9-2001

Inferring and Testing Hypotheses of Cladistic Character Dependence by Using Character Compatibility

F. Robin O'Keefe
Marshall University, okeefef@marshall.edu

Peter J. Wagner

Recommended Citation
O'Keefe, F. R., and P. J. Wagner. 2001. Inferring and testing hypotheses of cladistic character dependence by using character compatibility. Syst. Biol. 50:657–675.

Syst. Biol. 50(5):657–675, 2001

Inferring and Testing Hypotheses of Cladistic Character Dependence by Using Character Compatibility

F. ROBIN O'KEEFE[1] AND PETER J. WAGNER[2]

[1]*Department of Anatomy, New York College of Osteopathic Medicine, New York Institute of Technology, Old Westbury, New York 11568-8000, USA; E-mail: rokeefe@iris.nyit.edu*
[2]*Department of Geology, Field Museum of Natural History, Roosevelt Road at Lake Shore Drive, Chicago, Illinois 60605, USA; E-mail: pwagner@fmnh.org*

Abstract.—The notion that two characters evolve independently is of interest for two reasons. First, theories of biological integration often predict that change in one character requires complementary change in another. Second, character independence is a basic assumption of most phylogenetic inference methods, and dependent characters might confound attempts at phylogenetic inference. Previously proposed tests of correlated character evolution require a model phylogeny and therefore assume that nonphylogenetic correlation has a negligible effect on initial tree construction. This paper develops "tree-free" methods for testing the independence of cladistic characters. These methods can test the character independence model as a hypothesis before phylogeny reconstruction, or can be used simply to test for correlated evolution. We first develop an approach for visualizing suites of correlated characters by using character compatibility. Two characters are compatible if they can be used to construct a tree without homoplasy. The approach is based on the examination of mutual compatibilities between characters. The number of times two characters i and j share compatibility with a third character is calculated, and a pairwise shared compatibility matrix is constructed. From this matrix, an association matrix analogous to a dissimilarity matrix is derived. Eigenvector analyses of this association matrix reveal suites of characters with similar compatibility patterns. A priori character subsets can be tested for significant correlation on these axes. Monte Carlo tests are performed to determine the expected distribution of mutual compatibilities, given various criteria from the original data set. These simulated distributions are then used to test whether the observed amounts of nonphylogenetic correlation in character suites can be attributed to chance alone. We have applied these methods to published morphological data for caecilian amphibians. The analyses corroborate instances of dependent evolution hypothesized by previous workers and also identify novel partitions. Phylogenetic analysis is performed after reducing correlated suites to single characters. The resulting cladogram has greater topological resolution and implies appreciably less change among the remaining characters than does a tree derived from the raw data matrix. [Character independence; character weighting; compatibility; correlated character evolution; similarity coefficient.]

A hypothesis of correlated character evolution, that is, that change in one character depends on conditions of another character, is of interest for both theoretical and methodological reasons. Theories from developmental (Wake, 1989), functional (Wainwright et al., 1975), architectural (Raup, 1966), and molecular (Huelsenbeck and Nielsen, 1999) biology all predict that correlated change should be common. However, phylogenetic inference methods such as parsimony (Edwards and Cavalli-Sforza, 1964; Kluge and Farris, 1969)and simple maximum likelihood (Felsenstein, 1973) assume independent character change. Thus, workers have long recognized that character correlation (also termed lack of character independence or character oversplitting) is a central issue in character selection (Sneath and Sokal, 1973). If characters evolve in a correlated manner, the characters in the correlated suite are effectively overweighted (de Queiroz, 1993; Chippindale and Wiens, 1994). Simulations indicate that parsimony tree topologies and tree lengths are less accurate when character evolution is correlated rather than independent (Wagner, 1998; Huelsenbeck and Nielsen, 1999). Correlated character evolution might also exaggerate bootstrap and Bremer support values for some nodes by inflating apparent numbers of synapomorphies.

Sneath and Sokal (1973) give two reasons why characters might be correlated: simple logical correlation arising from the definitions of the characters themselves, and correlation arising from the biology of the organisms under study. Judicious character selection can eliminate logical correlations among characters. Biological correlations are more difficult to identify but have been suggested for many groups of organisms (see Emerson and Hastings [1998] for a review and discussion). Examples include

tooth characters in hyaenids (Werdelin and Solounias, 1991; Wagner, 1998), eye characters in caecilians (Wilkinson, 1997), reproductive characters in wasps (Quicke and Belshaw, 1999), diving characters in aquatic birds (McCracken et al., 1999), and gross body morphology among plesiosaurs (O'Keefe, 2000). Turner (1974), Shaffer et al. (1991), and McCracken et al. (1999) both suggest that conflict between different data partitions is evidence of correlated character change within one or more partitions. Suter (1994) suggests that correlated character evolution might explain multiple tree "islands"—sets of very distinct topologies implying the same or similar amounts of change (Maddison, 1991). The study of correlated change is therefore important for methodological as well as biological reasons.

One approach to dealing with this problem is to partition characters into suspicious subsets and find the parsimony trees supported by each character subset (Wray, 1996; Wilkinson, 1997; Emerson and Hastings, 1998; McCracken et al., 1999; Quicke and Belshaw, 1999). One then reconciles the resulting trees in some way (e.g., taxonomic congruence; Mickevich, 1978). Unfortunately, such approaches are laborious and again require prior hypotheses of character correlation. Workers have proposed tree-based tests for detecting correlated character evolution (e.g., Felsenstein, 1985; Maddison, 1990; Pagel, 1994). However, tree-based tests require a model phylogeny. Because most phylogenetic methods assume character independence, then to avoid circularity, one should infer the model phylogeny by using some character set other than the one being tested. Ultimately, one must assume independence for some subset of characters. If only one data set is available (e.g., only morphologic characters for fossil taxa), then one must make reliability assumptions about particular characters (O'Leary and Geisler, 1999). A second problem is that if taxa are extinct or otherwise poorly known, we often lack a firm basis for suspecting correlated change.

This paper develops a method designed to discover correlated character suites before tree building without requiring prior biological knowledge. Structure among characters such as character compatibility (Camin and Sokal, 1965; Le Quesne, 1969) should be in large part the result of phylogenetic autocorrelation among characters (Sneath and Sokal, 1973; Raup and Gould, 1974; Felsenstein, 1985). However, correlated change should induce secondary signals among suites of correlated characters. The primary focus of this paper is to determine whether these secondary signals exceed the expectations of phylogeny and independent character evolution. First, we focus on inferring such suites, using multivariate analyses of compatibility patterns among characters. Second, we develop Monte Carlo methods to determine whether observed secondary signals exceed those expected given independent change over a phylogeny.

Inferring Suites of Correlated Characters

Character Compatibility: A Review

Two cladistic characters are compatible if a tree exists upon which they can be mapped without homoplasy (Camin and Sokal, 1965; Le Quesne, 1969, 1982) (Fig. 1). Estabrook et al. (1976) proved two theorems relating to character compatibility. First, two binary characters are compatible if and only if they do not possess all four possible character state distributions (Fig. 1), or, if all four relations exist, they form a circuit and the characters are not compatible. Second, two sets of binary characters are compatible only if they are pairwise compatible. Because one can extend both theorems to ordered multistate characters (McMorris, 1975), one can

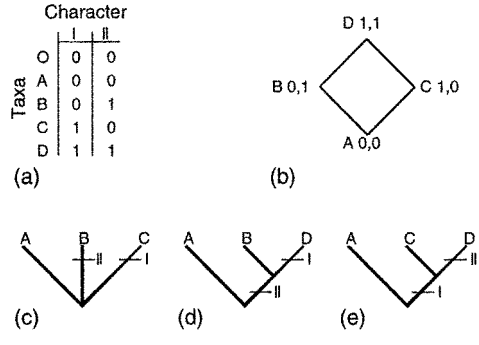

FIGURE 1. Definition of noncompatible characters. (a) Example matrix with 4 taxa and 2 characters. All four possible conditions of the two characters are observed, forming a circuit (<>) (b). Although any three taxa can be placed on a tree without invoking homoplasy, addition of the fourth always requires a parallelism or a reversal (c–e). Modified from Estabrook et al. (1976).

R

Characters

Taxon	A	B	C	D	E	F	G	H	I	J
I	0	0	0	0	0	0	0	0	0	0
II	0	0	0	0	0	0	0	1	0	0
III	1	0	0	0	0	0	0	0	0	0
IV	1	1	0	0	0	0	0	1	1	0
V	1	1	1	0	0	0	0	0	0	0
VI	1	1	1	1	0	0	0	1	1	1
VII	1	1	1	1	1	0	0	0	0	0
VIII	1	1	1	1	1	1	0	1	1	1
IX	1	1	1	1	1	1	1	0	0	0
X	1	1	1	1	1	1	1	1	1	1

C

Characters

Char.	A	B	C	D	E	F	G	H	I	J
A	•	1	1	1	1	1	1	0	1	1
B	1	•	1	1	1	1	1	0	1	1
C	1	1	•	1	1	1	1	0	0	1
D	1	1	1	•	1	1	1	0	0	1
E	1	1	1	1	•	1	1	0	0	0
F	1	1	1	1	1	•	1	0	0	0
G	1	1	1	1	1	1	•	0	0	0
H	0	0	0	0	0	0	0	•	1	1
I	1	1	0	0	0	0	0	1	•	1
J	1	1	1	1	0	0	0	1	1	•

M

Characters

Char.	A	B	C	D	E	F	G	H	I	J
A	•	7	6	6	5	5	5	2	2	4
B	7	•	6	6	5	5	5	2	2	4
C	6	6	•	6	5	5	5	1	3	3
D	6	6	6	•	5	5	5	1	3	3
E	5	5	5	5	•	5	5	0	2	4
F	5	5	5	5	5	•	5	0	2	4
G	5	5	5	5	5	5	•	0	2	4
H	2	2	1	1	0	0	0	•	1	1
I	2	2	3	3	2	2	2	1	•	3
J	4	4	3	3	4	4	4	1	3	•

D

Characters

Char.	A	B	C	D	E	F	G	H	I	J
A	•	0.13	0.25	0.25	0.38	0.38	0.38	0.75	0.75	0.50
B	0.13	•	0.25	0.25	0.38	0.38	0.38	0.75	0.75	0.50
C	0.25	0.25	•	0.25	0.38	0.38	0.38	0.88	0.63	0.63
D	0.25	0.25	0.25	•	0.38	0.38	0.38	0.88	0.63	0.63
E	0.38	0.38	0.38	0.38	•	0.38	0.38	1.00	0.75	0.50
F	0.38	0.38	0.38	0.38	0.38	•	0.38	1.00	0.75	0.50
G	0.38	0.38	0.38	0.38	0.38	0.38	•	1.00	0.75	0.50
H	0.75	0.75	0.88	0.88	1.00	1.00	1.00	•	0.88	0.88
I	0.75	0.75	0.63	0.63	0.75	0.75	0.75	0.88	•	0.63
J	0.50	0.50	0.63	0.63	0.50	0.50	0.50	0.88	0.63	•

FIGURE 2. Derivation of a dissimilarity matrix based on compatibility. The analysis begins with an initial taxon-by-character matrix (**R**) such as used in phylogenetic studies. A character-by-character matrix of pairwise compatibility then is constructed (**C**), where 1 indicates that characters i and j are compatible, and 0 indicates that i and j are incompatible. This is converted to a matrix of mutual compatibilities (**M**), in which each value $m_{i \cdot j}$ gives the number of characters with which both i and j are compatible in matrix **C**. (This equals $\mathbf{CC}^{\mathrm{T}}$, which is simply $\mathbf{C}^2$ for a square matrix). A dissimilarity matrix, **D**, is then constructed, $d_{i \cdot j} = 1 - \frac{m_{i \cdot j}}{n-2}$, where n is the number of characters (here, 10). **D** then is Gower-transformed (Gower, 1966) to matrix **T**, which is subjected to principal coordinates analysis.

determine character compatibility from character state distributions without reference to phylogeny.

These two theorems justify the construction of the pairwise compatibility matrix **C** (= **G** of Sneath et al., 1975) for use in clique analysis (e.g., Meacham, 1980). Given a cladistic data set **R** of p taxa and n characters, the pairwise compatibility matrix **C** is defined as an $n \times n$ matrix in which

$$c_{i \cdot j} = 1 \text{ if } i <|> j$$
$$= 0 \text{ if } i <> j$$

where <> represents a circuit and hence incompatible, and <|> represents no circuit and hence compatible. Figure 2 gives an example of **C** generated from artificial data.

The **C** matrix is the starting point for clique analysis (Le Quesne, 1969; Meacham, 1980). It also expresses global character compatibility (i.e., compatibility among all characters), which is the basis for tests evaluating hierarchical signal among characters (Le Quesne, 1969, 1982; Meacham, 1984, 1994; Sharkey, 1989, 1994) or within whole matrices (Alroy, 1994). Here it will be the starting point for both inferring and testing correlated charactersuites (Fig. 3).

In general, we expect characters that change frequently to have lower compatibilities than do characters that change rarely.

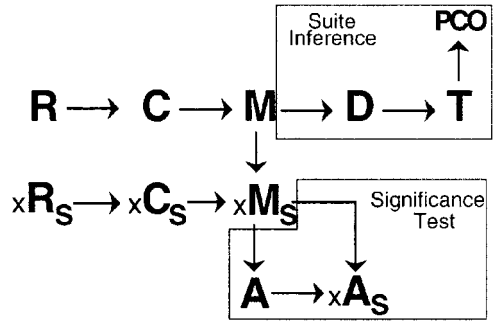

FIGURE 3. Summary flow chart of the matrices and analyses described in this paper. **R** is an empirical character matrix, **C** is the pairwise compatibility matrix, **M** is the mutual compatibility matrix, **D** is the dissimilarity matrix, and **T** is the Gower-transformed dissimilarity matrix. Principal coordinates analysis (PCO) of **T** provides a basis for inferring suites of correlated characters. For derivation of each matrix, see text and Figure 2. Monte Carlo simulations first produce simulated character matrices ($\mathbf{R}_S$) for the same number of taxa and characters. If global compatibility within $\mathbf{R}_S$ matches that of **R**, then $\mathbf{C}_S$ and $\mathbf{M}_S$ are calculated (see Fig. 2). This is repeated several thousand times. The significance of each observed $m_{i \cdot j}$ is determined by the frequencies of simulated m that equal or exceed the observed $m_{i \cdot j}$ when the simulated characters have the same compatibilities as the real characters. To account for multiple pairwise comparisons, the matrix of P-values (**A**) then is compared with a matirx of P-values from a second set of simulations.

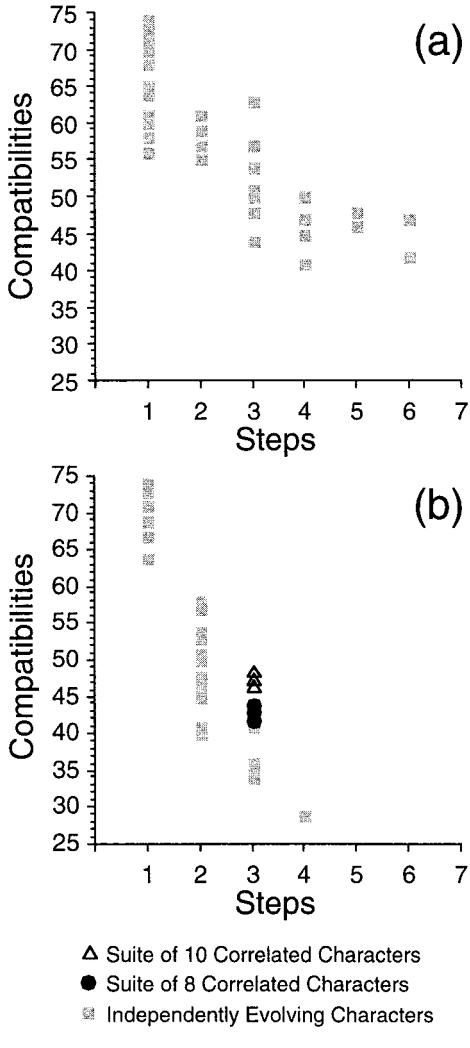

FIGURE 4. Properties of compatibility as shown by simulated character evolution with 75 binary characters among 25 taxa. (a) When all characters evolve independently, compatibility is strongly correlated with the number of steps. (b) When two suites of correlated characters exist, two patterns emerge. In one, correlated characters show greater compatibility than do other characters with the same number of changes. In the second, global compatibility decreases because of conflict among correlated suites and independent characters.

Simulations corroborate this. The example illustrated in Figure 4a uses the compatibilities of 75 independently evolving binary characters with 150 steps among 25 terminal taxa. (Simulations with multistate characters yield almost identical conclusions.)

We also expect characters in correlated suites to have greater compatibilities than do independently evolving characters that change the same numbers of times, because correlated characters will tend to be compatible with each other. Simulations also corroborate this expectation. The example shown in Figure 4b has the same number of taxa, characters, and steps as the previous example. However, the second example has two correlated character suites—one with 10 characters, the other with eight. Correlated change here is probabilistic, with the probability of change for the "dependent" characters greatly increased on branches once the first character in the suite changes. Even if the dependent characters fail to change on that branch, the probability of change remains greater on descendant branches. Because of this, correlated characters did not always have identical distributions. Note how characters from each correlated suite changing three times all have greater compatibility than do independently evolving characters changing three times. Simulations show that this is a general pattern.

Finally, simulations reveal that correlated character evolution lowers global compatibility (Fig. 4). This also makes sense, because we now have conflicting patterns among multiple characters suites instead of only a phylogenetic pattern plus random noise.

Shared Compatibility and a Coefficient of Association

Characters i and j have corresponding sets, i^c and j^c, consisting of the characters with which each is compatible. Thus, mutual compatibility is simply the intersection of i^c and j^c (see below). If mutual compatibilities are large relative to i^c and j^c, then characters i and j might have some affinity in terms of a compatible clique and would indicate support of a similar tree topology. If mutual compatibility is small relative to i^c and j^c, then characters i and j would not be members of the same clique and might have stronger affinities to other characters.

The largest clique of pairwise compatible characters in the compatibility matrix **C** should reflect phylogeny, whereas smaller cliques of pairwise compatible characters might represent subsidiary signals (Sneath et al., 1975). We propose to identify these cliques by using shared, or mutual, compatibility (Figs. 2, 3). The mutual compatibility between characters i and j ($m_{i,j}$) is defined as the sum of all characters that are compatible with both i and j. If characters i and j are

correlated, then we expect $m_{i \cdot j}$ to be large relative to the total compatibilities for characters i and j ($\mathbf{c}^i$ and $\mathbf{c}^j$, respectively). Accordingly we create a mutual compatibility matrix, **M**, in which cell values range from 0 to $n - 2$ (the maximum number of possible mutual compatibilities). One can calculate **M** either by summing mutual compatibilities or by using the cross-product matrix of **C** (i.e., $\mathbf{C}^2$) (Sneath et al., 1975).

Dissimilarity Matrices and Ordination: Inferring Correlated Suites

The question of interest actually concerns the multivariate structure of cladistic character data. Eigenvector analyses express multiple patterns of association among analyzed objects, and thus are appropriate here. However, **M** must be converted into a dissimilarity matrix (**D**) for eigenvector analysis. Each cell in **D** is

$$d_{i \cdot j} = 1 - \frac{m_j}{n - 2} \qquad (1)$$

(see Fig. 2). As mutual compatibilities between characters i and j increase, $d_{i \cdot j}$ converges on 0.

The dissimilarity matrix **D** is Gower-transformed (Gower, 1966) before eigenvector decomposition (Fig. 3). Each cell in the new matrix **T** first subtracts the mean dissimilarities of both characters i and j from each $d_{i \cdot j}$ and then adds twice the global average in dissimilarity to each $d_{i \cdot j}$. Gower transformation offers some standardization for the number of overall compatibilities: The transformation inflates the relative importance of high mutual compatibilities among characters with few overall compatibilities and deflates the relative importance of high mutual compatibilities among characters with numerous overall compatibilities. Principal coordinates analysis (PCO) is then used to extract eigenvalues and eigenvectors from **T**.

PCO may begin with other coefficients of similarity, such as the Jaccard (Cheetham and Hazel, 1969), Simpson (Simpson, 1960), and Ochiai (Ochiai, 1957) coefficients. These metrics would compare $m_{i \cdot j}$ with the total possible $m_{i \cdot j}$ for each two characters, rather than with the theoretical limit $n - 2$, and would thus emphasize high mutual compatibility among generally incompatible characters. However, this also means that the denominator will vary from one dissimilarity measure to the next. Variable denominators in a similarity coefficient yield matrices with triangle inequalities (i.e., one in which variables cannot be plotted in a Euclidean space given the measured distances; Strang, 1980), which in turn generate negative eigenvalues. For example, given an Euclidean space and distances of 5 from A to B and 6 from A to C, the distance from B to C can be no greater than 11 and no less than 1. Such matrices of Euclidean distances are semidefinite (Reyment and Jöreskog, 1996:141). Any space in which the distance from B to C is 0 would be non-Euclidean and not positive semidefinite. Large negative eigenvalues indicate vectors in imaginary space and complex warping of the remaining real vectors (Gower, 1966, 1971). Major triangle inequalities can "warp" distributions along eigenvectors with positive eigenvalues. Because we wish to use these distributions to infer correlated character suites, that result is undesirable.

Sneath et al. (1975:330) predicted that the first eigenvector (PO 1) of matrix **T** should reflect the overall compatibility of characters. We therefore expect highly compatible characters will "load" on one extreme of this axis, and the least compatible characters will "load" on the other extreme. In other words, PO 1 should be analogous to an axis of general size in a principal components analyses of morphometric data (see, e.g., Bookstein et al., 1985). This result is found in simulation matrices (Fig. 5a) and empirically (see below).

Sneath et al. (1975) also predicted that the effects of correlated evolution should be apparent on PO axes below the first. This is corroborated by analyzing the simulated matrix used in Figure 5b. Here, the first axis reflects overall compatibility and thus separates the most compatible characters from the least compatible. Axis two separates the first correlated suite (including 10 characters) from the least compatible characters. Axis three separates the second correlated (including 8 characters) from both the least compatible characters and the first suite. Thus, all three axes separate characters showing a pattern from those showing either no pattern or a conflicting pattern. The eigenvalues from this and other matrices simulated under the same parameters show two inflection points, one between the first and second axes and

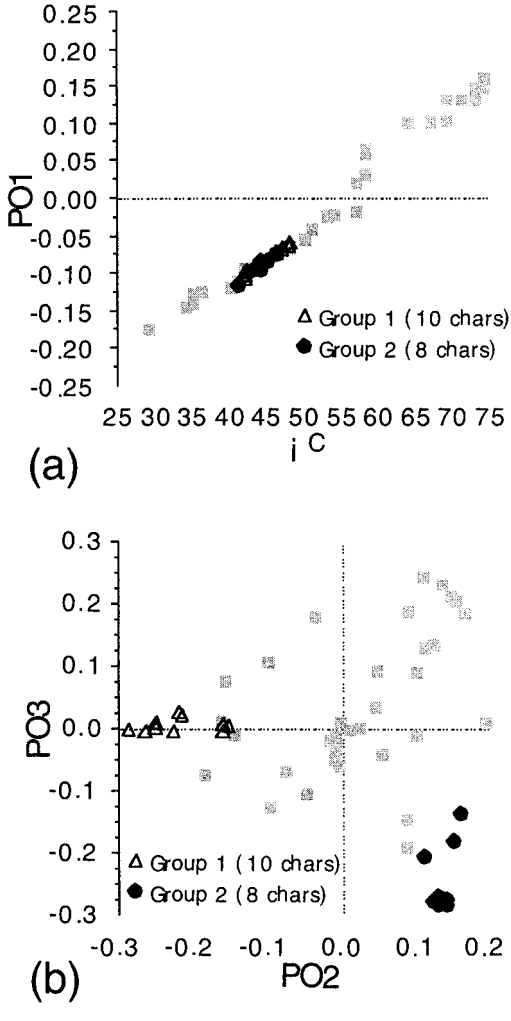

FIGURE 5. Multivariate properties of mutual compatibility, as revealed by PCO of mutual compatibility. This example uses the simulated matrix used for Figure 4b. (a) The primary signal reflects the relative compatibility of each character. (b) Axes 2 and 3 separate the two suites of correlated characters.

one between the third and fourth axes. In contrast, eigenvalues derived from matrices simulated under independent evolution (such as illustrated in Fig. 4a) show only one inflection (between axes one and two). For importance of this finding see below.

Eigenvector analyses always return a number of principal axes equal to the rank of **D**, regardless of whether these axes are meaningful. Jackson (1993) summarized several methods testing the null hypothesis that secondary axes form a multidimensional "sphere" and thus do not deviate from a random expectation. However, none of these tests was entirely satisfactory because the null hypothesis here is not random association, but rather that phylogenetic autocorrelation and chance alone produce the observed nonrandom association. Therefore, we use Monte Carlo simulations to generate sample data sets in which the total taxa, characters, and global compatibility match those of the original data set. The range of eigenvalues derived from such matrices provides a null distribution for evaluating the null hypothesis of independent character change.

These simulations evolve independent characters across phylogenies of the same diversity as the real clade. The simulations are based on bifurcating speciation and extinction, with the extinction rate being three-quarters of the speciation rate. If an extant clade is analyzed, then the simulations end when the number of coexisting "taxa" equals that of the real matrix. If a clade that includes fossils is analyzed, then sampling is done over time until the sampled number of "taxa" equals that of the real matrix. The number of changes per character is preset and usually is derived from the parsimony tree. This ensures that some heterogeneity exists both in rates and compatibilities among characters. The simulations use the same number of states as observed in the corresponding character in the real data. Because missing data affect compatibility, the number of taxa with unknown states for each character is also maintained. Character change occurs by randomly drawing one of the branches. If that branch is nodal, then the state change is distributed to all descendant branches. The probability of drawing a branch is determined by the number of unsampled ancestors along that branch, which makes change more probable along long branches. Long branches increase the probability of characters changing on the same branch, which mimics the effect of nonindependence and increases mutual compatibilities among independent characters (Wagner, 2000). This in turn increases the eigenvalues of the high secondary axes. The actual effect is very slight, but it has a conservative effect on our analyses. If global compatibility in a given simulated data set is within 1% of the number of real compatibilities, the character matrix is then converted into a **T** matrix by using the steps outlined above. Thus, difference in global compatibility should not affect the probabilities of mutual compatibility.

Eigenanalyses of **T** matrices derived from those simulations provide the range of eigenvalues expected, given a null hypothesis that phylogeny and rates of homoplasy are the sole determinants of compatibility. Thus, the distribution tests whether the secondary axes for the real data set deviate significantly from the expectation of independent character evolution.

Monte Carlo Simulations: Testing Hypotheses of Correlation

Numerous methods exist for testing whether previously defined sets of characters show random distributions along PCO axes (e.g., the Mann–Whitney test [Sokal and Rohlf, 1995:432] for two-character partitions or the Kruskal–Wallis test [Sokal and Rohlf, 1981:429] for three or more character partitions). However, if one uses the loadings of characters on PCO axes to infer sets of correlated characters a posteriori, it is logically circular to then examine the distributions of those loadings. Having developed an exploratory analysis of the compatibility structure, we now would like to recast the inference as a hypothesis and test a null hypothesis of independent character evolution. A procedure to do this is developed here, based again on Monte Carlo simulation (Fig. 2).

The dissimilarity information in **D** (and hence the compatibility structure in the initial character matrix, **R**) arises from the combination of at least three sources: (1) phylogenetic autocorrelation; (2) rates of homoplasy; and (3) correlated character change. Because we are concerned with only the third parameter here, we need an estimate of the expected distribution of mutual compatibilities for two characters i and j, given parameters i^c, j^c, the total number of characters and states, and a phylogeny. Superficially, this appears to be testable by combinatorics. However, the probabilities derived from combinatorics would assume completely random distributions of mutual compatibilities and would not account for nonrandom compatibility attributable to phylogeny alone. The phylogeny parameter causes the same problem here as in the generation of expected eigenvalue distributions described above; again, the appropriate null hypothesis is not random compatibility but nonrandom compatibility generated by independent character evolution. Phylogeny must therefore be modeled in some way so that it can be factored out.

In the absence of an analytic solution, we generated expected distributions of mutual compatibilities by using Monte Carlo simulations in which characters are evolved independently across simulated trees. The simulated distributions were calculated as follows:

1. n characters are evolved independently across a phylogeny of p taxa, where n and p are derived from the original matrix. These simulations are identical to those used in the test of eigenvalues (see above).
2. The compatibility of each simulated character is tallied (i.e., i^c).
3. If global compatibility (i.e., total number of compatible pairs) is within 1% of the observed compatibility, then each possible i and j pairwise comparison is examined. Two distributions are recorded:
   - 3a. The shared compatibilities between two compatible simulated characters i and j with i^c and j^c.
   - 3b. The shared compatibilities between two incompatible simulated characters i and j with i^c and j^c.

Separate distributions are tallied for compatible and incompatible character pairs. Suppose that character i has 60 compatibilities and character j has 45 compatibilities. If the characters are compatible, then they can share 44 compatibilities among the remaining $n-2$ characters. However, if they are incompatible, they can have 45 mutual compatibilities among the remaining $n-2$ characters. Moreover, the pairwise comparisons are not independent. If i and j are compatible and i and k are compatible, then the probability that j and k are compatible is greater than j^c and k^c alone would predict. Similarly, if i and j are incompatible whereas i and k are compatible, then the probability that j and k are compatible is less than j^c and k^c alone would predict. Thus, the distribution of expected mutual compatibilities differs slightly between compatible and incompatible pairs and should be examined.

The resulting distributions are the expectations for a null hypothesis in which phylogeny and random homoplasy determine

all mutual compatibilities. If mutual compatibilities exceed expectations derived from the simulations, then we reject the null hypothesis of independent character evolution in favor of correlated evolution. Two scenarios predict characters to have fewer mutual compatibilities than expected given independent evolution. One is negatively correlated evolution, in which the presence of a state in character i prohibits the evolution of a state in character j. A second is two characters belonging to different correlated suites, suites that, in turn, share few compatibilities. This situation will artificially inflate compatibility for both characters (because of compatibility within their suites) but will discourage mutual compatibilities. The actual mutual compatibilities of the two characters with characters outside the two suites should fit the expectations of two independent characters with far fewer overall compatibilities than are observed in either character. Thus, a significantly low value for $m_{i \cdot j}$ rejects the independent null hypothesis in favor of the hypothesis that the characters belong to conflicting character sets.

Multiple Pairwise Comparisons

Numerous unplanned comparisons present problems for significance tests (see discussion in Sokal and Rohlf, 1995:230). A pairwise compatibility matrix derived from n characters will contain $n^2/n - 2$ pairwise comparisons; a matrix of 78 characters will therefore contain 2,964 separate comparisons between characters, of which 5% (148) can be expected to be significant at $P = 0.05$ due to chance alone. The validity of individual tests of significance (i.e., individual comparisons between an observed $m_{i \cdot j}$ and the simulated distribution for that comparison, $m_{Si \cdot j}$,, is therefore impossible to establish. Furthermore, we expect some suites of three or more characters will have "significantly" high mutual compatibility simply by chance. Bonferonni corrections of various types are often used (e.g., Rice, 1989) to lower the P-values at which the null hypothesis can be rejected. However, Sokal and Rohlf (1995) note that Bonferonni corrections are overly conservative and thus promote Type II errors. Moreover, levels of significance derived from the Monte Carlo tests are limited by the number of characters and taxa. When there are few characters (e.g., $n \leq 20$), P-values rarely are <0.1, even when there are maximum mutual compatibilities. The statistical power of the tests described here is proportional to n and to the magnitudes of i^c and j^c.

Our solution is to repeat the Monte Carlo simulations. However, instead of recording distributions of mutual compatibilities, the second simulation uses the results of the first Monte Carlo test to determine P-values of each pairwise mutual compatibility for each simulated character. After determining the sets of characters in which all pairwise comparisons have P-values less than a set value (e.g., 0.05), the total number of characters belonging to such sets, the sizes of those sets, and the total number of steps per matrix are recorded. These distributions of sets and set sizes represent the expected number and size of "significantly" correlated subsets given independent character evolution. This expectation can then be compared with the actual profile of correlated set number and size, and the null hypothesis of independent change can be accepted or rejected.

Comparisons with Previous Compatibility Tests

Our Monte Carlo tests differ in important ways from permutation tests that use compatibility. Permutation analyses test whether compatibility deviates from an expectation generated from a null hypothesis of random data, either for particular characters (e.g., Meacham, 1984, 1994; Sharkey, 1989, 1994) or for groups of characters (e.g., Alroy, 1994; Wilkinson, 1998). The Monte Carlo analyses test whether the expectations of independent change across phylogeny are adequately met. Because phylogeny underlies the null distribution, characters and character compatibilities are not distributed randomly with respect to each other. When combined with the multivariate analyses described above, the analyses presented here differ from comparisons of matrices using compatibility (e.g., Wilkinson, 1998) by not requiring a priori definitions of character sets. The multivariate analyses can provide inferences of correlated characters, which then are recast as hypotheses and tested by the Monte Carlo methods.

A Test Case: Caecilians

Data

To demonstrate the above methods, we analyzed phylogenetic data from caecilian amphibians found in Wilkinson (1997). The caecilian data matrix contains 25 taxa scored for 78 morphological characters, 52 of which are "traditional" morphological characters supporting the accepted caecilian phylogeny derived from both morphological and molecular evidence. The remaining 26 "neuroanatomical" characters concern the presence and innervation of sensory structures in the head. A subset of these neuroanatomical characters involving the eye supports a different tree topology from that supported by the balance of characters. Wilkinson (1997) exhaustively analyzed this data set by using compatibility tests of data structure and parsimony analyses of different character partitions, and found a set of characters relating to the eye that exhibited correlated change. Tests such as the permutation tail probability (Faith, 1991) and compatibility tests (Alroy, 1994; Meacham, 1994) indicate that the neuroanatomical subset has less hierarchical structure than the traditional subset, although its signal still deviates significantly from random. These results led Wilkinson (1997) to conclude that the neuroanatomical subset records the convergent loss of eye structures as different caecilian groups evolved rudimentary eyes in response to a fossorial lifestyle. We reanalyzed these data to explore whether the methods proposed herein could identify sets of correlated characters similar to those hypothesized by Wilkinson.

Multivariate Analysis

We constructed the pairwise compatibility matrix **C** by tabulating pairwise compatibilities among all 78 characters (matrix available on Society of Systematic Biologists website: www.utexas.edu/ftp/depts/systbiol/). Compatibilities were assessed by assuming unordered evolution for multistate characters, which maximizes the compatibility of multistate characters (McMorris, 1975) and is conservative. The pairwise compatibility matrix was then used to construct matrices **M** (matrix available on SSB website), **D**, and **T** (see Fig. 2a), and **T** was analyzed by PCO. The matrix was positive semidefinite, with the scree plot (plot of the magnitude of eigenvalues by number) indicating inflections after the first axis and again after the fourth axis (Fig. 6a). This pattern of two inflections is observed in several other data sets we have analyzed (O'Keefe and Wagner, unpubl. analyses).

The eigenvalue distribution of caecilian compatibility deviates in several ways from the predictions of independent character evolution (Table 1). One difference is that the first eigenvalue (λ_1) summarizes more total "association" than expected, given the hypothesis of independent change. The same

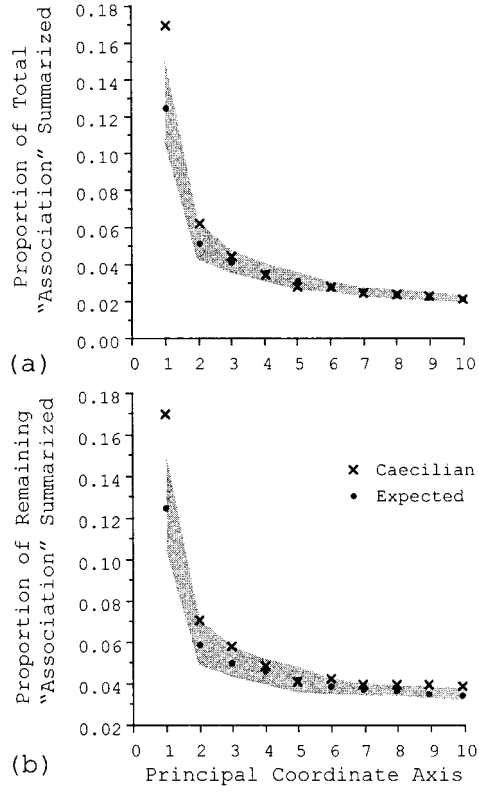

FIGURE 6. (a) Scree plot showing the distribution of eigenvalues for the first 10 eigenvectors. Dots mark the expected eigenvalue for a matrix with 2,327 ± 11 compatibilities, the gray area indicating the 95% confidence envelopes. X-eigenvalues for caecilian data. Only eigenvalue 1 exceeds this appreciably. Eigenvalues 2, 3, and 4 all exceed the expectation, but not significantly. (b) Describing eigenvalues as proportions of the remaining sum of eigenvalues tells a different story. For eigenvalue 2, this is the sum of eigenvalues 2 through 78; for eigenvalue three, this is the sum of eigenvalues 3 through 78; and so forth. Eigenvalues 2–4 all are larger than expected, given the remaining eigenvalues, and eigenvalues 2 and 3 are significantly larger than expected.

TABLE 1. The multivariate structure of mutual compatibilities. Observed eigenvalues (λ), standardized by the sum of all eigenvalues ($\sum_{i=1}^{i=77} \lambda_i$), are from real data. Expected eigenvalues, again standardized by the sum of all eigenvalues, reflect the averages from 100 simulations in which 78 independently evolving characters yielded matrices of compatibility similar to that of the caecilian matrix (i.e., 2,327 ± 11 compatible pairs). The P-value gives the proportion of such runs in which simulated standardized λ_i exceeded real standardized λ_i. Significance is assessed by the amount of remaining "association" summarized (i.e., $\sum_{i=\mathrm{PO}}^{i=77} \lambda_i$ for λ_{PO}).

PO	Obs. $\frac{\lambda_{\mathrm{PO}}}{\sum_{i=1}^{i=78} \lambda_i}$	Exp. $\frac{\lambda_{\mathrm{PO}}}{\sum_{i=\mathrm{PO}}^{i=77} \lambda_i}$	p	Obs. $\frac{\lambda_{\mathrm{PO}}}{\sum_{i=\mathrm{PO}}^{i=77} \lambda_i}$	Exp. $\frac{\lambda_{\mathrm{PO}}}{\sum_{i=\mathrm{PO}}^{i=77} \lambda_i}$	p
1	0.170	0.125	>0.99	0.170	0.125	>0.99
2	0.063	0.052	0.96	0.071	0.059	0.96
3	0.045	0.042	0.84	0.059	0.051	0.96
4	0.035	0.036	0.41	0.049	0.046	0.75
5	0.029	0.031	0.15	0.042	0.042	0.54
6	0.028	0.028	0.60	0.043	0.039	0.98
7	0.025	0.026	0.22	0.040	0.038	0.89
8	0.024	0.024	0.65	0.040	0.037	0.98
9	0.023	0.023	0.73	0.040	0.036	0.99
10	0.022	0.022	0.81	0.039	0.035	0.98

PO, Principal coordinate.

is true for λ_2 but not λ_3. Considering the proportion of remaining "association" (i.e., association not summarized by higher eigenvectors), one finds that λ_1, λ_2, and λ_3 summarize significantly more than expected, and that λ_4 summarizes more association than expected, but the difference is not significant (Fig. 6b). Another important difference between the observed and expected eigenvalue distributions is the two major inflections in the scree plot: One separates λ_1 from eigenvalue λ_2, and another separates the trend λ_2–λ_4 from λ_5–λ_{77}. Monte Carlo simulations indicate that phylogeny alone predicts a single inflection between λ_1 from λ_2. As already stated, only axes 1–3 are statistically significant. However, we include information and plots of axis 4 because the high value of λ_4 and its position before the second inflection in the scree plot indicate it has some meaning.

PO 1: The Compatibility Axis

As predicted by Sneath et al. (1975), a strong relationship exists between number of compatibilities and character position on PO 1 (Fig. 7a). Autapomorphies plot in the lower right because they are compatible with all characters, whereas characters with few compatibilities plot in the upper left. This is not surprising in light of the discussion on similarity coefficients above. Characters scoring as derived for relatively few taxa have a greater probability of being compatible with another character simply because of fewer opportunities for a circuit to form (Meacham, 1981). Autapomorphies are the limit of this tendency and are compatible with all other characters. A character i scoring as derived for few taxa will have a large set i^c and a correspondingly low $d_{i \cdot j}$ as long as j^c is also relatively large. PO 1 has the desirable property of recording the compatibility structure related to a character's overall compatibility, and hence is indirectly related to the differences in the number of taxa with the derived state per character. More importantly, dropping this axis corrects for the variation in the sizes of i^c and j^c in the calculation of $d_{i \cdot j}$ without sacrificing the semidefinite nature of the matrix. PO 1 also reveals differences in compatibility among Wilkinson's subsets (Fig. 7a).

PO 2–4: Patterns on Secondary Axes

Lower axes (e.g., PO 2) show no association with individual character compatibility (Fig. 7b). In concordance with Wilkinson (1997), the subset of eye characters clusters on PO 2 (Fig. 8; Table 2) and has a significantly different distribution on PO 2 than do the balance of characters (Mann–Whitney test, $P < 10^{-4}$). This result is expected for a set of correlated characters, and demonstrates how the axes can be used to test a priori hypotheses of character correlation. The ordination can also be used to explore the structure of data set compatibility, especially in concert with the simulation results outlined below. Based on the simulation analyses, three traditional characters also cluster within the eye character subset

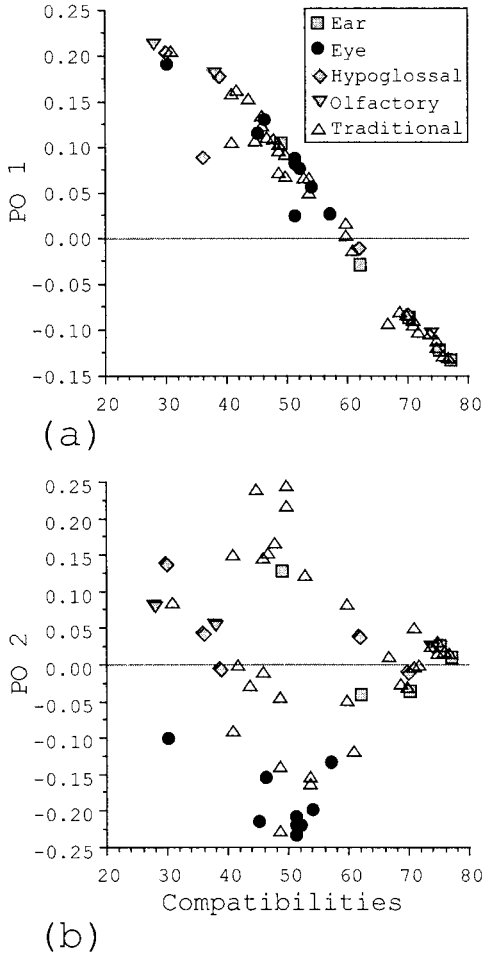

FIGURE 7. Association between raw compatibility and "loadings" on principal coordinate axes. (a) As predicted by Sneath et al. (1975), PO 1 reflects basic compatibility, with highly compatible characters receiving high, negative coefficients (note that the signs on eigenvectors are arbitrary). (b) No association is apparent between compatibility and PO 2.

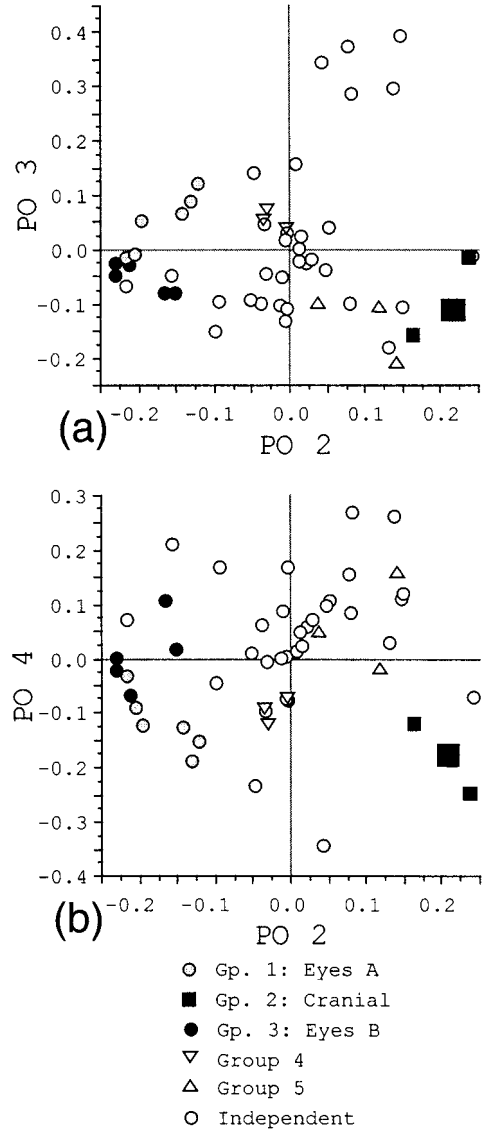

FIGURE 8. Separation of characters by PCO. Two other small partitions suggested by Monte Carlo tests (see Fig. 9) also are illustrated. PO 2 separates eye character suites from a cranial character suite. PO 3 and PO 4 show weak separation of two eye character suites.

(Table 2). One, the partial covering of the orbit by bone (T28), is also an eye character of sorts, but the other two, larval versus direct development (T43) and bicuspid or monocuspid anterior dentary teeth (T57), are not obviously connected to eyes. The axes also illustrate a more complicated pattern than postulated byWilkinson (1997). First, the eye character subset actually shows two separate suites that are weakly separated on PO 3 and PO 4. Second, a large suite of primarily cranial characters (e.g., T4: fusion of the premaxillae and nasals; T5: presence/absence of the septomaxillae; T6: presence/absence of the prefrontals; T16: Basipterygoid process strength) is opposed to the two suites of eye characters. The primary signal of PO 2 is to separate this cranial suite from the eye suites (Table 2). In contrast, PO 3 and PO 4 appear to separate the character suites from characters with low compatibility.

Minor Axes

The PO axes below the fourth (i.e., those associated with λ_5–λ_{77}) tend to separate out individual characters and seem to have little

TABLE 2. Characters clustering on the second, third, and fourth principal coordinate (PO) axes. Note the strong separation of eye characters from cranial characters on PO 2 and the weaker separation of different eye characters on PO 3 and PO 4. Character numbers after Wilkinson (1997).

No.	Character	PO 2	PO 3	PO 4
E1.2	Rectus internus: present/absent	−0.232	−0.022	−0.021
T28	Orbit covered by bone	−0.231	−0.048	0.003
E10	Lens: present/rudimentary/lost	−0.218	−0.066	0.072
E1.4	Rectus oblique: present/absent	−0.218	−0.014	−0.030
E1.3	Rectus superior: present/absent	−0.213	−0.026	−0.067
E1.6	Inferior oblique: present/absent	−0.207	−0.009	−0.089
E1.1	Rectus externus: present/absent	−0.207	−0.009	−0.089
E1.5	Superior oblique: present/absent	−0.196	0.054	−0.121
T40	Premaxillary: maxillary teeth size	−0.167	−0.079	0.110
T54	Terminal keel: absent/present	−0.158	−0.046	0.211
E6	Retinal cells: >5000/ <5000	−0.152	−0.078	0.018
T57	Anterior dentaries: bi/monocuspid	−0.144	0.066	−0.126
E3	Well-developed optic nerve	−0.133	0.089	−0.186
T43	Larva versus direct development	−0.122	0.123	−0.152
T56	*M. interhyoideous* posterior length	0.163	−0.156	−0.117
T16	Basipterygoid process strength	0.212	−0.109	−0.174
T4	Premaxillae: nasals unfused/fused	0.212	−0.109	−0.174
T5	Septomaxillae: present/absent	0.212	−0.109	−0.174
T31	Vent longitudinal/circular	0.212	−0.109	−0.174
T6	Prefrontals: present/absent	0.235	−0.013	−0.243
T20b	Ceratobranchials: fused/unfused	0.240	−0.011	−0.069

utility. However, the observed axes deviate from the expectations of independent evolution in two noteworthy ways. First, the real data have only 76 eigenvalues >0, whereas all simulated sets have 77 positive eigenvalues. A single eigenvalue of 0 is expected (i.e., λ_{78}), because the coefficient $d_{i \cdot j}$ is calculated with 77 comparisons rather than 78. However, the additional eigenvalue of 0 indicates that that some redundancy exists in the matrix, as would be expected if correlated change occurred.

Second, λ_6–λ_{15} all are slightly greater than expected, especially given the reduced amount of "association" remaining to be summarized. Conversely, eigenvalues λ_{20}–λ_{52} all are less than expected. The corresponding eigenvectors all separate single characters from the remaining characters for the real data, which is a pattern not seen on comparable axes when evolution is simulated. Thus, the patterns probably are not germane to the question of independent character evolution. However, they do suggest that actual evolution was more complicated among caecilians than in the Monte Carlo simulations.

Monte Carlo Tests of Character Independence

The first set of Monte Carlo tests reveals five suites of three or more characters with improbably high mutual compatibilities. The largest of these (the first eye suite) includes seven characters (Table 3), whereas

TABLE 3. Largest partition of characters with improbably high mutual compatibilities. Character labels and numbers after Wilkinson (1997). *P*-values based on the proportion of Monte Carlo simulations where characters had $m_{i \cdot j}$ or more mutual compatibilities, given that they had i^c and j^c compatibilities and were either compatible or incompatible (see compatibility matrices on SSB web site: www.utexas.edu/ftp/depts/systbiol/).

			P						
No.	Character	i^c	E1.1	E1.3	E1.5	E1.6	E3	T43	T57
E1.1	Rectus externus: present/absent	51		0.004	0.018	0.046	0.031	0.047	0.022
E1.3	Rectus superior: present/absent	45	0.004		0.036	0.004	0.041	0.049	0.047
E1.5	Superior oblique: present/absent	54	0.018	0.036		0.018	0.015	0.023	0.004
E1.6	Inferior oblique: present/absent	51	0.046	0.004	0.018		0.031	0.047	0.022
E3	Well-developed optic nerve	57	0.031	0.041	0.015	0.031		0.018	0.022
T43	Larva vs. direct development	61	0.047	0.049	0.023	0.047	0.018		0.017
T57	Anterior dentaries: bi/monocuspid	49	0.022	0.047	0.004	0.022	0.022	0.017	

TABLE 4. Second largest partition of characters with improbably high mutual compatibilities. See Table 3 for details. Character labels after Wilkinson (1997).

			P					
No.	Character	i^c	T4	T5	T6	T16	T31	T56
T4	Premaxillae: nasals unfused/fused	50		0.018	0.013	0.018	0.018	0.017
T5	Septomaxillae: present/absent	50	0.018		0.013	0.018	0.018	0.017
T6	Prefrontals: present/absent	45	0.013	0.013		0.013	0.013	0.050
T16	Basipterygoid process strength	50	0.018	0.018	0.013		0.018	0.017
T31	Vent longitudinal/circular	50	0.018	0.018	0.013	0.018		0.017
T56	*M. interhyoideous* posterior length	48	0.017	0.017	0.050	0.017	0.017	

the cranial suite includes six characters (Table 4), and the second eye suite includes five (Table 5). The remaining partitions include three characters each.

The second set of Monte Carlo simulations indicates that five partitions are not an unusually high number, given independent character evolution ($P = 0.31$). However, the total number of characters in these partitions (23) is unusually high ($P = 0.023$; Fig. 9a) as is a single suite of 7+ characters ($P = 0.015$; Fig. 9b). Partitions of 6+ characters are fairly probable ($P = 0.11$), but not given a partition of 7+ characters ($P < 0.001$, Fig. 9c). Finally, a third partition of 5+ characters is also improbable, given two partitions of 6+ characters ($P < 0.01$; Fig. 9d). However, a fourth or fifth partition of 3+ characters is not improbable ($P = 0.57$ and 0.39, respectively).

DISCUSSION

Implications for Caecilian Phylogeny: Weighting and Character Choice

Suites of characters that evolve dependently violate the assumption of character independence underlying phylogenetic reconstruction. The presence of dependent characters will exaggerate the apparent support for some nodes, whether those nodes are correct or incorrect. The possibility that parsimony will link taxa incorrectly also increases with character correlation, because correlated parallelisms will mimic phylogenetic autocorrelation. The obvious solution is to deweight the characters in the suite to remove the bias introduced by the correlation (Chippindale and Wiens, 1994); this approach has been taken by molecular workers (e.g., Wheeler and Honeycutt, 1988; Dixon and Hillis, 1993) to account for nonindependence in ribosomal DNA sequence data.

Any method that can identify dependent characters before tree building can provide a basis for objective character weighting (Sneath et al., 1975). However, one cannot easily do this while coding correlated characters separately. Parsimony approximates a likelihood solution in which the probability of change for each character is the same on each branch (Edwards and Cavalli-Sforza, 1964; Felsenstein, 1981). However, if characters are correlated, then the probability of change is partially determined by the states or state changes (or both) of other characters and therefore will vary over the tree. An alternative is to code correlated suites as a single compound character, with different combinations of states from each character representing a state in the compound character. The step-matrix of the compound character (Sankoff and Rousseau, 1975) would weight transitions such as $\{00\} \rightarrow \{01\}$ and $\{00\} \rightarrow \{11\}$ as little more than a single step.

TABLE 5. Third largest partition of characters with improbably high mutual Compatibilities. See Table 3 for details. Character labels after Wilkinson (1997).

			P				
No.	Character	i^c	E1.2	E1.4	E6	T28	T40
E1.2	Rectus internus: present/absent	51		0.008	0.009	0.009	0.038
E1.4	Rectus oblique: present/absent	52	0.008		0.014	0.010	0.045
E6	Retinal cells: >5000/<5000	46	0.009	0.014		0.005	0.021
T28	Orbit covered by bone	49	0.009	0.010	0.005		0.010
T40	Premaxillary: maxillary teeth size	54	0.038	0.045	0.021	0.010	

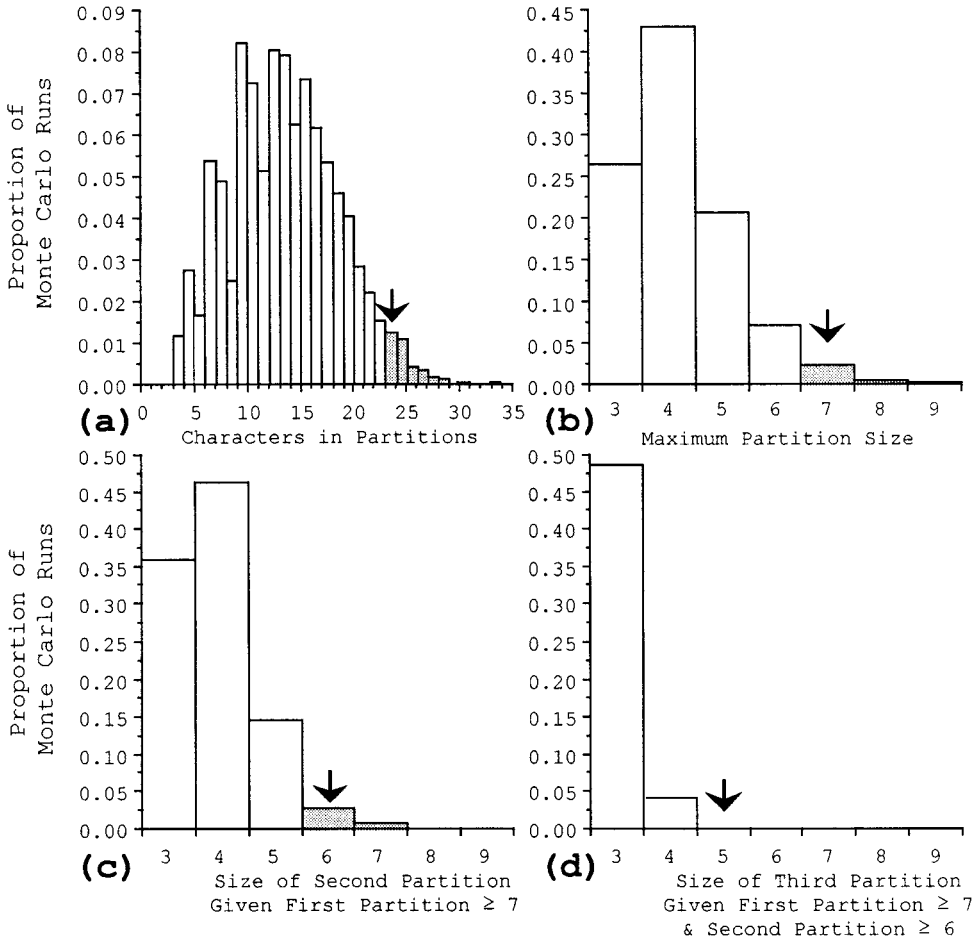

FIGURE 9. Monte Carlo distributions of partition sizes for 78 characters and 2,338 ± 23 compatible character pairs. The probability of pairwise mutual compatibilities of all characters in each partition is 0.05 or less. Arrows denote conditions for the caecilian matrix. (a) Expected numbers of characters in partitions. (b) Expected size of largest partition. (c) Expected size of second partition, given that the first partition has seven or more characters. (d) Expected size of the third partition, given that the first and second partitions have six or more characters each.

However, transitions such as {01} → {11} or {10} → {11} would cost much less than one step, which would greatly decrease the effects of correlated homoplasy. Methods for objectively assigning weights within such a step matrix have yet to be developed.

Another alternative is the reduction of each correlated character set to a single exemplar character (see, e.g., Werdelin and Solounias, 1991). In the case of caecilians, this approach reduces three sets to three characters. In all three cases, the character that could diagnose the most taxa as a synapomorphy was chosen. Reducing the correlated suites decreases the apparent homoplasy slightly but results in a topology somewhat different from that of the original parsimony tree (compare Figs. 10a and 10b). The reduced homoplasy does not, however, increase resolution: Resolution actually decreases, resulting in 1,616 most-parsimonious trees (MPTs), whereas the untreated matrix results in only 35 MPTs. This difference, however, is not surprising; as noted above, an expected effect of correlated characters is stronger support for erroneous nodes. Improved support also should reduce the chance that alternative nodes will be equally parsimonious. Hillis and Huelsenbeck (1992) demonstrated that the addition of even random characters will increase resolution, so reducing the number of characters should decrease resolution, regardless of the value of those characters.

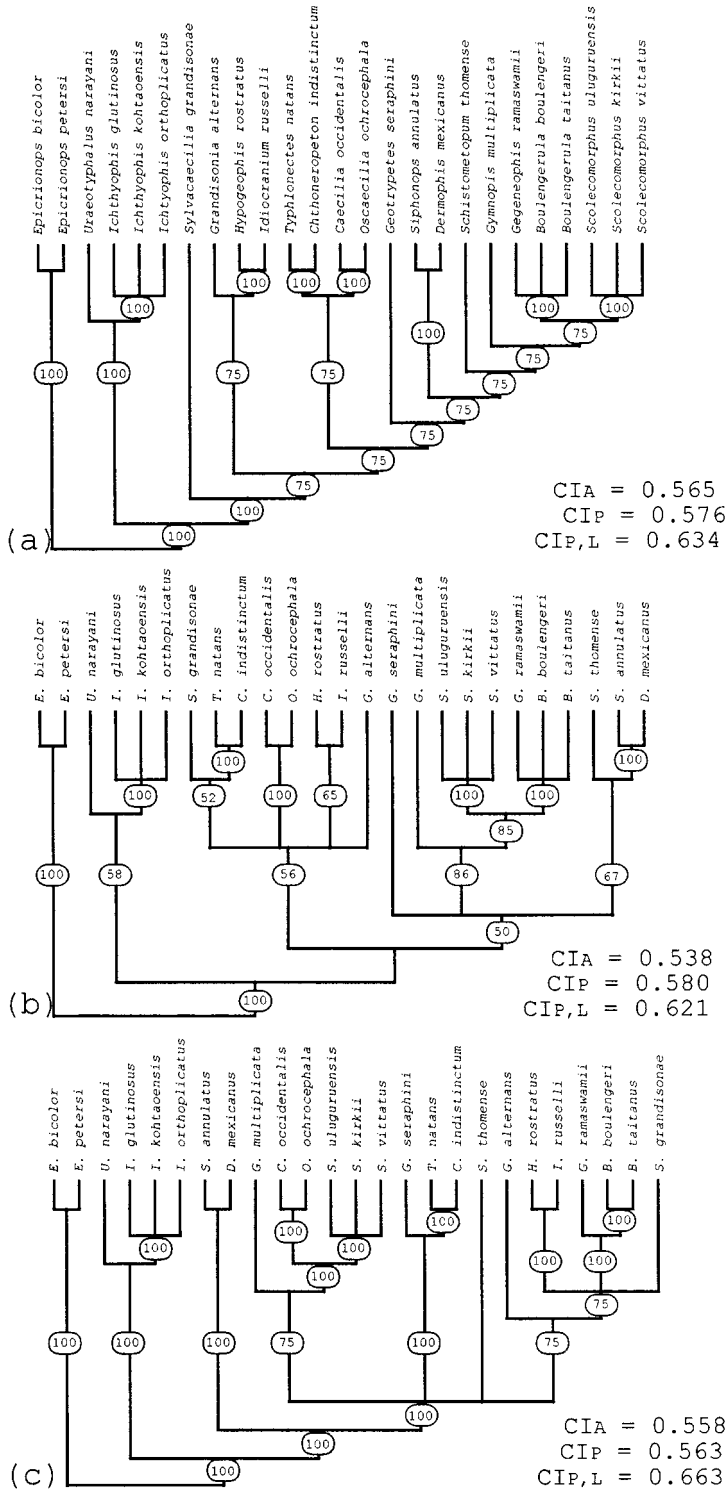

FIGURE 10. The effects of correlated characters on inferred phylogeny and rates of character evolution. Topologies depicted are the majority-rule consensus trees of all MPTs from each analysis. CI_A gives the consistency index for all 78 characters on each tree. CI_P gives the consistency index for 62 characters—that is, those not in one of the three partitions plus one character from each partition. $CI_{P,L}$ gives the consistency index for 50 characters not in partitions, the compatibilities of which are significantly greater than predicted by random distributions. (a) Majority rule consensus of 35 MPTs based on all 78 characters. (b) Majority rule consensus of 1,616 MPTs based on 62 characters, with correlated suites reduced to one character each. (c) Majority rule consensus of 8 MPTs based on 50 characters, with correlated suites reduced to one character each, and all characters for which the hypothesis of random distribution is not rejected (from Wilkinson, 1997) are excluded.

Le Quesne permutation tests by Wilkinson (1997) revealed that 18 caecilian characters had compatibilities so low that the hypothesis in which they were distributed at random with respect to the other characters could not be rejected. Because characters bearing phylogenetic signal should not be distributed randomly within a matrix, Wilkinson's final analyses simply eliminated those characters. We eliminated those characters from this analysis after reducing the correlated suites to single characters, and then reran the parsimony analysis. This analysis yielded far fewer trees (8) and showed several additional topology changes (Fig. 10c). The combined results suggest that highly homoplastic characters obfuscated phylogenetic inference (as suggested by the large numbers of trees), whereas homoplasy among correlated suites presented strongly misleading signal (as suggested by the small number of different trees in the initial analysis). This conclusion is supported by the existence of a subclade supported by "suspicious" eye characters (*Gymnopis*, *Gegeneophis*, *Boulengerula*, and *Scolecomorphus*; Wilkinson, 1997, pers. comm.). That clade is present in the first and second analyses but is broken up in the third (preferred) analysis. This result demonstrates how partitions identified by the methods developed here can be used to improve a phylogenetic hypothesis through character weighting.

Implications for Total Evidence

Kluge and Wolf (1993:112) acknowledge correlated character evolution might mislead parsimony analyses, but they deny that one can recognize such patterns even with known phylogenies. Kluge and Wolf's claim leads to the fundamental assumption of the "total evidence" paradigm; that is, there are no natural partitions among characters. However, the analyses presented here show that one can both recognize suites of potentially correlated characters and test alternative hypotheses of independent character evolution. These analyses recast a basic assumption of parsimony analyses as a testable hypothesis. If all character divisions are truly arbitrary, then data sets such as the caecilian example should not exist. The fact that they do falsifies Kluge and Wolf's premise. Some workers might worry that eliminating characters will eliminate phylogenetic signal (e.g., O'Leary and Geisler, 1999). However, if the hypothesis of independent character evolution is refuted, then retaining dependent characters increases the risk that character congruence will reflect homoplasy rather than homology. Simulation studies indicate that this risk is substantial even if character change is independent. Correlated change can only exacerbate this problem.

Limitations of Multivariate and Monte Carlo Analyses

One limitation of the methods we propose is that they apply only to matrices possessing a substantial amount of compatibility. With luck, this will not be a concern for most phylogenetic data sets. However, examples do exist of matrices with insufficient compatibility for these tests to be applicable. Wagner (2000) documented trilobite clades in which fewer than 10% of the character pairs were compatible. In some of these clades, most of the characters were incompatible with every other character. Even if correlated evolution occurred among these characters, it could not be detected by these tests. A possible solution to this problem is to describe compatibility not as a Boolean character, but instead as a fraction, that is, the largest subset of taxa in which the characters remain compatible divided by the total number of taxa. Difficulties arising with similarity matrix calculation remain unresolved for this approach. "Fuzzy compatibility" is an area for further research.

Several cautions about the use of these tests arise for technical reasons. The first concerns the statistical power of the Monte Carlo tests, which is determined by the magnitude of n. If a data set contains fewer than about 20 characters, then the tests lack the power to reject the null hypothesis of character independence. The number of taxa in a data set is also important; at least four taxa are required for assessment of compatibility, and more are advisable—although the sensitivity of the analyses to taxon number has not yet been investigated. A simulation study investigating the performance of the methods developed here over a range of evolutionary parameters would be useful and is an area of further study. Lastly, preliminary analysis of a large data set (34 taxa, 166 characters) of plesiosaurs (O'Keefe, 2000) has indicated that autapomorphic and invariant

characters slightly affect ordinations. Such characters, therefore, should probably be removed before analysis to ensure maximum clarity in the ordination. Autapomorphies do not affect Monte Carlo tests.

Conclusions

The methods developed in this paper can test a priori hypotheses of character correlation without reference to a cladogram. They also allow one to infer correlated character suites and test them without reference to a specific tree topology. Multivariate ordination of caecilian characters based on mutual compatibility reproduces Wilkinson's (1997) hypothesis that caecilian eye characters did not evolve independently. Subsequent Monte Carlo tests reject the hypothesis of character independence. Our methods suggest that two suites include eye characters as well as a few "traditional" characters, and an additional suite includes cranial characters. Monte Carlo tests also reject hypotheses of independent character evolution within these sets. Reanalysis of the caecilian data indicates that the correlated suites affect parsimony inferences about caecilian relationships. Reduction of correlated subsets and of homoplastic characters results in better phylogenetic resolution in the resulting cladogram and breaks up a subclade supported by correlated homoplastic characters.

In the case of the caecilians, there was prior reason to suspect correlated character evolution. However, the methods we used here require neither prior suspicion to hypothesize correlated character suites nor a cladogram topology to test those hypotheses. This ability is very important for both theoretical and methodological reasons. Methods that can identify interesting suites of characters could be useful for testing a variety of macroevolutionary hypotheses concerning the importance of development and function. This will be especially true when studying taxa for which very little information is available about developmental and functional biology (e.g., extinct or rare taxa). For systematists, the tests proposed herein can identify and reject hypotheses pertaining to particular sets of characters in comparison with an expectation of independent character evolution. Deweighting correlated characters suites offers a means of improving phylogenetic inferences. For biologists in general, objective means of identifying and testing correlated character suites will be useful for testing a range of developmental, functional, and architectural hypotheses.

Acknowledgments

The concepts regarding the ordination of compatibility-based coefficients of association were devised by the senior author as part of his dissertation research at the University of Chicago. We note that John Alroy independently developed a similar multivariate approach using principal components analyses of correlation coefficients, which leads us to reject a null hypothesis of independent lines of reasoning among U.C. alumni. M. Wilkinson, J. Wiens, and an anonymous reviewer provided thorough and very helpful reviews. For discussion and comments, we thank B. Chernoff, R. Cipriani, A. Dudek, T. Felix, M. Foote, J. Hopson, G. Hunt, D. Jablonski, M. LaBarbera, R. Lupia, J. Marcot, and O. Rieppel. A course taught by the late J. J. Sepkoski, Jr., introduced both authors to many of the concepts in this paper. The contributions of P.J.W. were funded by NSF Research Grant EAR-9903238. An Area 271 publication.

References

ALROY, J. 1994. Four permutation tests for the presence of phylogenetic structure. Syst. Biol. 43:430–437.

BOOKSTEIN, F. L., B. CHERNOFF, R. ELDER, J. HUMPHRIES, G. SMITH, AND R. STRAUSS. 1985. Morphometrics in evolutionary biology—the geometry of size and shape changes, with examples from fishes. Academy of Natural Sciences of Philadelphia, Philadelphia.

CAMIN, J. H., AND R. R. SOKAL. 1965. A method for deducing branching sequences in phylogeny. Evolution. 19:311–326.

CHEETHAM, A. H., AND J. E. HAZEL. 1969. Binary (presence–absence) similarity coefficients. J. Paleontol. 43:1130–1136.

CHIPPINDALE, P. T., AND J. J. WIENS. 1994. Weighting, partitioning, and combining characters in phylogenetic analysis. Syst. Biol. 43:278–287.

DE QUEIROZ, A. 1993. For consensus (sometimes). Syst. Biol. 42:368–372.

DIXON, M. T., AND D. M. HILLIS. 1993. Ribosomal RNA secondary structure: Compensatory mutations and implications for phylogenetic analysis. Mol. Biol. Evol. 10:256–267.

EDWARDS, A. W. F., AND L. L. CAVALLI-SFORZA. 1964. Reconstruction of evolutionary trees. Pages 67–76 *in* Phenetic and phylogenetic classification (J.H. Heywood and J. McNeil, eds.). Systematic Association, London.

EMERSON, S. B., AND P. A. HASTINGS. 1998. Morphological correlations in evolution: Consequences for phylogenetic analysis. Q. Rev. Biol. 73:141–162.

ESTABROOK, G. F., C. S. JOHNSON, AND F. R. MCMORRIS. 1976. A mathematical foundation for the analysis of cladistic character compatibility. Math. Biosci. 29:181–187.

FAITH, D. P. 1991. Cladistic permutation tests for monophyly and nonmonophyly. Syst. Zool. 40:366–375.

FELSENSTEIN, J. 1973. Maximum-likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. Syst. Zool. 22:240–249.

FELSENSTEIN, J. 1981. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. Biol. J. Linn. Soc. 16:183–196.

FELSENSTEIN, J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

GOWER, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–338.

GOWER, J. C. 1971. A general coefficient of similarity and some of its properties. Biometrics 27:857–874.

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

HUELSENBECK, J. P., AND R. NIELSEN. 1999. Effects of nonindependent substitution on phylogenetic accuracy. Syst. Biol. 48:317–328.

JACKSON, D. A. 1993. Stopping rules in principal components analysis: A comparison of heuristical and statistical approaches. Ecology 74:2204–2215.

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

KLUGE, A. G., AND A. J. WOLF. 1993. Cladistics: What's in a word? Cladistics 9:183–199.

LE QUESNE, W. J. 1969. A method of selection of characters in numerical taxonomy. Syst. Zool. 18:201–205.

LE QUESNE, W. J. 1982. Compatibility analysis and its applications. Zool. J. Linn. Soc. 74:267–275.

MADDISON, D. R. 1991. The discovery and importance of multiple islands of most-parsimonious trees. Syst. Zool. 40:315–328.

MADDISON, W. P. 1990. A method for testing the correlated evolution of two binary characters: Are gains or losses concentrated on certain branches of a phylogenetic tree? Evolution 44:539–557.

MCCRACKEN, K. G., J. HARSHMAN, D. A. MCCLELLAN, AND A. D. AFTON. 1999. Data set incongruence and correlated character evolution: An example of functional convergence in the hind-limbs of stifftail diving ducks. Syst. Biol. 48:683–714.

MCMORRIS, F. R. 1975. Compatibility criteria for cladistic and qualitative taxonomic characters. Pages 399–415 *in* Proceedings of the Eighth Annual Freeman International Congress on Numerical Taxonomy (B. F. Estabrook, ed.). W. H. Freeman, San Francisco.

MEACHAM, C. A. 1980. Phylogeny of the Berberidaceae with an evaluation of classifications. Syst. Bot. 5:149–172.

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

MEACHAM, C. A. 1984. Evaluating characters by character compatibility analysis. Pages 152–165 *in* Cladistics: Perspectives on the reconstruction of evolutionary history (T. Duncan, and T. F. Stuessy, eds.). Columbia Univ. Press, New York.

MEACHAM, C. A. 1994. Phylogenetic relationships at the basal radiation of angiosperms: Further study by probability of character compatibility. Syst. Bot. 19:506–522.

MICKEVICH, M. F. 1978. Taxonomic congruence. Syst. Zool. 27:143–158.

OCHIAI, A. 1957. Zoogeographic studies on soleoid fishes found in Japan and its neighbouring regions. Bull. Jpn. Soc. Sci. Fish 22:526–530.

O'KEEFE, F. R. 2000. Phylogeny and convergence in the Plesiosauria (Reptilia: Sauropterygia). Ph.D. Thesis, Univ. of Chicago, Chicago.

O'LEARY, M. A., AND J. H. GEISLER. 1999. The position of Cetacea within Mammalia: Phylogenetic analysis of morphological data from extinct and extant taxa. Syst. Biol. 48:455–490.

PAGEL, M. D. 1994. Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. Proc. R. Soc. Lond. B 255:37–45.

QUICKE, D. L. J., AND R. BELSHAW. 1999. Incongruence between morphological data sets: An example from the evolution of endoparasitism among parastic wasps (Hymenoptera: Braconidae). Syst. Biol. 48:436–454.

RAUP, D. M. 1966. Geometric analysis of shell coiling: General problems. J. Paleontol. 40:1178–1190.

RAUP, D. M., AND S. J. GOULD. 1974. Stochastic simulation and evolution of morphology—Towards a nomothetic paleontology. Syst. Zool. 23:305–322.

REYMENT, R. A., AND K. G. JÖRESKOG. 1996. Applied factor analysis in the natural sciences. Cambridge Univ. Press, Cambridge, England.

RICE, W. R. 1989. Analyzing tables of statistical data. Evolution 43:223–225.

SANKOFF, D., AND P. ROUSSEAU. 1975. Locating the vertices of a Steiner tree in arbitrary space. Math. Progr. 9:240–246.

SHAFFER, H. B., J. M. CLARK, AND F. KRAUS. 1991. When molecules and morphology clash: A phylogenetic analysis of the North American ambystomatid salamanders (Caudata: Ambystomatidae). Syst. Zool. 40:284–303.

SHARKEY, M. J. 1989. A hypothesis-independent method of character weighting for cladistic analysis. Cladistics 5:63–86.

SHARKEY, M. J. 1994. Discriminate compatibility measures and the reduction routine. Syst. Biol. 43:526–542.

SIMPSON, G. G. 1960. Notes on the measurement of faunal resemblance. Am. J. Sci. 258a:300–311.

SNEATH, P. H. A., M. J. SACKIN, AND R. P. AMBLER. 1975. Detecting evolutionary incompatibilities from protein sequences. Syst. Zool. 24:311–332.

SNEATH, P. H. A., AND R. R. SOKAL. 1973. Numerical taxonomy. W. H. Freeman, San Francisco.

SOKAL, R. R., AND F. J. ROHLF. 1995. Biometry, 3rd edition. W. H. Freeman, New York.

STRANG, G. 1980. Linear algebra, 2nd edition. Academic Press, New York.

SUTER, S. J. 1994. Cladistic analysis of cassiduloid echinoids: Trying to see the phylogeny for the trees. Biol. J. Linn. Soc. 53:31–72.

TURNER, B. J. 1974. Genetic divergence of Death Valley pupfish species: Biochemical versus morphological evidence. Evolution 28:281–284.

WAGNER, P. J. 1998. A likelihood approach for estimating phylogenetic relationships among fossil taxa. Paleobiology 24:430–449.

WAGNER, P. J. 2000. Exhaustion of cladistic character states among fossil taxa. Evolution 54:365–386.

WAINWRIGHT, S. A., W. D. BIGGS, J. D. CURREY, AND J. M. GOSLINE. 1975. Mechanical design in organisms. John Wiley and Sons, New York.

WAKE, D. B. 1989. Phylogenetic implications of ontogenetic data. Geobios, Mém. Spéc. 12:369–378.

WERDELIN, L., AND N. SOLOUNIAS. 1991. The Hyaenidae: Taxonomic systematics and evolution. Fossils Strata 30:1–104.

WHEELER, W. C., AND R. L. HONEYCUTT. 1988. Paired sequence difference in ribosomal RNAs: Evolutionary and phylogenetic implications. Mol. Biol. Evol. 5:90–96.

WILKINSON, M. 1997. Characters, congruence and quality: A study of neuroanatomical and traditional data in caecilian phylogeny. Biol. Rev. 72:423–470.

WILKINSON, M. 1998. Split support and split conflict randomization tests in phylogenetic inference. Syst. Biol. 47:673–695.

WRAY, G. A. 1996. Parallel evolution of nonfeeding larvae in echinoids. Syst. Biol. 45:308–322.

Received 24 March 2000; accepted 23 August 2000
Associate Editor: J. Wiens