8	84	0	0.89	Gauthier, 1986
46	animal	morph	class	9	39	11	0.93	Brooks et al., 1985
47	animal	molec[a]	species	6	118	87	0.56	George and Ryder, 1986
48	animal	molec[a]	species	32	76	55	0.66	Hillis and Davis, 1986
49	animal	molec[b]	order	13	19	0	0.59	Wyss et al., 1987
50	animal	molec[b]	order	14	18	0	0.49	Wyss et al., 1987
51	animal	molec[b]	order	14	27	0	0.60	Wyss et al., 1987
52	animal	molec[b]	order	19	17	0	0.63	Wyss et al., 1987
53	animal	molec[c]	species	12	21	2	0.95	Miyamoto, 1983
54	animal	molec[c]	species	25	90	38	0.38	Sites et al., 1984
55	animal	molec[c]	species	26	68	0	0.37	W. Heed, pers. comm.
56	animal	molec[c]	genus	14	92	46	0.47	Straney, 1981

TABLE 1. Continued.

Data set	Kingdom	Data type	Taxonomic rank	Number of taxa	Number of characters	Number of autapomorphies	Corrected consistency index	Source
57	other	morph	species	11	25	5	0.53	Vilgalys, 1986
58	other	morph	genus	17	24	0	0.53	Baum and Saville, 1985
59	other	morph	?	18	20	8	0.63	Churchill et al., 1984
60	other	morph	class	36	105	38	0.34	Lipscomb, 1985

[a] Restriction-fragment analysis.
[b] Nucleotide-sequence and amino-acid-sequence analyses.
[c] Protein-polymorphism electrophoretic analysis.

nation for their recurrence may differ from an explanation that is adequate to account for broad convergences that are not positionally or developmentally similar (Rensch, 1959 p. 191).

Assertions about homoplasy are widespread in the literature (see Discussion). Statements regarding levels of homoplasy in plants versus animals and in morphological versus molecular data have seldom been tested. Hypothesized relationships between homoplasy and the number of characters, taxonomic rank, or number of taxa included in cladistic studies have received even less attention. Here, we explore general patterns of variation in levels of homoplasy utilizing data obtained from recent cladistic analyses of a wide variety of organisms. We focus on the relationship between consistency index and the variables mentioned above and then consider the evolutionary implications of the patterns observed.

MATERIALS AND METHODS

Cladistic studies were selected according to the following criteria. First, in order to exclude analyses utilizing earlier versions of parsimony algorithms, we considered only studies conducted since 1980. Second, only those analyses based on discrete "qualitative" characters were selected; studies involving continuous (morphometric) characters and distance data (e.g., DNA-hybridization data) were not considered, because there is no clear analog of consistency index for such data. Third, only data sets analyzed under Wagner parsimony (allowing both forward state changes and reversals; Felsenstein, 1982) were selected for comparison; analyses using restricted parsimony approaches (e.g., "Dollo parsimony"), character compatibility, or phenetic algorithms were excluded. Finally, analyses were included only if sufficient information was provided to permit us to standardize consistency indexes in the manner described in the Appendix.

Sixty data sets that satisfied these requirements were selected (Table 1), encompassing a broad range of variation in the type of organism and the type of characters involved. Of these, 26 are analyses of plant taxa, 30 are analyses of animals (both "plant" and "animal" are interpreted broadly), and four are analyses of "other" groups (two of fungi, one of all major groups of eukaryotes, and one involving invented "organisms," the Dendrogramaceae). Forty-two of the studies are analyses of "morphological" data (interpreted broadly to include one study of secondary chemical characters), while the remaining 18 are analyses of "molecular" data (also interpreted broadly to include four protein-electrophoretic studies and 14 restriction-fragment, amino-acid, and nucleotide-sequence studies). The smaller number of molecular studies reflects the fact that fewer molecular studies that meet the requirements specified above are currently available. The selected studies involve a wide range of numbers of taxa (4–68), numbers of characters (15–868), and taxonomic ranks of terminal taxa (species–classes/phyla).

The consistency index (CI; Kluge and Farris, 1969) was used as the standard for comparison of levels of homoplasy among studies. Although CI measures only the overall amount of homoplasy, its simplicity and widespread use make it an obvious choice for comparing data sets. A character is perfectly consistent (shows no homoplasy on a cladogram) if all state changes occur only once. For example, a perfectly consis-

tent binary (two-state) character would entail only a single state change (step), say from state "0" to state "1". If more than one state change is required to fit parsimoniously, or to "optimize," the character on the tree (for example, two changes from "0" to "1" or a change from "0" to "1" and then a reversal to "0"), then the character is not perfectly consistent; it is inconsistent, or homoplastic. The consistency index is defined as the minimum number of character-state changes required by a particular data set (summed over all characters) divided by the total number of state changes required to most parsimoniously fit all of the characters on the tree under consideration. If homoplasy is absent the consistency index is 1.0, and CI decreases toward 0 as homoplasy increases.

Possible Sources of Bias

Despite its frequent use, it is not widely appreciated that CI is affected by the inclusion of characters that are necessarily perfectly consistent on all trees (Carpenter, 1988). Inclusion of characters in which a single state is possessed by all groups under consideration (invariant characters) and/or characters in which only one of the included taxa possesses a particular derived state (autapomorphies), will inflate CI by adding one unit to both the numerator and denominator in the calculation. Unfortunately, cladists have not been entirely consistent in omitting these phylogenetically uninformative characters from their data sets. Recognizing the possibility that this inconsistency might seriously bias our comparison of CI's obtained from different studies, we recalculated CI's after eliminating necessarily consistent characters according to a set of rules detailed in the Appendix.

Several other potential biases are more difficult to eliminate. We are unaware of any systematic bias associated with these factors in relation to type of organism, type of data, or any other variable discussed below and, therefore, doubt that they exert any significant effect on our results. Nevertheless, these factors are worthy of mention as possible sources of residual variation in CI's, and their effects should be tested when a larger number of appropriate cladistic studies become available.

Character Distribution.—The effects of autapomorphies and invariant characters on CI are best seen as extreme cases of the influence of the distribution of character states among the taxa. Characters in which an almost equal number of taxa possess the alternate states of a binary character have a greater probability of showing lower consistencies than characters in which the distribution of states is markedly unequal, simply because there are more opportunities for homoplasy in the first case (Meacham, 1981, 1984). Thus, data sets that include a large proportion of characters with highly unequal distributions of states will tend to have higher CI's than those in which most characters have nearly equal distributions among the taxa.

Multistate Characters.—Studies involving ordered multistate characters will tend to have lower consistencies than those in which the multistate characters are unordered, because cladograms may necessitate state transitions in ordered characters that add more than one step (e.g., from state 1 to 3). This effect is apparent in Baum's (1984) compatibility analysis of *Avena,* which involved ordered characters with as many as 16 states. Not surprisingly, he found that the largest sets of mutually compatible characters (cliques) were exceedingly small, indicating a very high level of inconsistency. Most molecular studies utilize either binary coding for presence/absence of restriction fragments or unordered multistate coding for nucleotide data. Such data are therefore less susceptible to this kind of bias than are ordered morphological data.

Missing Data.—Parsimony algorithms assign characters scored as "unknown" or "missing" to whichever state is most parsimonious, given the position of the taxon in the tree based on "known" characters. Since such coding can never increase inconsistency, whereas homoplasy might be introduced if "unknowns" were replaced by definite scores, data sets with more unknowns will tend to have higher consistencies than those with fewer unknowns. This is likely to be a significant factor only for data sets in which many taxa are scored as unknown due to high levels of polymorphism or in studies that include fossil groups for which information on many characters

TABLE 2. Total correlation matrix.

Variable	Variable		
	Number of taxa	Number of characters	Taxonomic rank
log(CI)	−0.68***	−0.12	−0.01
Number of taxa		0.38**	−0.04
Number of characters			−0.05

** $P < 0.01$; *** $P < 0.001$.

is lacking (e.g., Gauthier, 1986; Cracraft, 1986*b*) or about which there are uncertainties regarding initial homology assessments (e.g., Doyle and Donoghue, 1986).

Statistical Analyses

We utilized multiple linear regression to test for the dependence of CI on several variates of interest, including the number of taxa included in an analysis, the number of characters used, and the taxonomic rank of the terminal taxa (using a linear scale corresponding to the categories of the Linnaean hierarchy: from species = 1 to class or above = 5). Because preliminary analyses suggested a curvilinear relationship of CI to one of the independent variates, number of taxa (see Fig. 1A), and because the CI is a ratio, which becomes increasingly insensitive to homoplasy as homoplasy increases, log-transformed CI's were used in subsequent analyses (Fig. 1B). Log transformation was only partially effective in linearizing the data.

Differences in regression parameters between subsets of the data were examined using analysis of covariance. Class variables considered were type of organism ("plant" versus "animal") and type of character data ("morphological" versus "molecular"). All statistical analyses were performed using SAS version 6.02 on an IBM PC-AT microcomputer.

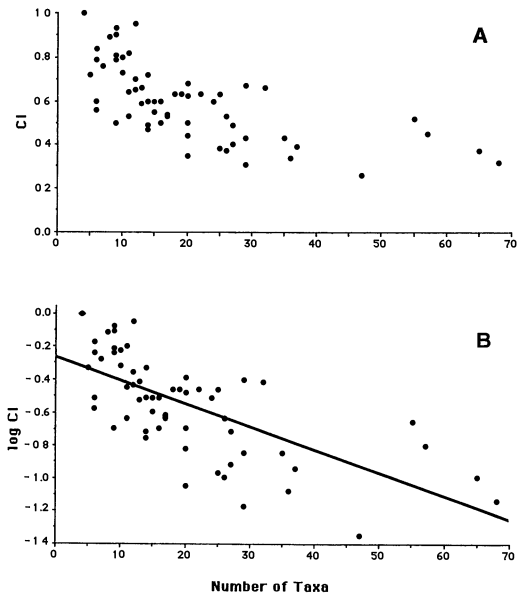

FIG. 1. Graphs of consistency index (CI) versus number of taxa included in 60 cladistic analyses (Table 1): A) untransformed consistency indexes versus number of taxa; B) log-transformed consistency indexes and linear regression of CI on number of taxa (see Table 3). The formula CI = 0.90 − 0.022 (number of taxa) + 0.000213 (number of taxa)^2, derived from polynomial regression analysis of Table 1, can be used to estimate expected CI's for a study with a given number of taxa. The regression cannot be extended over about 60 taxa.

RESULTS

Relationships among the variables are shown in Table 2 and Figures 1–3. Consistency index is highly correlated with number of taxa ($P < 0.001$) but not with number of characters or taxonomic rank. Number of characters is strongly correlated with number of taxa ($P < 0.01$) but not with CI, despite the strong correlation between CI and number of taxa. Multiple-regression analysis applied to the entire data set confirmed the pattern evident from the corre-

TABLE 3. Regression-parameter estimates of log-transformed consistency indexes for subsets of the data.

Subset	Sample size	Intercept	Regression slope		
			Number of taxa	Number of characters	Taxonomic rank
All data	60	−0.2528**	−0.0158***	0.0003	−0.0092
Animals	30	−0.2507*	−0.0160**	0.0001	0.0151
Plants	26	−0.2740*	−0.0143***	0.0004	−0.0236
Morphological data	42	−0.2126*	−0.0170***	0.0002	−0.0029
Molecular data	18	−0.3552*	−0.0104	0.0003	−0.0191

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

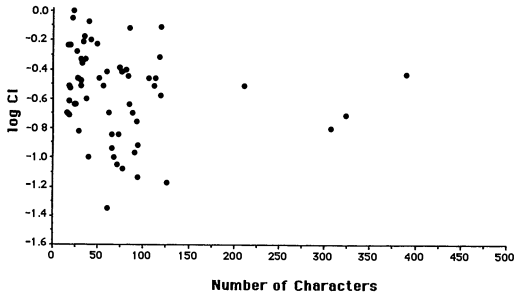

FIG. 2. Log-transformed consistency index versus number of characters in 60 cladistic analyses (Table 1). Studies with more than 400 characters have been omitted from the graph (see text).

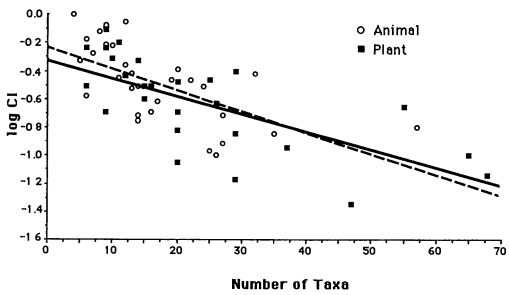

FIG. 4. Comparison of log-transformed consistency index versus number of taxa in plant versus animal data sets (Table 1). Simple regressions of CI on number of taxa are plotted for each group. The solid line is the regression for plants, and the dashed line is the regression for animals.

lation matrix (Table 3). The intercept and the slope of number of taxa are both highly significant components of the regression ($P < 0.001$), while other variates are not statistically significant. The linear-regression model explains 50% of the variance in CI. Because of potential sensitivity of the regression model to extreme outliers for character number (those few studies with between 200 and 800 characters), an additional analysis was conducted omitting these studies. A negative, but nonsignificant ($R^2 = 9\%$), relationship was found between CI and character number in a simple regression analysis; however, when other variables were included in a multiple-regression model, number of characters was still not a significant component.

Conclusions based on the entire data set are the most robust results of our study. When the data set is subdivided, sampling error becomes a more serious problem, and we hesitate to draw strong conclusions. For this reason, we did not further subdivide the "molecular" subset of studies. We suspected that number of taxa would be the most important factor affecting CI in subsets of the data (as it was in the entire data set), and multiple regression applied separately to plants and animals, and to morphological and molecular data, provided no reason to question this expectation. Regressions for plants and animals are strikingly similar, as are estimates for morphological and molecular data; in each case, only number of taxa is a significant variate. Although the regression parameters for the molecular data set are not significantly different from zero, they nevertheless suggest a negative relationship between CI and number of taxa.

Plots of CI versus number of taxa for subsets of the data are shown in Figures 4 and 5. Simple regressions of a dependent variate on a single independent variate may be misleading in the context of multiple regression

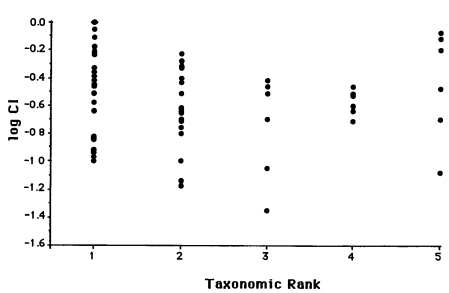

FIG. 3. Log-transformed consistency index versus taxonomic rank (Table 1), in which ranks are assigned an integer value corresponding to the Linnaean hierarchy, from species = 1 to classes (and above) = 5.

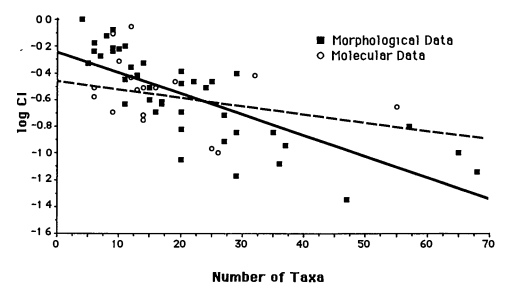

FIG. 5. Comparison of log-transformed consistency index versus number of taxa in morphological versus molecular data sets (Table 1). Simple regressions of CI on number of taxa are plotted for each group. The solid line is the regression for morphological data, and the dashed line is the regression for molecular data.

if other variates exert important effects (Montgomery and Peck, 1982). However, the similarity between plant and animal data sets in the multiple-regression analyses (Table 3) is mimicked by a similarity in the simple regressions (Fig. 4). Analysis of covariance for number of taxa alone showed no significant difference in the slopes of these two regressions. The comparison of morphological and molecular characters was less clear. Multiple-regression parameters suggest that there are some differences in the patterns for the two kinds of data (Table 3), but the simple regressions of CI on number of taxa differ from one another only at a significance level of $P = 0.057$ (Fig. 5). More molecular studies must be considered before any importance can be attached to this "almost significant" difference.

Discussion

Number of Taxa

Our results indicate that as the number of taxa included in an analysis increases, the consistency index decreases (Fig. 1). Therefore, the average number of state changes per character increases with the addition of taxa. Although this relationship has been suggested previously both on theoretical (e.g., Riggins and Farris, 1983) and empirical grounds (Archie's [1985] analysis of seven morphometric data sets), the present analysis, based on 60 studies, provides definitive support for this relationship.

Several taxonomic artifacts might yield the observed results, but in our opinion, none provides a compelling explanation. For example, systematists working on large groups might be more prone to errors in a priori homology assessments and in scoring taxa (perhaps because of the additional work required), leading to an increase in homoplasy. There is, however, sociological pressure opposing this bias, since low CI values are widely considered to be symptomatic of poor character analysis and dubious results. Furthermore, systematists sometimes remove excessively homoplastic characters after consulting a preliminary parsimony analysis (at least in morphological studies), either as part of a formal successive weighting scheme (Farris, 1969) or because initial indications of homoplasy lead to a reconsideration of certain characters and the discovery of mistakes. Even so, systematists have been unable to maintain high CI's in large studies.

Bias would also be introduced if systematists studying large groups included more homoplastic characters than they would in studying smaller groups. A study of a large number of taxa necessarily requires the identification of a larger number of individual relationships, and this requires more characters to achieve the same degree of resolution. The strong positive correlation seen between number of taxa and number of characters (Table 2) may reflect a conscious or unconscious desire to include more characters, even if this means settling for characters that are more likely to show homoplasy. However, there is no reason why additional characters must necessarily be more homoplastic; in fact, larger numbers of characters do not appear to be correlated with lower CI's (see below).

A third possible bias may result from the algorithms used to generate most-parsimonious trees. For small data sets (fewer than 15 to 20 taxa) "branch-and-bound" algorithms are available that guarantee the recovery of the most-parsimonious tree(s). For larger data sets, however, there are no efficient algorithms that always identify the shortest trees, and therefore larger studies may report tree lengths somewhat longer than the actual minimal tree(s), therefore entailing more homoplasy. However, we consider it unlikely that this accounts for the large differences in CI observed in Figure 1. One way to investigate this bias would be to use branch-and-bound algorithms to study variation in CI among small studies. We predict that the observed relationship between CI and number of taxa would still hold.

Another bias stemming from algorithms used to reconstruct trees may be important in studies with very few taxa. Parsimony algorithms place an absolute upper limit on the amount of homoplasy that can be detected. For example, in the case of three taxa, a consistency index below 0.60 is not possible, because of the way parsimony algorithms minimize character-state changes on a tree. However, this minimum CI approaches zero rapidly as number of taxa in-

creases and is unlikely to be an important bias over the large range of number of taxa studied here.

An obvious alternative to all of the explanations given above is that systematists studying large groups are unable to find characters with low levels of homoplasy because such characters are rare. In other words, the correlation of CI with number of taxa may be primarily a reflection of the evolutionary process itself, not of some taxonomic artifact. The increase in homoplasy with number of taxa is consistent with a model of character evolution in which the probability that a character will change somewhere on a tree is monotonically related to the total number of internodes (branch segments) in the tree and, hence, to the number of taxa. Given an average nonzero probability of character-state change along internodes as branches are added, the number of state changes will tend to increase, and this will result in increased homoplasy if the number of possible states is limited. This last point is critical. If there were an unlimited number of states, then an increase in the amount of change in a character might not result in more homoplasy; instead, all changes might be to completely new states. It is the existence of a limited number of alternative states (due to some biological constraint; e.g., the existence of only four nucleotides) or the imposition of a limited number of discrete states by the systematist that may be responsible for the basic result.

The data compiled here do not allow a direct test of this model. What is needed are studies of nested subsets of taxa that are taken from within the same larger group and that differ in the numbers of terminal taxa they contain. In any case, the view that systematic characters should indicate more homoplasy in broader analyses accords well with empirical generalizations suggested by systematists. It is widely recognized, for example, that characters found to be useful in delimiting groups in one taxon will often be found less useful when additional taxa are examined (e.g., Mayr, 1969; Davis and Heywood, 1973 p. 116). In addition, characters thought to be conservative because they delimit "higher" taxa are often found to vary among closely related species when additional taxa are considered (e.g., Stebbins, 1974).

Number of Characters

There has been little explicit discussion of the relationship between homoplasy and number of characters. Although it is clear that the number of character incompatibilities in a data set can only increase as characters are added, the consistency index can increase or decrease depending on the distribution of the new character states among taxa (in contrast, when taxa are added to an analysis, CI can decline or stay the same). Archie (1985) found that CI declined slightly with the addition of characters in seven morphometric data sets.

Despite a strong correlation between number of taxa and CI and between number of characters and number of taxa, we did not find a significant correlation between number of characters and CI (Fig. 2). Nor does number of characters enter significantly into the regression estimates. When the effect of number of taxa is taken into account, studies with hundreds of characters exhibit levels of homoplasy similar to those in studies employing only a handful of characters (e.g., compare Fink [1985] with Wighton and Wilson [1987]).

This result might be interpreted as a refutation of the widespread view that the quality or reliability of phylogenetic analyses increases with the addition of characters. Indeed, if CI were considered a measure of quality, our results would cast doubt on the intuition that it is better to include more characters. However, the discussion above on the decline of CI with an increasing number of taxa suggests that this is an inappropriate interpretation, unless we suppose that studies involving more taxa are necessarily inferior. We suggest that CI is not a measure of quality, but simply a measure of the overall level of homoplasy. These attributes are logically distinct.

The standard view on the desirability of adding characters can be defended on other grounds. The reason to gather more data is not to increase the consistency index (which it does not) but, rather, to take advantage of "statistical consistency" (Felsenstein, 1978), that is, convergence on truth as a more complete sample of characters is con-

sidered. There is also a relationship between the number of characters and the confidence that can be associated with particular clades, at least when confidence is measured by the bootstrap technique (Felsenstein, 1985; Sanderson, 1989). It is important to recognize that CI and confidence are not directly related. Although confidence may in some cases be associated with the CI, it need not be. In Figure 6, we present a data set possessing a CI much lower than the expectation for that number of taxa, which nevertheless has component clades supported with high levels of confidence as measured by bootstrap resampling—higher levels in fact than are often found in real data sets (Sanderson, 1989).

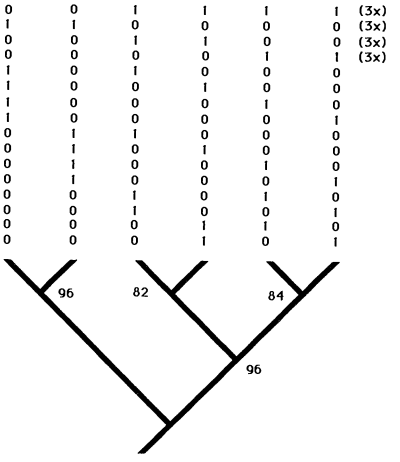

FIG. 6. Example of a data set with high homoplasy and yet high confidence, as measured by Felsenstein's (1985) bootstrap confidence assessment. The numbers along the nodes are the numbers of replicates out of 100 in which the clade occurred. The first four characters in the data set are represented three times each, so that 24 characters are included in all. The tree was constructed using the branch-and-bound algorithm in PAUP (Swofford, 1985), which produced one minimal tree of length 36 steps. The CI of 0.67 is well below empirically observed values for six taxa (see Fig. 1A), indicating substantial homoplasy.

Taxonomic Rank

There is little consensus in the literature regarding the expected relationship between homoplasy and taxonomic rank. Some authors have suggested that unravelling phylogenetic relationships among higher taxa, such as angiosperm families or mammalian orders, will be particularly difficult, owing to rampant parallelism and convergence at such levels (e.g., Cronquist, 1987 p. 46). Other authors suggest that parallelism ought to be more prevalent at lower taxonomic levels, because very closely related organisms are genetically and developmentally more similar and, therefore, more prone to parallel responses to similar selective or mutational pressures (e.g., Vavilov, 1922; Rensch, 1959; Davis and Heywood, 1973; Arnold, 1981; Stevens, 1986).

Our analysis suggests that levels of homoplasy are largely independent of taxonomic rank (Fig. 3). Homoplasy in characters used to assess relationships among higher taxa is similar to that in characters used at lower taxonomic levels. This result might simply reflect the fact that ranks are not equivalent, especially among divergent groups (e.g., species or families of angiosperms versus mammals or insects; Mishler and Donoghue, 1982; Gauthier, et al. 1988*a*). Perhaps if rank were adjusted to reflect age, as suggested by Hennig (1966), a correlation between homoplasy and rank would be discovered, with older groups showing more homoplasy. However, this expectation may be misguided. Higher taxa, if monophyletic, are conceptually equivalent to single ancestral species, and a cladistic analysis of a set of higher taxa can be viewed as an analysis of their ancestral species. If there is no reason to expect homoplasy to be any more or less prevalent among such ancestral species than it is among a set of recent species, then levels of homoplasy should be about equal at different ranks.

We suspect that intuition on the subject of rank has often been mistaken for two reasons. First, higher taxa are not viewed as ancestral species. Instead the myriad polymorphisms, parallelisms, and reversals, which seem characteristic of all large taxa, are taken into consideration, even if only subconsciously. Consideration of such variation is operationally equivalent to the inclusion of all the taxa within the higher group, which leads to an increase in homoplasy owing to the increase in number of taxa (see above). Also, when higher taxa reflect old lineages, it might be more likely that extensive homoplasy would have time to appear. However, if there are few terminal taxa, then that homoplasy (present or

not) would be difficult to detect, given the upper limit on homoplasy detectable by parsimony.

These conclusions do not contradict the observation that any particular character might be expected to show progressively higher levels of homoplasy as the scope of a study is expanded upward in rank. In practice, systematists undoubtedly choose their character sets to reflect an appropriate amount of variation for the rank of the terminal taxa used. Although workers using molecular data are more constrained than morphologists in this regard, they too can choose the appropriate molecule or the appropriate set of restriction enzymes (some restriction mutations appear more or less frequently in a given genome). Hence, in all cases, the relative invariance of levels of homoplasy to taxonomic rank may reflect compensation on the part of taxonomists.

Plant versus Animal Data

In contrast to the factors discussed above, there has been much commentary on the level of homoplasy in plants versus animals. In particular, it is widely believed that plants are more homoplastic than most animals. Thus, according to Wagner (1984 p. 115), "plants are evidently unusually inclined towards parallelisms." Likewise, Cronquist (1987 p. 24) suggested that ". . . the relative lack of morphological integration in plants, and poor correlation of evolutionary advances with adaptive zones and ecologic niches, combine to permit rampant parallelism, in contrast to the more rigid (though imperfect) evolutionary channelling in animals." Accounting for increased homoplasy by reference to the relative simplicity and indeterminate growth of plants has been an especially popular argument. For example, Stebbins (1974 p. 143), observing that ". . . patterns of development of individual organs are, in general, much simpler in plants than in animals," concluded that ". . . similar but independent evolutionary modifications of structures . . . are much more common in plants than in animals." Similarly, Gottleib (1984) postulated that relatively few genes of major effect may underlie character differences between plant species and that plants may therefore be more prone to parallelism than animals.

Such sentiments have led to the widespread belief that phylogeny reconstruction is more difficult in plants than in animals (e.g., Funk, 1981 p. 73; Davis and Heywood, 1973 p. 112). Stebbins (1950 p. 506) presumed, for instance, that ". . . morphological similarity is much less indicative of phylogenetic relationships in plants than it is in animals," and Cronquist (1987 p. 24) pointed to the "difference in amount of parallelism in plants and animals" in commenting on the reluctance among botanists to embrace cladistic methods. The perception of rampant homoplasy in plant morphological characters has even led to the suggestion that only molecular data will provide characters with acceptable levels of homoplasy (Sytsma and Gottlieb, 1986*b*).

Despite considerable interest in this issue, levels of homoplasy in plants and animals have never been compared quantitatively. Surprisingly, our comparison indicates that levels of homoplasy are remarkably similar; the regressions of consistency index on number of taxa match closely (Fig. 4). We conclude from this that there is little difference in the average level of homoplasy between plants and animals, at least among those characters actually selected by systematists in analyzing cladistic relationships.

We can think of two explanations for this result. First, it is possible that CI fails to capture what most biologists have in mind when they use the term "homoplasy." Perhaps if all characters (or a truly random sample of characters) were considered, plants would indeed be more homoplastic than animals. Systematists often eliminate characters prior to phylogenetic analysis, and it is possible that plant systematists eliminate more characters, because a higher proportion of characters are excessively variable in plants than in animals. This is a difficult proposition to test, because the number of "rejected" characters is not usually reported in published studies. Unfortunately, one cannot simply compare the average number of characters included in plant studies with the average number included in animal studies. Plant studies are generally based on fewer characters, but it is well known that animals (at least the "higher" metazoa) have many more basic tissue types than plants

(e.g., Gottlieb, 1984) and consequently are likely to have many more taxonomic characters. Alternatively, preconceived ideas about homoplasy may have been incorrect. Animals and plants may in fact be approximately equally homoplastic. It is not possible at this time to distinguish between these two alternative hypotheses, which is unfortunate from the perspective of attempting to understand the interesting evolutionary implications of homoplasy. On the other hand, botanists should find these results encouraging, because regardless of which of the above explanations is correct, it is clearly possible to gather character data on plants that exhibit as little homoplasy as data on animals. Indeed, plant systematists have been doing it for years.

Morphological versus Molecular Data

The value of morphological data in elucidating phylogeny has frequently been questioned because of the susceptibility of morphology to convergent evolution by natural selection. Some have suggested that molecular data can and will resolve phylogenetic questions that morphology cannot (e.g., Hillis, 1987; Patterson, 1987; Sibley and Ahlquist, 1987 p. 118; Bobrova et al., 1987; Olmstead, 1989). For example, Sytsma and Gottlieb (1986*b* p. 5556) state that a ". . . primary reliance on morphological data to model phylogenetic relationships may be misleading, no matter how many characters are examined." Furthermore, molecular systematists have touted high CI's in support of the contention that such data are less homoplastic (e.g., Jansen and Palmer, 1988 p. 764), although these CI's are substantially lower when numerous autapomorphies are removed (Table 1). On the other hand, some systematists have pointed out difficulties in establishing homology in molecular characters (e.g., Patterson, 1987 p. 18), and it has even been suggested (Mishler et al., 1988) that molecular data may show *more* homoplasy, due to the limited number of character states per character (locus) and the difficulty in eliminating homoplasy by careful a priori character analysis (in contrast to morphological characters which undergo development).

A few studies have compared levels of homoplasy in molecular data with those in morphological data for particular taxa. Wyss et al. (1987) examined homoplasy in amino-acid sequences of mammal orders and found it comparable to that seen in morphological data. In a similar analysis of amino-acid sequences in angiosperms, Bremer (1988) concluded that levels of homoplasy were too high to yield robust phylogenies. Our study is more comprehensive in that it includes 18 molecular studies and 42 morphological studies, and it considers the effects of other factors, such as the numbers of taxa and characters. We find little support for the belief that molecular data show either lower or higher levels of homoplasy (Fig. 5). However, because our sample of molecular studies is still small and because of the high variability in consistency indexes, the regression estimates for molecular studies are not statistically significant (Table 3). Had they been significant, they would have implied that molecular data are more homoplastic than morphological data in smaller studies and less homoplastic in larger ones. Unfortunately, small sample sizes prevented us from examining further subsets of the "molecular" data, to see, for example, whether sequence and restriction-site data are more or less homoplastic than protein electrophoretic data. We do find it suggestive, however, that the large study of Jansen and Palmer (with 55 taxa) has a much lower CI than other restriction-fragment studies with fewer taxa.

In general, our result is not a criticism of the use of molecular data. Molecular data may prove to be better than morphological data for reasons other than lower levels of homoplasy. Perhaps the ability to generate large numbers of characters will improve the statistical consistency of the tree-estimation process (Felsenstein, 1978), and less ambiguous character-state assignments may reduce artifacts introduced through human error. However, there is no evidence to date that molecular data are less homoplastic than morphological data.

Conclusions

Our comparison of consistency indexes, which has become possible with the availability of large numbers of cladistic studies, allows a test of hypotheses about character evolution and exposes unexpected patterns

that require evolutionary explanation. We have found that CI is strongly correlated with number of taxa but not with number of characters or taxonomic rank. Moreover, when other factors are taken into account, there does not appear to be any significant difference in homoplasy between plants and animals or between morphological and molecular data.

Many other comparisons are suggested by this study. First, more refined analyses based on a larger sample of studies will soon be possible. The contrast of "morphological" versus "molecular" data sets is crude and can be refined as studies become available. It would be useful to compare morphological versus behavioral, reproductive versus nonreproductive, sequence versus electrophoretic, and nuclear versus organellar genome data. Second, CI could be decomposed into its components (parallelism and reversal) as has been done in some individual studies (Eckenwalder and Barrett, 1986; Gauthier, et al. 1988*b*). Finally, it should be possible to compare patterns in the topological distribution of homoplasy: for instance, the localization of homoplasy to particular regions of a tree and the topological correlation of homoplastic changes in different characters.

Analyses along these lines, and any other quantitative comparisons of homoplasy using the consistency index, should take into account the strong relationship of CI to the number of taxa studied. Our results provide an initial bench mark for comparative purposes. Inferences based on levels of homoplasy in particular groups must be drawn with the knowledge that the size of the group is a critical factor. This has rarely been recognized. Thorpe and Dickinson (1988) studied the use of regulatory sequences in reconstructing phylogenies in *Drosophila* and concluded that ". . . the amount of homoplasy within the groups of closely related species is sufficiently small that regulatory proteins may be useful for inferring relationships at this level . . ." (p. 97). Their conclusions were based on a comparison of CI's in small subsets of taxa relative to the CI of the overall data set, which in large part reflect differences in the numbers of taxa involved. Systematists studying large groups frequently encounter what seem to be inordinately high levels of homoplasy (Cronquist, 1987 p. 32). For example, Funk (1981 p. 82) suggested that the level of parallelism in the genus *Montanoa* had reached "staggering proportions." In fact, the consistency index that we calculate from her cladogram (0.46) is only slightly below the value of 0.48 expected with the 25 taxa included in her study. The observed deviation is well within the standard error of the predicted value based on the regression parameters in Table 3 (Snedecor and Cochran, 1980 p. 166). The conclusion that homoplasy is extreme should not be based on the absolute value of CI, but on a significant deviation from the regression line in Figure 1.

Acknowledgments

J. D. Palmer, R. K. Jansen, P. F. Stevens, E. Zimmer, and W. B. Heed kindly provided information on their analyses in press. We are grateful to R. E. Strauss and B. Walsh for statistical advice and helpful comments on the manuscript. W. H. Wagner, P. F. Stevens, J. M. Porter, A. Harvey, and S. Suter provided stimulating comments on the ideas proposed here, and two anonymous reviewers provided useful criticisms and suggestions. This research was facilitated by an NSF grant awarded to M.J.D. (BSR-8414450).

Literature Cited

Anderberg, A. 1986. The genus *Pegolettia* (Compositae:Inuleae). Cladistics 2:158–186.

Archie, J. 1985. Methods for coding variable morphological features for numerical taxonomic analyses. Syst. Zool. 34:326–345.

Arnold, E. N. 1981. Estimating phylogenies at low taxonomic levels. Zeitschr. Zool. Syst. Evol-Forsch. 19:1–35.

Baum, B. R. 1983. A phylogenetic analysis of the tribe Triticeae (Poaceae) based on morphological characters of the genera. Can. J. Bot. 61:518–535.

———. 1984. Application of compatibility and parsimony methods at the infraspecific, specific, and generic levels in Poaceae, pp. 192–220. *In* T. Duncan and T. Stuessy (eds.), Cladistics: Perspectives on the Reconstruction of Evolutionary History. Columbia Univ. Press, N.Y.

Baum, B. R., and D. B. O. Saville. 1985. Rusts (Uredinales) of Triticeae: Evolution and extent of coevolution, a cladistic analysis. Bot. J. Linn. Soc. 91:367–394.

Bobrova, V., A. Troitsky, A. Ponomareo, and A. Antonov. 1987. Low molecular weight rRNA sequences and plant phylogeny reconstruction: Nu-

cleotide sequences of chloroplast 4.5S rRNAs from *Acoris calamus* (Araceae) and *Ligularia calthifolia* (Asteraceae). Plant Syst. Evol. 156:13–27.

BREMER, K. 1987. Tribal interrelationships of the Asteraceae. Cladistics 3:210–253.

———. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution 42:795–803.

BREMER, K., C. J. HUMPHRIES, B. D. MISHLER, AND S. P. CHURCHILL. 1987. On cladistic relationships in green plants. Taxon 36:339–349.

BROOKS, D., R. DANIEL, R. T. O'GRADY, AND D. R. GLEN. 1985. The phylogeny of the Cercomeria Brooks, 1982 (Platyhelminthes). Proc. Helminthol. Soc. Wash. 52:1–20.

CAMPBELL, C. S. 1986. Phylogenetic reconstructions and two new varieties in the *Andropogon virginicus* complex (Poaceae: Andropogoneae). Syst. Bot. 11: 280–292.

CANE, J. H. 1983. Preliminary chemosystematics of the Andrenidae and exocrine lipid evolution of the short-tongued bees (Hymenoptera:Apoideae). Syst. Zool. 32:417–430.

CARPENTER, J. M. 1987. Phylogenetic relationships and classification of the Vespinae (Hymenoptera: Vespidae). Syst. Entomol. 12:413–431.

———. 1988. Choosing among multiple equally parsimonious trees. Cladistics 4:291–296.

CHURCHILL, S. P., E. O. WILEY, AND L. A. HAUSER. 1984. A critique of Wagner groundplan-divergence studies and a comparison with other methods of phylogenetic analysis. Taxon 33:212–232.

COLLETTE, B. B., AND J. L. RUSSO. 1985. Interrelationships of the Spanish mackerels (Pisces: Scombridae: *Scomberomorus*) and their copepod parasites. Cladistics 1:141–158.

CRACRAFT, J. 1985. Monophyly and phylogenetic relationships of the Pelicaniformes: A numerical cladistic analysis. Auk 102:835–853.

———. 1986*a*. Origin and evolution of continental biotas: Speciation and historical congruence within the Australian avifauna. Evolution 40:977–996.

———. 1986*b*. The origin and early diversification of birds. Paleobiology 12:383–399.

CRANE, P. R. 1985. Phylogenetic relationships in seed plants. Cladistics 1:329–348.

CRONQUIST, A. 1987. A botanical critique of cladism. Bot. Rev. 53:1–52.

CROTHER, B. I., M. M. MIYAMOTA, AND W. F. PRESCH. 1986. Phylogeny and biogeography of the lizard family Xantusidae. Syst. Zool. 35:37–45.

CUTLER, E. B., AND P. E. GIBBS. 1985. A phylogenetic analysis of higher taxa in the Sipuncula. Syst. Zool. 34:162–173.

DAHLGREN, R., AND K. BREMER. 1985. Major clades of the angiosperms. Cladistics 1:349–368.

DAHLGREN, R., AND F. RASMUSSEN. 1983. Monocotyledon evolution: Characters and phylogenetic estimation. Evol. Biol. 16:255–395.

DAVIS, P., AND V. HEYWOOD. 1973. Principles of Angiosperm Taxonomy. Krieger, Huntington, NY.

DOYLE, J. A., AND M. J. DONOGHUE. 1986. Seed plant phylogeny and the origin of angiosperms: An experimental cladistic approach. Bot. Rev. 52:321–431.

ECKENWALDER, J. E., AND S. C. H. BARRETT. 1986. Phylogenetic systematics of the Pontederiaceae. Syst. Bot. 11:373–391.

EMERSON, S. B. 1988. Testing for historical patterns of change: A case study with frog pectoral girdles. Paleobiology 14:174–186.

FARRIS, J. S. 1969. A successive approximations approach to character weighting. Syst. Zool. 18:374–385.

———. 1983. The logical basis of phylogenetic analysis, pp. 7–36. *In* N. Platnick and V. Funk (eds.), Advances in Cladistics, Vol. II. Columbia Univ. Press, N.Y.

FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility will be positively misleading. Syst. Zool. 34:152–161.

———. 1982. Numerical methods for inferring evolutionary trees. Quart. Rev. Biol. 57:127–141.

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

FINK, W. L. 1985. Phylogenetic relationships of the stomiid fishes (Teleostei:Stomiiformes). Misc. Publ. Mus. Zool. Univ. Mich. 171:1–127.

FINK, W. L., AND S. V. FINK. 1986. A phylogenetic analysis of the genus *Stomias,* including the synonymy of *Macrostomias.* Copeia 1986:494–503.

FUNK, V. A. 1981. Special concerns in estimating plant phylogenies, pp. 73–86. *In* V. Funk and D. Brooks (eds.), Advances in Cladistics. Oxford Univ. Press, Oxford, U.K.

———. 1982. The systematics of *Montanoa* (Asteraceae: Heliantheae). Mem. New York Bot. Gard. 36:1–133.

FUTUYMA, D. J. 1986. Evolutionary Biology, 2nd ed. Sinauer, Sunderland, MA.

GABRIELSON, P. W., AND D. J. GARBARY. 1987. A cladistic analysis of Rhodophyta: Floridophycean orders. Brit. Phycol. J. 22:125–138.

GAUTHIER, J. A. 1986. Saurischian monophyly and the origin of birds, pp. 1–55. *In* K. Padian (ed.), The Origin of Birds and the Evolution of Flight. California Academy of Sciences, San Francisco.

GAUTHIER, J., R. ESTES, AND K. DE QUEIROZ. 1988*a*. A phylogenetic analysis of Lepidosauromorpha, pp. 15–98. *In* R. Estes and G. Pregill (eds.), Phylogenetic Relationships of the Lizard Families: Essays Commemorating Charles L. Camp. Stanford Univ. Press, Stanford, CA.

GAUTHIER, J. A., A. G. KLUGE, AND T. ROWE. 1988*b*. Amniote phylogeny and the importance of fossils. Cladistics 4:105–209.

GEORGE, M., AND O. A. RYDER. 1986. Mitochondrial DNA evolution in the genus *Equus.* Molec. Biol. Evol. 3:535–546.

GOTTLIEB, L. D. 1984. Genetics and morphological evolution in plants. Amer. Natur. 123:681–709.

GRISMER, L. 1983. A reevaluation of the North American Gekkonid genus *Anarbylus* Murphy and its cladistic relationships to *Coleonyx* Gray. Herpetologica 39:394–399.

GUYER, C., AND J. M. SAVAGE. 1986. Cladistic relationships among anoles (Sauria:Iguanidae). Syst. Zool. 35:509–531.

HAMBY, R. K., AND E. A. ZIMMER. 1988. Ribosomal RNA sequences for inferring phylogeny within the

grass family (Poaceae). Plant Syst. Evol. 34:393–400.

HENNIG, W. 1966. Phylogenetic Systematics. Univ. Illinois Press, Urbana, IL.

HERMAN, L. H. 1986. Revision of *Bledius,* Part IV. Classification of species groups, phylogeny, natural history, and catalogue (Coleoptera, Staphylinidae, Oxytelenae). Bull. Amer. Mus. Nat. Hist. 184:1–368.

HILLIS, D. M. 1987. Molecular versus morphological approaches to systematics. Ann. Rev. Ecol. Syst. 18:23–42.

HILLIS, D. M., AND S. K. DAVIS. 1986. Evolution of ribosomal DNA: Fifty million years of recorded history in the frog genus *Rana.* Evolution 40:1275–1288.

HUEY, R., AND A. BENNETT. 1987. Phylogenetic studies of coadaptation: Preferred temperatures versus optimal performance temperatures of lizards. Evolution 41:1098–1115.

HUFFORD, L. D. 1988. The evolution of floral morphological diversity in *Eucnide* (Loasaceae): The implications of modes and timing of ontogenetic changes on phylogenetic diversification, pp. 103–119. *In* P. Leins, S. C. Tucker, and P. K. Endress (eds.), Aspects of Floral Development. Cramer, Berlin, W. Ger.

JAMIESON, B. G. M., C. ERSEUS, AND M. FERRAGUTI. 1987. Parsimony analysis of the phylogeny of some Oligochaeta (Annelida) using spermatazoal ultrastructure. Cladistics 3:145–155.

JANSEN, R. K. 1981. Systematics of *Spilanthes* (Compositae: Heliantheae). Syst. Bot. 6:231–257.

JANSEN, R. K., AND J. PALMER. 1988. Phylogenetic implications of chloroplast DNA restriction site variation in the Mutisieae (Asteraceae). Amer. J. Bot. 75:753–766.

KELLOGG, E., AND C. CAMPBELL. 1987. Phylogenetic analysis of the Gramineae, pp. 310–322. *In* T. Soderstrom, K. W. Hilu, C. S. Campbell, and M. E. Barkworth (eds.), Grass Systematics and Evolution. Smithsonian Inst. Press, Wash., DC.

KITCHING, I. 1987. Spectacles and silver Y's: A synthesis of the systematics, cladistics, and biology of the Plusiinae (Lepidoptera: Noctuidae). Bull. British Mus. Nat. Hist. (Entomol.) 54:75–261.

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

LADIGES, P. Y., AND C. J. HUMPHRIES. 1983. A cladistic study of *Arillastrum, Angophora,* and *Eucalyptus* (Myrtaceae). Bot. J. Linn. Soc. 87:105–134.

LADIGES, P. Y., C. J. HUMPHRIES, AND M. BROOKES. 1987. Cladistic and biogeographic analysis of Western Australian species of *Eucalyptus* L'Herit, informal subgenus *Monocalyptus* Pryor and Johnson. Austral. J. Bot. 35:251–281.

LANKESTER, E. 1870. On the use of the term homology in modern zoology. Ann. Mag. Nat. Hist. (Ser. 4) 6:34–43.

LAUDER, G. 1981. Form and function: Structural analysis in evolutionary morphology. Paleobiology 7:430–442.

———. 1982. Historical biology and the problem of design. J. Theoret. Biol. 97:57–67.

LIPSCOMB, D. L. 1985. The eukaryotic kingdoms. Cladistics 1:127–140.

LIVEZEY, B. C. 1986. Phylogeny and historical biogeography of steamer-ducks (Anatidae:Tachyeres). Syst. Zool. 35:458–469.

MAYR, E. 1969. Principles of Systematic Zoology. McGraw-Hill, N.Y.

MEACHAM, C. A. 1981. A probability measure for character compatibility. Math. Biosci. 57:1–18.

———. 1984. Evaluating characters by character compatibility analysis, pp. 152–165. *In* T. Duncan and T. Stuessy (eds.), Cladistics: Perspectives on the Reconstruction of Evolutionary History. Columbia Univ. Press, N.Y.

MISHLER, B., K. BREMER, C. HUMPHRIES, AND S. CHURCHILL. 1988. The use of nucleic acid sequence data in phylogenetic reconstruction. Taxon 37:391–395.

MISHLER, B., AND S. P. CHURCHILL. 1985. Transition to a land flora: Phylogenetic relationships of the green algae and bryophytes. Cladistics 1:305–328.

MISHLER, B., AND M. J. DONOGHUE. 1982. Species concepts: A case for pluralism. Syst. Zool. 31:491–503.

MIYAMOTO, M. M. 1983. Biochemical variation in the frog *Eleutherodactylus brandfordii*: Geographical patterns and cryptic species. Syst. Zool. 32:43–51.

MONTGOMERY, D., AND E. PECK. 1982. Introduction to Linear Regression Analysis. Wiley, N.Y.

NELSON, C. H. 1984. Numerical cladistic analysis of phylogenetic relationships in Plecoptera. Ann. Entomol. Soc. Amer. 77:466–473.

OLMSTEAD, R. 1989. Phylogeny, phenotypic evolution, and biogeography of the *Scutellaria angustifolia* complex (Lamiaceae): Inference from morphological and molecular data. Syst. Bot. 14:320–338.

PATTERSON, C. 1982. Morphological characters and homology, pp. 21–74. *In* K. A. Joysey and A. E. Friday (eds.), Problems of Phylogenetic Reconstruction. Academic Press, London, U.K.

———. 1987. Introduction, pp. 1–22. *In* C. Patterson (ed.), Molecules and Morphology in Evolution: Conflict or Compromise? Cambridge Univ. Press, Cambridge, U.K.

REMANE, A. 1952. Die Grundlagen des Natürlichen Systems, der vergleichenden Anatomie und der Phylogenetik. Akademische Verlagsgesellschaft, Leipzig, E. Ger.

RENSCH, B. 1959. Evolution Above the Species Level. Columbia Univ. Press, N.Y.

RIDLEY, M. 1983. The Explanation of Organic Diversity: The Comparative Method and Adaptations for Mating. Clarendon, Oxford, U.K.

RIGGINS, R., AND J. S. FARRIS. 1983. Cladistics and the root of angiosperms. Syst. Bot. 8:96–101.

RODMAN, J. E., M. OLIVER, R. NAKAMURA, J. U. MCCLAMMER, JR., AND A. H. BLEDSOE. 1984. A taxonomic analysis and revised classification of Centrospermae. Syst. Bot. 9:297–323.

SANDERSON, M. J. 1989. Confidence limits in phylogenies: The bootstrap revisted. Cladistics 5:113–129.

SCHUH, R. T., AND J. T. POLHEMUS. 1980. Analysis of taxonomic congruence among morphological, ecological, and biogeographic data sets for the Leptopodomorpha (Hemiptera). Syst. Zool. 29:1–26.

SESSIONS, S., AND A. LARSON. 1987. Developmental correlates of genome size in Plethodontid salaman-

ders and their implications for genome evolution. Evolution 41:1239–1251.

SIBLEY, C., AND J. AHLQUIST. 1987. Avian phylogeny reconstructed from comparisons of the genetic material, DNA, pp. 95–121. *In* C. Patterson (ed.), Molecules and Morphology in Evolution: Conflict or Compromise? Cambridge Univ. Press, Cambridge, U.K.

SILLÉN-TULLBERG, B. 1988. Evolution of gregariousness in aposematic butterfly larvae: A phylogenetic analysis. Evolution 42:293–305.

SITES, J. W., J. W. BICKHAM, B. A. PYTEL, I. F. GREENBAUM, AND B. A. BATES. 1984. Biochemical characters and reconstruction of turtle phylogenies: Relationships among Batagurine genera. Syst. Zool. 33:137–158.

SNEDECOR, G. W., AND W. G. COCHRAN. 1980. Statistical Methods. Iowa State Univ. Press, Ames.

SOKAL, R., AND K. SHAO. 1985. Character stability in 39 data sets. Syst. Zool. 34:83–89.

STEBBINS, G. L. 1950. Variation and Evolution in Plants. Columbia Univ. Press, N.Y.

———. 1974. Flowering Plants. Evolution Above the Species Level. Harvard Univ. Press, Cambridge, MA.

STEVENS, P. F. 1986. Evolutionary classification in botany, 1960–1985. J. Arnold Arbor. 67:313–339.

STRANEY, D. 1981. The stream of heredity: Genetics in the study of phylogeny, pp. 100–138. *In* M. Smith and J. Joule (eds.), Mammalian Population Genetics. Univ. Georgia Press, Athens.

SWOFFORD, D. 1985. PAUP: Phylogenetic Analysis Using Parsimony, Version 2.4. Illinois Natural History Survey, Champaign, IL.

SYTSMA, K. J., AND L. D. GOTTLIEB. 1986*a*. Chloroplast DNA evolution and phylogenetic relationships in *Clarkia* sect. *Peripetasma* (Onagraceae). Evolution 40:1248–1261.

———. 1986*b*. Chloroplast DNA evidence for the origin of the genus *Heterogaura* from a species of *Clarkia* (Onagraceae). Proc. Nat. Acad. Sci. USA 83:5554–5557.

THORPE, P. A., AND W. J. DICKINSON. 1988. The use of regulatory patterns in constructing phylogenies. Syst. Zool. 37:97–105.

VAVILOV, N. 1922. The law of homologous series in variation. J. Genet. 12:47–89.

VILGALYS, R. 1986. Phenetic and cladistic relationships in *Collybia* sect. *Levipedes* (Fungi:Basidiomycetes). Taxon 35:225–233.

WAGNER, W. H. 1984. Applications of the concepts of groundplan-divergence, pp. 95–118. *In* T. Duncan and T. Stuessy (eds.), Cladistics: Perspectives on the Reconstruction of Evolutionary History. Columbia Univ. Press, N.Y.

WAKE, D. B., AND A. LARSON. 1987. Multidimensional analysis of an evolving lineage. Science 238: 42–48.

WIGHTON, D. C., AND M. V. H. WILSON. 1987. The Gomphaeshninae (Odonata: Aeshnidae): New fossil genus, reconstructed phylogeny, and geographical history. Syst. Entomol. 11:505–522.

WYSS, A. R., M. J. NOVACEK, AND M. C. MCKENNA. 1987. Amino acid sequence versus morphological data and the interordinal relationships of mammals. Molec. Biol. Evol. 4:99–116.

Corresponding Editor: K. J. Niklas

APPENDIX

Correction of Consistency Index for Autapomorphies and Invariant Characters

In the case of binary characters, a character was omitted from the calculation if all of the taxa in the study were scored as having the same state (invariant) or if only one taxon was scored as having the presumed derived state (autapomorphic). This was done even when one or more taxa were scored as "unknown" for the character in question. In such cases, parsimony algorithms assign the most parsimonious state to the questionable taxon based upon its position in the tree determined from other character data, and thus there is no possibility of inconsistency in that character. If only one taxon possessed the presumed ancestral (plesiomorphic) state, the character was retained, on the grounds that it might still be inconsistent if it were later found to be most parsimonious to assume a transition to the derived state between the outgroup(s) and the ingroup and a reversal to the outgroup condition within the ingroup cladogram. We assumed that outgroups were consulted in assessing character polarities, unless it was explicitly stated that some other method was employed. If a character was explicitly treated as undirected (polarity unknown), then it was eliminated if only one taxon had a particular state, whether this was coded as 0 or 1 in the matrix.

The same rules were also applied in standardizing multistate characters that were treated as *unordered,* that is, entailing only one step between any pair of states. However, in such cases entire characters were not eliminated; instead, a step was subtracted from the numerator and denominator in calculating the consistency index when only one taxon possessed a particular derived state. In contrast, in the case of *ordered* multistate characters, steps were not subtracted when only one taxon possessed a derived state, on the grounds that the derivation of that state could entail more than one step. This would be the case if the states possessed by closely related taxa in the tree under consideration were more than one step removed in the ordered transformation series.